

Preparing for the Worst But Hoping for the Best: Robust (Bayesian) Persuasion*

Piotr Dworczak and Alessandro Pavan[†]

July 20, 2022

Abstract

We propose a robust solution concept for Bayesian persuasion that accounts for the Sender’s concern that her Bayesian belief about the environment—which we call the *conjecture*—may be false. Specifically, the Sender is uncertain about the exogenous sources of information the Receivers may learn from, and about strategy selection. She first identifies all information policies that yield the largest payoff in the “worst-case scenario,” i.e., when Nature provides information and coordinates the Receivers’ play to minimize the Sender’s payoff. Then, she uses the conjecture to pick the optimal policy among the worst-case optimal ones. We characterize properties of robust solutions, identify conditions under which robustness requires separation of certain states, and qualify in what sense robustness calls for more information disclosure than standard Bayesian persuasion. Finally, we discuss how some of the results in the Bayesian persuasion literature change once robustness is accounted for, and develop a few new applications.

Keywords: persuasion, information design, robustness, worst-case optimality

JEL codes: D83, G28, G33

*For comments and useful suggestions, we thank Emir Kamenica, Shaowei Ke, Stephen Morris, Eran Shmaya, Ron Siegel, and seminar participants at various institutions where the paper was presented. Pavan also thanks NSF for financial support under the grant SES-1730483. Matteo Camboni provided excellent research assistance. The usual disclaimer applies.

[†]Department of Economics, Northwestern University

1 Introduction

“I am prepared for the worst but hope for the best,” Benjamin Disraeli, 1st Earl of Beaconsfield, UK Prime Minister.

In the canonical Bayesian persuasion model, a Sender designs an information structure to influence the behavior of a Receiver. The Sender is Bayesian, and has beliefs over the Receiver’s prior information as well as the additional information the Receiver might acquire after observing the realization of the Sender’s signal. As a result, the Sender’s optimal signal typically depends on the details of her belief about the Receiver’s learning environment.

In many applications, however, the Sender may be concerned that her belief—which we call a *conjecture*—is wrong. In such cases, the Sender may prefer to choose a policy that is not optimal under her conjecture but that protects her well in the event her conjecture turns out to be false.

In this paper, we propose a solution concept for the persuasion problem that accounts for the uncertainty that the Sender may face over the Receiver’s learning environment and that incorporates the Sender’s concern for the validity of her conjecture. Specifically, we assume that the Sender discards all policies that do not provide her with the optimal payoff guarantee when her conjecture is wrong. The payoff guarantee is computed conservatively by considering all possible learning environments for the Receiver, without assuming that the Sender is last to speak or that the Receiver will break indifferences in the Sender’s favor. We characterize properties of “robust solutions” which we define as policies that maximize the Sender’s payoff under her conjecture among those that provide the optimal payoff guarantee.

The following example (inspired by the “judge example” from [Kamenica and Gentzkow, 2011](#)) illustrates our main ideas.

Example 1. The Receiver is a judge, the Sender is a prosecutor, and there are three relevant states of the world, $\omega \in \{i, m, f\}$, corresponding to a defendant being innocent, guilty of a misdemeanor, or guilty of a felony, respectively. The prior μ_0 is given by $\mu_0(i) = 1/2$ and $\mu_0(m) = \mu_0(f) = 1/4$. The judge, who initially only knows the prior distribution, will convict if her posterior belief that the defendant is guilty (that is, that $\omega \in \{m, f\}$) is at least $2/3$. In that case, she also chooses a sentence. Let $x \in [\underline{x}, \bar{x}]$, with $\underline{x} > 0$, be the range of the number of years in prison that the

judge can select from. The maximal sentence \bar{x} is chosen if the judge’s posterior belief that a felony was committed conditional on the defendant being guilty is at least $1/2$. Otherwise, the sentence is linearly increasing in the conditional probability of the state f . The prosecutor attempts to maximize the expected sentence (with acquitting modeled as a sentence of $x = 0$). Formally, if μ is the induced posterior belief of the judge, with $\mu(\omega)$ denoting the probability of state ω , the Sender’s payoff is given by

$$\widehat{V}(\mu) = \mathbf{1}_{\{\mu(m)+\mu(f)\geq\frac{2}{3}\}} \min\{\bar{x}, \underline{x} + \frac{2\mu(f)}{\mu(f) + \mu(m)}(\bar{x} - \underline{x})\},$$

where $\mathbf{1}_{\{a\}}$ is a function taking value 1 when the statement $\{a\}$ is true and 0 otherwise.

The Bayesian solution, as defined by [Kamenica and Gentzkow \(2011\)](#), is as follows: The prosecutor induces the posterior belief $(\mu(i), \mu(m), \mu(f)) = (1, 0, 0)$ with probability $1/4$ and the belief $(1/3, 1/3, 1/3)$ with probability $3/4$ (by saying “innocent” with probability $1/2$ conditional on the state being i , and “guilty” in all other cases). The expected payoff for the prosecutor is $(3/4)\bar{x}$.

In the above situation, the prosecutor’s conjecture is that she is the sole provider of information. However, this could turn out to be false. For example, after the prosecutor presents her arguments, the judge could call a witness. The prosecutor might not know the likelihood of this scenario, the amount of information that the witness has about the state, or the witness’ motives.¹

When confronted with such uncertainty, it is common to consider the *worst case*: Suppose that the witness knows the true state and strategically reveals information to minimize the sentence. Under this scenario, the prosecutor cannot do better than fully revealing the state. Indeed, if the prosecutor chose a disclosure policy yielding a strictly higher expected payoff, the adversarial witness could respond by fully revealing the state, lowering the prosecutor’s expected payoff back to the full-disclosure payoff of $(1/4)\underline{x} + (3/4)\bar{x}$.

The key observation of our paper is that the prosecutor—even if she is primarily concerned about the worst-case scenario—should *not* fully disclose the state. Consider the following alternative partitional signal: reveal the state “innocent,” and pool together the remaining two states. Suppose that the witness is adversarial. When it is already revealed that the defendant is innocent, the witness has no information

¹The prosecutor may have beliefs over these events, in which case such beliefs are part of what we called the “conjecture.” Our results allow for arbitrary beliefs, not necessarily that the Receiver is uninformed. What is important is that the Sender does not fully trust her beliefs.

left to reveal. In the opposite case, because conditional on the state being m or f the prosecutor’s payoff is concave, the adversarial witness will choose to disclose the state. Thus, in the worst case, the prosecutor’s expected payoff under this policy is the same as under full disclosure. At the same time, the policy is superior if the prosecutor’s conjecture turns out to be true. Indeed, if the prosecutor really is the sole provider of information, then the alternative policy yields $(1/2)\bar{x}$, which is strictly greater than the full-disclosure payoff.

It may be tempting to conclude from similar reasoning that the prosecutor might as well stick to the Bayesian solution, even if she is concerned about the worst-case scenario. After all, if the witness discloses the state in case she is adversarial, then it is irrelevant what signal the prosecutor selects, so shouldn’t she focus on maximizing her payoff under her conjecture? The problem with that argument is that the most adversarial scenario is not always that the witness fully discloses the state. In the Bayesian solution, when the prosecutor induces the posterior $(1/3, 1/3, 1/3)$, the witness may instead reveal the state f with some small probability $\epsilon > 0$. With remaining probability, the judge’s posterior belief that the defendant is guilty will then shift just below the threshold of $2/3$. As a result, the judge acquits the defendant with probability arbitrarily close to one, not just when the latter is innocent but also when they are guilty. Thus, the payoff guarantee for the prosecutor from selecting the Bayesian solution is in fact 0, implying that the Bayesian solution need not be robust to misspecifications in the conjecture.

When the prosecutor is unable to attach probability assessments to all relevant events (like the appearance of a witness), any policy she chooses results in a range of expected payoffs generated by the set of all possible scenarios. Thus, there are many ways in which any two information policies can be compared. Our solution concept is based on two pragmatic premises that are captured by a lexicographic solution. First, and foremost, the Sender would like to secure the best possible payoff guarantee. She does so by dismissing any policy that is not optimal in the “worst-case scenario.” Second, when there are multiple policies that are worst-case optimal, the Sender acts as in the standard Bayesian persuasion model. That is, she selects the policy that, among those that are worst-case optimal, maximizes her expected payoff under the conjecture. We refer to the case described by the conjecture as the *base-case scenario*. The base-case scenario may correspond to the specification that the Sender considers most plausible (for example, after calibrating on some data), focal, or a good

approximation (obtained, for example, by ignoring events that appear unlikely).

The combination of these two properties defines a *robust solution*: a policy that is base-case optimal among those that are worst-case optimal.² The alternative policy described above is in fact a robust solution for the prosecutor. ■

Our baseline model studies a generalization of the above example to arbitrary Sender-Receiver games with finite action and state spaces. To ease the exposition, we initially assume that the base-case scenario is that the Receiver does not have any exogenous information other than that contained in the common prior (the case considered in most of the literature).³ We capture the Sender’s concern about the validity of her conjecture by introducing a third player, Nature, that may send an additional signal to the Receiver. We assume that Nature can condition on the Sender’s signal realization, reflecting the Sender’s uncertainty over the order in which signals are observed.

Worst-case optimal policies maximize the Sender’s expected payoff when Nature’s objective is to minimize the Sender’s payoff. Robust solutions maximize the Sender’s base-case payoff among all worst-case optimal policies.

Despite the fact that robust solutions involve worst-case optimality, they exist under standard conditions, and can be characterized by applying techniques similar to those used to identify Bayesian solutions (e.g., concavification of the value function). However, the economic properties of robust solutions can be quite different from those of Bayesian solutions. Our main result identifies states that cannot appear together in the support of *any* of the posterior beliefs induced by a robust solution. Separation of such states is both necessary and sufficient for worst-case optimality. Robust solutions thus maximize the same objective function as Bayesian solutions but subject to the additional constraint that the induced posteriors have admissible supports.

The separation theorem permits us to qualify in what sense more information is disclosed under robust solutions than under standard Bayesian solutions: For any Bayesian solution, there exists a robust solution that is either Blackwell more informative or not comparable in the Blackwell order. A naive intuition for why robustness

²In Section 6, we show that the lexicographic nature of our solution concept is not essential for its properties: If the Sender instead maximizes a weighted sum of her payoffs in the worst-case and the base-case scenarios, then, under permissive regularity conditions, the solutions coincide with robust solutions as long as the weight on the worst-case scenario is sufficiently large.

³In the single-Receiver case, the scenario in which the Receiver is uninformed happens to be the best possible case for the Sender, that is, the “base-case scenario” is also the “best-case scenario.” Later in the analysis, though, we allow for arbitrary conjectures.

calls for more information disclosure is that, because Nature can always reveal the state, the Sender may opt for revealing the state herself. This intuition, however, is not correct, as we already indicated in the example above. While fully revealing the state is always worst-case optimal, it need not be a robust solution. In fact, if Nature’s most adversarial response to any selection by the Sender is to fully disclose the state, then any signal chosen by the Sender yields the same payoff guarantee and hence is worst-case optimal—the Sender then optimally selects the same signal as in the standard Bayesian persuasion model. Instead, the reason why robustness calls for more information disclosure than standard Bayesian persuasion is that, if certain states are not separated, Nature could push the Sender’s payoff *strictly below* what the Sender would obtain by fully disclosing these states herself. This is the reason why the Sender always reveals the state “innocent” in the robust solution in Example 1, whereas the Bayesian solution sometimes pools that state with the other two.

When the Sender faces non-Bayesian uncertainty, it is natural for her to want to avoid dominated policies. A dominated policy performs weakly (and sometimes strictly) worse than some alternative policy that the Sender could adopt, no matter how Nature responds. We show that at least one robust solution is undominated, and that, provided that the conjecture satisfies a certain condition, all robust solutions are undominated. Thus, robust solutions are desirable even if the Sender attaches no significance to any particular conjecture; they can be used to generate solutions that are worst-case optimal and undominated. Example 1 shows that focusing on worst-case optimal solutions is not enough for this purpose: Full disclosure is worst-case optimal but dominated.

While we focus on a simple model to highlight the main ideas, we argue in Section 4 that our approach and results extend to more general persuasion problems, and can accommodate various assumptions about the Sender’s conjecture and the worst case. With a single Receiver, we can allow the Sender to conjecture that the Receiver observes a particular exogenous signal that is informative about her type or the state; the non-Bayesian uncertainty is created by the possibility that the actual signal observed by the Receiver is different from the one conjectured by the Sender.

Our results also generalize to the case of multiple Receivers under the assumption that the Sender uses a public signal. In the standard persuasion framework, it is typical to assume that the Sender not only controls the information that the Receivers observe but also coordinates their play on the strategy profile most favorable to

her, in case there are multiple profiles consistent with the assumed solution concept and the induced information structure.⁴ In this case, a policy is worst-case optimal if it maximizes the Sender’s payoff under the assumption that Nature responds to the information provided by the Sender by revealing additional information to the Receivers (possibly in a discriminatory fashion) and coordinating their play (in a way consistent with the assumed solution concept) to minimize the Sender’s payoff. In contrast, if the Sender’s conjecture turns out to be correct, the Receivers’ exogenous information and the equilibrium selection are the ones consistent with the Sender’s beliefs. As a result, robust solutions are a flexible tool that can accommodate various assumptions about the environment. For example, a Sender may conjecture that play will constitute a Bayes Nash equilibrium under the information structure induced by her signal. However, she may first impose a “robustness test” to rule out policies that deliver a suboptimal payoff in the worst Bayes correlated equilibrium. For any given specification of the worst-case and base-case Sender’s payoffs, our separation theorem characterizes the resulting robust solutions.

The rest of the paper is organized as follows. We review the related literature next. In Section 2, we present the baseline model, and then we derive the main properties of robust solutions in Section 3. Section 4 extends the model and the results to general persuasion problems, and Section 5 illustrates the results with applications. Finally, in Section 6, we discuss how our solution concept relates to alternative notions of robustness. Most proofs are collected in Appendix A. The Online Appendix contains supplementary results, most notably a discussion of a version of our model in which Nature chooses her signal simultaneously with the Sender, rather than conditioning on the Sender’s signal realization.

Related literature. Our paper contributes to the fast-growing literature on Bayesian persuasion and information design (see, among others, [Bergemann and Morris, 2019](#), and [Kamenica, 2019](#) for surveys). Several recent papers adopt a robust approach to the design of the optimal information structure. [Inostroza and Pavan \(2022\)](#), [Morris et al. \(2020\)](#), [Ziegler \(2020\)](#), and [Li et al. \(2021\)](#) focus on the adversarial selection of the continuation strategy profile of the Receivers. [Babichenko et al. \(2021\)](#) characterize regret-minimizing signals for a Sender who does not know the Re-

⁴Of course, this issue is already present in the single-Receiver case when the Receiver is indifferent between multiple actions; however, with a single Receiver, this is typically a non-generic phenomenon which can be avoided at an arbitrarily low cost for the Sender.

ceiver’s utility function. Most closely related are [Hu and Weng \(2021\)](#) and [Kosterina \(2021\)](#) who study signals that maximize the Sender’s payoff in the worst-case scenario, when the Sender faces uncertainty over the Receivers’ exogenous private information. [Hu and Weng \(2021\)](#) observe that full disclosure maximizes the Sender’s payoff in the worst-case scenario, when the Sender faces full ambiguity over the Receivers’ exogenous information (as in our solution concept). They also consider the opposite case of a Sender that faces small local ambiguity over the Receivers’ exogenous information and show robustness of Bayesian solutions in this case. [Kosterina \(2021\)](#) considers a setting in which the Sender faces ambiguity over the Receiver’s prior. This is similar to the version of our model (analyzed in the Online Appendix) in which the Sender and Nature move simultaneously; however, an important difference is that Nature in [Kosterina’s](#) model chooses the Receiver’s prior, while Nature in our model chooses a distribution of posteriors induced from a fixed (and known) prior.⁵

Our results are different from those in any of the above papers, and reflect a different approach to the design of the optimal signal. Once the Sender identifies all signals that are worst-case optimal, she looks at their performance under the base-case scenario (as in the canonical Bayesian persuasion model). In particular, our solution concept reflects the idea that there is no reason for the Sender to fully disclose the state if she can benefit by withholding some information under the conjectured scenario while still guaranteeing the same worst-case payoff. Our lexicographic approach to the assessment of different information structures is in the same spirit as the one proposed by [Börgers \(2017\)](#) in the context of robust mechanism design.

The literature on Bayesian persuasion with multiple Senders is also related, in that Nature is effectively a second Sender in the persuasion game that we study. [Gentzkow and Kamenica \(2016, 2017\)](#) consider persuasion games in which multiple Senders move simultaneously and identify conditions under which competition leads to more information being disclosed in equilibrium. [Board and Lu \(2018\)](#) consider a search model and provide conditions for the existence of a fully-revealing equilibrium. [Au and Kawai \(2020\)](#) study multi-Sender simultaneous-move games where each Sender discloses information about the quality of her product (with the qualities

⁵The above papers consider robustness for the Sender. [Nikzad \(2021\)](#) studies a model with a non-Bayesian Receiver who takes an action that guarantees the highest possible payoff guarantee. Beyond information design, other papers look at the consequences of the designer’s ambiguity over the agents’ information sources; for example, [Carroll \(2019\)](#) and [Brooks and Du \(2021\)](#) consider informationally-robust design of trading mechanisms.

drawn independently across Senders). They show that, as the number of Senders increases, each Sender discloses more information, with the information disclosed by each Sender converging to full disclosure as the number of Senders goes to infinity. [Cui and Ravindran \(2020\)](#) consider persuasion by competing Senders in zero-sum games and identify conditions under which full disclosure is the unique outcome.⁶ [Li and Norman \(2021\)](#), and [Wu \(2021\)](#), instead, analyze games in which Senders move sequentially and, among other things, identify conditions under which (1) full information revelation can be supported in equilibrium and (2) silent equilibria (that is, equilibria in which all Senders but one remain silent) sustain all equilibrium outcomes. These papers focus on equilibrium outcomes under competition, rather than robustness of the policy chosen by a single Sender. A key element in [Li and Norman \(2021\)](#)’s equilibrium analysis is the optimality for each Sender of inducing “stable beliefs” that are not further split by downstream Senders. Our paper shows that imposing a zero-sum payoff assumption generates a sharp implication on the structure of stable beliefs in terms of states that are separated under any of the induced posteriors.

[Kolotilin et al. \(2017\)](#), [Laclau and Renou \(2017\)](#), and [Guo and Shmaya \(2019\)](#), among others, consider persuasion of privately informed Receivers. Additionally, [Matysková and Montes \(2021\)](#) and [Ye \(2021\)](#) study models in which the Receiver optimally acquires information in response to the signal selected by the Sender. Contrary to the present paper, in that literature, the distribution of the Receivers’ private information is known to the Sender.

2 Model

A payoff-relevant state ω is drawn from a finite set Ω according to a distribution $\mu_0 \in \Delta\Omega$ that is common knowledge between a Sender and a Receiver. The Receiver has a continuous utility function $u(a, \omega)$ that depends on her action a , chosen from a compact set A , and the state ω . Let $A^*(\mu) := \operatorname{argmax}_{a \in A} \sum_{\omega \in \Omega} u(a, \omega)\mu(\omega)$ denote the set of actions that maximize the Receiver’s expected payoff when her posterior belief over the state is $\mu \in \Delta\Omega$. The Sender has a continuous utility function $v(a, \omega)$. She chooses an information structure $q : \Omega \rightarrow \Delta\mathcal{S}$ that maps states into probability

⁶In the Stackelberg version of the zero-sum game between competing Senders, [Cui and Ravindran \(2020\)](#) assume that the follower cannot condition its information on the realization of the leader’s signal. This scenario corresponds to the version of our model analyzed in the Online Appendix.

distributions of signal realizations in some finite signal space \mathcal{S} : We denote by $q(s|\omega)$ the probability of signal realization $s \in \mathcal{S}$ in state ω . Hereafter, we abuse terminology and refer to q as the Sender’s signal.

The Sender faces uncertainty about the exogenous sources of information the Receiver may have access to, when learning about the state. We capture this uncertainty by allowing Nature to disclose additional information to the Receiver that can be correlated with both the state and the realization of the Sender’s signal. That is, in the eyes of the Sender, Nature chooses an information structure $\pi : \Omega \times \mathcal{S} \rightarrow \Delta\mathcal{R}$ that maps $(\omega, s) \in \Omega \times \mathcal{S}$ into a distribution over a set of signal realizations in some finite signal space \mathcal{R} . We denote by $\pi(r|\omega, s)$ the probability of signal realization $r \in \mathcal{R}$ when the state is ω and the realization of the Sender’s signal is s . The possibility for Nature to condition her signal on the realization of the Sender’s signal reflects the Sender’s concern that the Receiver may be able to acquire additional information after seeing the realization of her signal.

Hereafter, we treat the signal spaces \mathcal{S} and \mathcal{R} as exogenous and assume that they are subsets of some sufficiently rich space. Because Ω is finite, it will become clear that, under our solution concept, the assumption of finite \mathcal{S} and \mathcal{R} is without loss of optimality for either the Sender or Nature. We denote by Q and Π the set of all feasible signals for the Sender and Nature, respectively. Fixing some set of signals, for any initial belief $\mu \in \Delta\Omega$, we denote by $\mu^x \in \Delta\Omega$ the posterior belief induced by the realization x of these signals, where x could be a vector. In particular, we denote by $\mu_0^{s,r} \in \Delta\Omega$ the posterior belief over Ω that is obtained starting from the prior belief μ_0 and conditioning on the realization (s, r) of the signals q and π .

In the standard Bayesian persuasion model, the Sender has a belief about the Receiver’s exogenous information and the way the Receiver plays in case of indifference. We refer to this belief as the Sender’s *conjecture*. We denote by $\widehat{V}(\mu)$ the Sender’s expected payoff when her induced posterior belief μ is paired with Nature’s disclosure and the Receiver adopts the conjectured tie-breaking rule. To simplify the exposition, we assume in this section that the Sender’s conjecture is that the Receiver has access to no information other than the one contained in the prior and, in case of indifference, chooses the action most favorable to the Sender, as in the baseline model of [Kamenica and Gentzkow \(2011\)](#). (This assumption is relaxed in Section 4, where we show that all our results extend to general conjectures.) Under this simplifying

assumption, we have that

$$\widehat{V}(\mu) := \max_{a \in A^*(\mu)} \sum_{\omega \in \Omega} v(a, \omega) \mu(\omega).$$

The Bayesian persuasion problem is to maximize

$$\widehat{v}(q) := \sum_{\omega \in \Omega, s \in \mathcal{S}} \widehat{V}(\mu_0^s) q(s|\omega) \mu_0(\omega)$$

over all signals $q \in Q$. We refer to the function $\widehat{v}(\cdot)$ as the *base-case payoff*.

In contrast, if the Sender is concerned about the robustness of her information policy, she may evaluate her expected payoff from choosing q as

$$\psi(q) := \inf_{\pi \in \Pi} \left\{ \sum_{\omega \in \Omega, s \in \mathcal{S}} \left(\sum_{r \in \mathcal{R}} \underline{V}(\mu_0^{s,r}) \pi(r|\omega, s) \right) q(s|\omega) \mu_0(\omega) \right\},$$

where

$$\underline{V}(\mu) := \min_{a \in A^*(\mu)} \sum_{\omega \in \Omega} v(a, \omega) \mu(\omega).$$

We refer to $\psi(\cdot)$ as the *worst-case payoff*. The “worst case” refers to the scenario in which Nature responds to the Sender’s choice of signal q by selecting a signal that minimizes the Sender’s payoff, as reflected by the infimum over all signals $\pi \in \Pi$. Moreover, in case the Receiver is indifferent between several actions, Nature breaks the ties against the Sender, as reflected by the definition of \underline{V} .

3 Robust solutions

We now define robust solutions and derive their properties.

Definition 1. A signal $q \in Q$ is *worst-case optimal* if it maximizes the worst-case payoff ψ over the set of all signals Q .

We let $W = \operatorname{argmax}_{q \in Q} \psi(q)$ denote the set of worst-case optimal signals (as we show below, the set is non-empty).

Definition 2. A signal $q \in Q$ is a *robust solution* if it maximizes the base-case payoff \widehat{v} over the set of all worst-case optimal signals W .

As foreshadowed in the Introduction, the definition of a robust solution reflects the Sender’s lexicographic attitude towards the uncertainty she faces. First, the Sender seeks a signal that is worst-case optimal, i.e., that is not outperformed by any other

signal, in case Nature plays adversarially. Second, if there are multiple signals that pass this test, the Sender seeks one among them that maximizes her payoff in case her conjecture is correct. In short, a robust solution is a signal that is base-case optimal among those that are worst-case optimal.

Because the Sender's payoff depends only on the induced posterior belief, it is natural to optimize directly over distributions of posterior beliefs (rather than signals). For any $\mu \in \Delta\Omega$, let

$$\underline{V}(\mu) := \inf_{\pi: \Omega \rightarrow \Delta\mathcal{R}} \left\{ \sum_{\omega \in \Omega, r \in \mathcal{R}} \underline{V}(\mu^r) \pi(r|\omega) \mu(\omega) \right\}$$

denote the expected payoff to the Sender conditional on inducing a posterior belief μ under the worst-case scenario, that is, when Nature responds to the induced belief μ by minimizing the Sender's payoff with the choice of π (and the Receiver breaks ties adversarially). Note that π no longer depends on the realization of the Sender's signal because the function $\underline{V}(\mu)$ is defined at the interim stage, conditional on the Sender inducing some belief μ with her signal realization.

Next, for any $\omega \in \Omega$, let δ_ω denote the Dirac distribution assigning measure one to the state ω and, for any induced posterior $\mu \in \Delta\Omega$, denote by

$$\underline{V}_{\text{full}}(\mu) := \sum_{\omega \in \Omega} \underline{V}(\delta_\omega) \mu(\omega)$$

the Sender's payoff when, starting from μ , the state is fully disclosed to the Receiver (hereafter, the "full-disclosure" payoff).

Lemma 1. *Let $\mathcal{W} \subset \Delta\Delta\Omega$ denote the set of all distributions ρ over posterior beliefs that satisfy*

$$\int \underline{V}(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0), \tag{WC}$$

and Bayes plausibility

$$\int \mu d\rho(\mu) = \mu_0. \tag{BP}$$

A signal $q \in Q$ is a robust solution if and only if the distribution over posterior beliefs $\rho_q \in \Delta\Delta\Omega$ that q induces maximizes $\int \widehat{V}(\mu) d\rho(\mu)$ over \mathcal{W} .

Lemma 1 is intuitive. Since Nature can always disclose the state, the Sender's payoff in the worst-case scenario is upper bounded by the full-disclosure payoff. Clearly, this upper bound can be achieved by the Sender disclosing the state herself. Hence, a signal q is worst-case optimal (i.e., $q \in W$) if and only if $\underline{v}(q) = \underline{V}_{\text{full}}(\mu_0)$. In

particular, full disclosure belongs to W . The lemma then expresses this observation in terms of distributions of posterior beliefs, using the property that a distribution can be induced by some signal if and only if it is *Bayes-plausible*, that is, it satisfies (BP). Hereafter, we abuse terminology and call ρ_{RS} a robust solution if it maximizes $\int \widehat{V}(\mu) d\rho(\mu)$ over all Bayes-plausible distributions $\rho \in \Delta\Delta\Omega$ satisfying (WC), with no further reference to the underlying signal q .

It is useful at this point to contrast a robust solution with a Bayesian-persuasion solution (henceforth, *Bayesian solution*; see [Kamenica and Gentzkow, 2011](#)).

Definition 3. A signal q_{BP} is a *Bayesian solution* if it maximizes the base-case payoff \widehat{v} over the set of all signals Q . This is the case if and only if the distribution $\rho_{BP} \in \Delta\Delta\Omega$ over posterior beliefs induced by q_{BP} maximizes $\int \widehat{V}(\mu) d\rho(\mu)$ over all ρ satisfying (BP).

By Lemma 1, the only difference between a Bayesian solution and a robust solution is that a robust solution must additionally satisfy constraint (WC).

While the result in Lemma 1 offers a useful perspective, to compute $\underline{V}(\mu)$, one must solve a Bayesian persuasion problem with μ as a prior. This problem consists in choosing a distribution over posterior beliefs averaging out to μ to minimize the Sender's expected payoff, with the latter given by $\underline{V}(\eta)$ for any posterior $\eta \in \Delta\Omega$ induced by Nature. Let $\text{lco}(\underline{V})$ denote the *lower convex closure* of \underline{V} , that is, $\text{lco}(\underline{V}) = -\text{co}(-\underline{V})$, where the concave closure $\text{co}(\cdot)$ of a function is defined as the lowest concave upper bound, as in [Kamenica and Gentzkow \(2011\)](#). It follows that $\underline{V}(\mu) = \text{lco}(\underline{V})(\mu)$. Note that $\underline{V}(\mu)$ is a convex function that coincides with $\underline{V}(\mu)$ at Dirac deltas $\mu = \delta_\omega$. Computing $\text{lco}(\underline{V})(\mu)$ can be difficult, especially when the state space is large. The next result shows an alternative characterization of \mathcal{W} .

For any function $V : \Delta\Omega \rightarrow \mathbb{R}$, and $Y \subseteq \Delta\Omega$, let $V|_Y$ denote a function defined on the domain Y that coincides with V on Y . Given any $\mu \in \Delta\Omega$, let $\text{supp}(\mu)$ denote the support of μ , that is, the smallest subset of Ω with measure one under μ .

Proposition 1. *Let*

$$\mathcal{F} := \{B \subseteq \Omega : \underline{V}|_{\Delta B} \geq \underline{V}_{full}|_{\Delta B}\}.$$

Then,

$$\mathcal{W} = \{\rho \in \Delta\Delta\Omega : \rho \text{ satisfies (BP) and, } \forall \mu \in \text{supp}(\rho), \text{supp}(\mu) \in \mathcal{F}\}.$$

Proposition 1 states that a Bayes-plausible distribution $\rho \in \Delta\Delta\Omega$ is worst-case optimal if and only if the support of any of the posteriors induced by ρ is “admissible.” Moreover, the proposition describes exactly what the admissible sets are: The set $B \subseteq \Omega$ is admissible if (and only if) *any* posterior supported on B gives the Sender an expected payoff no smaller than the one the Sender could obtain, starting from μ , by fully disclosing the state. Informally, B is admissible if the Sender prefers obfuscation to transparency on that set of states. Importantly, this condition is expressed effectively in terms of the primitives of the model (apart from solving for the Receiver’s best-response correspondence), and checking it does not require computing the lower convex closure of \underline{V} .

To gain intuition, fix a posterior belief $\mu \in \Delta\Omega$ in the support of the belief distribution chosen by the Sender. Then, for any belief $\eta \in \Delta\Omega$ with $\text{supp}(\eta) \subseteq \text{supp}(\mu)$, starting from μ , Nature can induce the belief η with positive probability, and disclose the state with remaining probability. Thus, if there exists an η such that $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$, then by inducing μ , the Sender exposes herself to a payoff strictly below what she would obtain by revealing the state. The only way for the Sender to avoid that exposure is to separate some states in the support of μ so that Nature can no longer induce η . Conversely, if no such η exists for which $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$, then, conditional on μ , Nature minimizes the Sender’s payoff by fully disclosing the states in the support of μ . Because the Sender’s payoff under the worst-case scenario is upper bounded by the payoff she obtains under full disclosure (by Lemma 1), any such μ can be part of a worst-case optimal distribution.

This logic is illustrated in Figure 3.1 where, for simplicity, we consider a binary state space. The solid line depicts the payoff function \underline{V} , while the dashed line (connecting the payoffs from inducing degenerate beliefs) represents the full-disclosure payoff $\underline{V}_{\text{full}}$. Because there exists a belief η at which \underline{V} lies strictly below $\underline{V}_{\text{full}}$, worst-case optimality rules out inducing *any* posterior belief whose support contains the support of η (in the figure, this means any belief in the interior of the simplex). To see this, suppose, for example, that the Sender induces the posterior μ (see Figure 3.1) that is part of the Bayesian solution. Then, conditional on μ realizing, Nature can split μ into η and a set of Dirac deltas. Because $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$, and because \underline{V} and $\underline{V}_{\text{full}}$ always coincide at degenerate beliefs, the resulting expected payoff for the Sender is strictly below her full-disclosure payoff. Hence, μ cannot be induced under any policy generating the optimal payoff guarantee. A similar argument applies to

any other posterior belief whose support contains the support of η .

The example also illustrates the convenience of Proposition 1. To identify worst-case optimal policies, one does not actually need to compute $\text{lco}(\underline{V})$, which can be difficult with more than two states. The existence of a belief η for which $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$ suffices to rule out all distributions $\rho \in \Delta\Delta\Omega$ that generate beliefs whose supports contain the support of η .

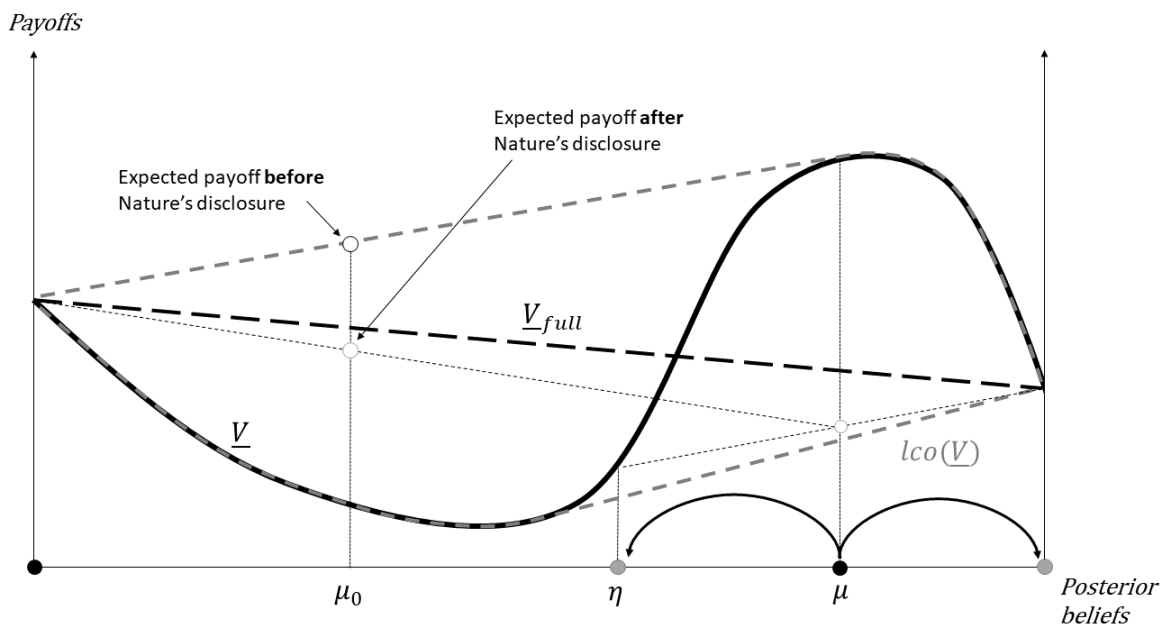


Figure 3.1: Illustration of Proposition 1

The following theorem, which is our main characterization result, then follows directly from what we established above.

Theorem 1 (Separation Theorem). $\rho_{RS} \in \Delta\Delta\Omega$ is a robust solution if and only if it maximizes

$$\int \widehat{V}(\mu) d\rho(\mu)$$

over all distributions $\rho \in \Delta\Delta\Omega$ satisfying (BP) and such that

$$\text{supp}(\rho) \subseteq \Delta_{\mathcal{F}}\Omega := \{\mu \in \Delta\Omega : \text{supp}(\mu) \in \mathcal{F}\}.$$

Theorem 1 implies that the only difference between a Bayesian solution and a robust solution is that the latter must satisfy an additional constraint on the supports of the posterior beliefs it induces: A robust solution can only attach positive

probability to posterior beliefs supported on “admissible” subsets of the state space, as described by the collection \mathcal{F} from Proposition 1.

For illustration, consider Example 1 from the Introduction. First, note that a robust solution cannot induce a posterior belief that mixes the state i with some other state from $\{m, f\}$; indeed, by Theorem 1, such supports are not admissible because any posterior belief that puts sufficiently high (but not full) probability on the state i yields a zero payoff for the Sender, which is strictly worse than the full-disclosure payoff. Second, states m and f need not be separated because, when restricted to beliefs supported on $\{m, f\}$, the Sender’s payoff \underline{V} is concave, and hence lies above the linear function $\underline{V}_{\text{full}}$ with which it coincides at the end points (degenerate beliefs). Thus, we have that $\mathcal{F} = \{\{i\}, \{m\}, \{f\}, \{m, f\}\}$. Theorem 1 then predicts that a robust solution must reveal the state i , and that it maximizes the Sender’s expected payoff \widehat{V} conditional on states $\{m, f\}$. Because \widehat{V} is concave on $\Delta\{m, f\}$, it is optimal not to reveal any information conditional on these states. This confirms our assertion that revealing i and pooling m and f is a robust solution for Example 1.

Theorem 1 yields a number of direct corollaries that we describe next.

Corollary 1 (Existence). *A robust solution exists.*

Indeed, the set \mathcal{W} of worst-case optimal distributions is closed, and thus compact (this follows because the collection \mathcal{F} is closed with respect to taking subsets, i.e., if $B \in \mathcal{F}$, then all subsets of B also belong to \mathcal{F}). It is non-empty because it contains a distribution corresponding to full disclosure of the state. Finally, the function \widehat{V} is upper semi-continuous, so existence follows from Weierstrass Theorem.

It is well-known that requiring exact worst-case optimality often precludes existence of solutions in related models. Indeed, we show in the Online Appendix that existence may fail when Nature selects a signal simultaneously with the Sender. However, when Nature can condition on the realization of the Sender’s signal, existence is guaranteed by the fact that Nature’s optimal response to each signal realization convexifies the Sender’s value function, hence making it continuous.

Hereafter, we will say that states ω and ω' are *separated* by a distribution $\rho \in \Delta\Delta\Omega$ if there is no posterior $\mu \in \text{supp}(\rho)$ such that $\{\omega, \omega'\} \subseteq \text{supp}(\mu)$. Intuitively, given any posterior belief μ induced by ρ , the Receiver never faces any uncertainty between ω and ω' .

Corollary 2 (State separation). *Suppose that there exists $\lambda \in (0, 1)$ and $\omega, \omega' \in \Omega$ such that $\underline{V}(\lambda\delta_\omega + (1 - \lambda)\delta_{\omega'}) < \lambda\underline{V}(\delta_\omega) + (1 - \lambda)\underline{V}(\delta_{\omega'})$. Then any robust solution must separate the states ω and ω' .*

Under the assumptions of Corollary 2, \mathcal{F} does not contain the set $\{\omega, \omega'\}$. Thus, by Theorem 1, a worst-case optimal distribution cannot induce posterior beliefs that have both of these states in their support. Note that the assumption is that there exists *some* belief supported on $\{\omega, \omega'\}$ under which full disclosure is strictly better for the Sender, while the conclusion says that a robust solution cannot induce *any* posterior belief that puts strictly positive mass on both ω and ω' .

In the special case when there are two states, Corollary 2 exhausts all possibilities.

Corollary 3 (Characterization for binary-state case). *Suppose that $\Omega = \{\omega_L, \omega_H\}$, and $\underline{V}(p)$ is the Sender's payoff when the posterior probability of state ω_H is p . Then,*

- *if for some p , $\underline{V}(p) < (1 - p)\underline{V}(0) + p\underline{V}(1)$, then full disclosure is the unique robust solution;*
- *otherwise, the set of robust solutions coincides with the set of Bayesian solutions.*

For a quick application of Corollary 3, consider the original judge example of [Kamenica and Gentzkow \(2011\)](#): For low posterior probabilities $p > 0$ of the defendant being guilty, the prosecutor's payoff is zero, while the prosecutor's expected payoff would be strictly positive under full disclosure at p . Thus, full disclosure is the unique robust solution for the prosecutor in the original judge example of [Kamenica and Gentzkow \(2011\)](#).

Beyond the binary-state case, by Corollary 2, full disclosure is the unique robust solution in any problem for which the separation condition holds for *any* pair of states. Similarly, we can extend the conditions under which robust solutions coincide with Bayesian solutions.

Corollary 4 (Robust and Bayesian solutions coincide). *All Bayes-plausible distributions are worst-case optimal if, and only if, $\Omega \in \mathcal{F}$; then, the set of robust solutions coincides with the set of Bayesian solutions.*

For an illustration of Corollary 4, consider the baseline model of [Bergemann et al. \(2015\)](#): A monopolistic seller quotes a price to a buyer who is privately informed about her value ω for the seller's good. A Sender reveals information to the seller

(who acts as a Receiver) about ω . When the Sender maximizes the buyer’s surplus, Corollary 4 applies. Because the buyer’s surplus is 0 at all degenerate beliefs, we have that $\underline{V}_{\text{full}}(\mu) = 0$ and $\underline{V}(\mu) \geq 0$ for all μ . Thus, $\Omega \in \mathcal{F}$, and the optimal signal identified by Bergemann et al. (2015)—although quite complicated—is in fact robust. If the Sender instead maximizes the seller’s profit, then Corollary 2 applies to any pair of states: If ω and ω' are not separated by the Sender, Nature can ensure that the seller does not extract all the surplus. Thus, in this case, \mathcal{F} only contains singletons, and full disclosure is the unique robust solution.

In all the examples discussed thus far, a robust solution discloses weakly more information than a Bayesian solution. To see whether this property holds generally, we use Blackwell dominance to formalize the idea that one distribution of posteriors is more informative than another one.⁷ We start with a useful observation.

Corollary 5 (Worst-case optimality preserved under more information disclosure). *\mathcal{W} is closed under Blackwell dominance: If $\rho' \in \mathcal{W}$, and ρ Blackwell dominates ρ' , then $\rho \in \mathcal{W}$.*

The conclusion follows directly from Theorem 1 by noting that if $B \in \mathcal{F}$, then any subset of B must also be in \mathcal{F} . An increase in the Blackwell order on $\Delta\Delta\Omega$ can only make the supports of posterior beliefs smaller, so such an increase cannot take a distribution out of the set \mathcal{W} .

Suppose that there exists a Bayesian solution that Blackwell dominates a robust solution. Then, by Corollary 5, that Bayesian solution must be worst-case optimal, and hence it is also a robust solution. Therefore, we obtain the following conclusion:

Corollary 6 (Comparison of informativeness). *Take any Bayesian solution ρ_{BP} . Then, there exists a robust solution ρ_{RS} such that either ρ_{RS} and ρ_{BP} are not comparable in the Blackwell order, or ρ_{RS} Blackwell dominates ρ_{BP} .*

Corollary 6 provides a formal sense in which (maximally informative) robust solutions provide (weakly) more information than Bayesian solutions.⁸ This is a relatively weak notion—it is certainly possible that the two solutions are not comparable in the

⁷Formally, we say that $\rho \in \Delta\Delta\Omega$ Blackwell dominates $\rho' \in \Delta\Delta\Omega$ if, for all convex functions $V : \Delta\Omega \rightarrow \mathbb{R}$, $\int V(\mu)d\rho(\mu) \geq \int V(\mu)d\rho'(\mu)$. Equivalently, ρ is a mean-preserving spread of ρ' .

⁸By a “maximally informative” solution we mean a solution that is not Blackwell dominated by any other robust solution. Note that without that qualifier the statement is obviously false. For example, when both \underline{V} and \widehat{V} are affine, all distributions are both robust and Bayesian solutions and hence there exist Bayesian solutions that strictly Blackwell dominate some robust solutions.

Blackwell order. However, it can never happen that a Bayesian solution strictly Blackwell dominates a maximally informative robust solution. Whenever a robust solution ρ_{RS} is strictly more informative than the Bayesian solution ρ_{BP} , Theorem 1 implies that ρ_{RS} separates states that are not separated under ρ_{BP} .

While the result in Corollary 6 is intuitive, we emphasize that it is not trivial. Because Nature can only provide additional information, one may expect more information to be disclosed overall under robust solutions than under Bayesian solutions. However, Corollary 6 says that the Sender *herself* will provide more (or at least not less) information than she would in the Bayesian-persuasion model. Moreover, we show in the Online Appendix that the conclusion of Corollary 6 actually fails in the version of the model in which Nature chooses a signal simultaneously with the Sender.⁹

Finally, we show that robust solutions can be found using the concavification technique (see Aumann and Maschler, 1995, and Kamenica and Gentzkow, 2011). Indeed, because the state-separation condition applies posterior by posterior, we can incorporate the constraints into the objective function \widehat{V} by modifying its value on $\Delta_{\mathcal{F}}^c\Omega := \Delta\Omega \setminus \Delta_{\mathcal{F}}\Omega$ (that is, on the set of posteriors not supported in \mathcal{F}) to be a sufficiently low number. Formally, let $v_{\text{low}} := \min_{\omega \in \Omega} \widehat{V}(\delta_{\omega}) - 1$, and define¹⁰

$$\widehat{V}_{\mathcal{F}}(\mu) := \begin{cases} \widehat{V}(\mu) & \text{if } \mu \in \Delta_{\mathcal{F}}\Omega \text{ and } \widehat{V}(\mu) \geq v_{\text{low}}, \\ v_{\text{low}} & \text{otherwise.} \end{cases} \quad (3.1)$$

Observe that posteriors μ with $\widehat{V}(\mu) \leq v_{\text{low}}$ are never induced in either a robust or a Bayesian solution because a strictly higher expected value for the Sender could be obtained by decomposing such μ into Dirac deltas, by the definition of v_{low} . Therefore, Bayesian solutions under the objective function $\widehat{V}_{\mathcal{F}}$ correspond exactly to robust solution with the original objective, by Theorem 1. Moreover, we have defined the modification $\widehat{V}_{\mathcal{F}}$ of \widehat{V} so that it remains upper-semi-continuous because the set $\{\mu \in \Delta\Omega : \text{supp}(\mu) \in \mathcal{F} \text{ and } \widehat{V}(\mu) \geq v_{\text{low}}\}$ is closed.

⁹Corollary 6 relates to the comparative statics in Li and Norman (2021). They show that, in general, adding a new Sender who speaks after the existing Senders need not increase the amount of information that is passed on to the Receiver; our result shows that adding a new Sender (Nature) to the end of the line never induces the first Sender to reduce the informativeness of her signal as long as the preferences of the new Sender are opposite to hers.

¹⁰Note that, in general, $\widehat{V}(\mu) \geq v_{\text{low}}$ is not implied by $\mu \in \Delta_{\mathcal{F}}\Omega$. The latter condition guarantees that $\underline{V}(\mu) = \underline{V}_{\text{full}}(\mu)$. However, it is possible that $\underline{V}_{\text{full}}(\mu) < v_{\text{low}}$. This is because the Receiver breaks ties adversarially under $\underline{V}_{\text{full}}(\mu)$ but favorably under $\widehat{V}(\mu)$.

Corollary 7 (Concavification). *A Bayes-plausible distribution $\rho \in \Delta\Delta\Omega$ is a robust solution if and only if $\int \widehat{V}_{\mathcal{F}}(\mu)d\rho(\mu) = \text{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$.*

Corollary 7 implies that the problem of finding a robust solution can be reduced to finding a Bayesian solution with a modified objective function. As a result, robust solutions inherit many of the properties of Bayesian solutions. For example, [Kamenica and Gentzkow \(2011\)](#) show that there always exists a Bayesian solution that uses at most as many signal realizations as there are states, implying in particular that the restriction to finite signal spaces is without loss of optimality for the Sender.

Corollary 8 (Support). *There always exists a robust solution ρ with $|\text{supp}(\rho)| \leq |\Omega|$.*

4 Extensions

By direct inspection of the proofs, all the results of the previous section rely only on the following properties of the reduced-form payoffs:

- $\underline{V} : \Delta\Omega \rightarrow \mathbb{R}$ is lower semi-continuous;
- $\underline{V} : \Delta\Omega \rightarrow \mathbb{R}$ is the lower convex closure of \underline{V} .¹¹
- $\widehat{V} : \Delta\Omega \rightarrow \mathbb{R}$ is upper semi-continuous.

By Lemma 1, robust solutions can be defined in terms of these reduced-form payoff functions. Because the specific micro-foundation for these payoffs plays no role, the conclusions established in the previous section extend to any primitive environment that generates reduced-form payoffs satisfying the above properties.

4.1 General conjectures in the single-Receiver model

In the baseline model, the Sender conjectures that the Receiver does not have any information other than that contained in the common prior. Moreover, she conjectures that, in case of indifference, the Receiver will resolve the indifference in her favor. Suppose, instead, that the Sender conjectures that Nature will respond to her disclosure with some signal $\pi_0 : \Omega \times \Delta\Omega \rightarrow \Delta\mathcal{R}$. That is, when the Sender's signal realization induces a posterior belief μ , the Sender conjectures that the Receiver will

¹¹Because \underline{V} is the lower convex closure of a lower semi-continuous function, it is also continuous; this follows from the so-called GKR Theorem (see [Gale et al., 1968](#)), which states that any convex function on a closed convex bounded polytope is upper semi-continuous.

observe an additional signal realization r drawn from \mathcal{R} with probability $\pi_0(r|\omega, \mu)$. The dependence of $\pi_0(\cdot|\omega, \mu)$ on μ captures the possibility that the additional information collected by the Receiver may depend on the Sender’s signal realization.¹² Moreover, the Sender conjectures that the Receiver will use a (potentially) stochastic and belief-dependent tie-breaking rule $\xi_0 : \Delta\Omega \rightarrow \Delta A$, where $\xi_0(\cdot|\mu')$ is a probability distribution over the Receiver’s actions when the final posterior belief is μ' , with the property that $\xi_0(A^*(\mu')|\mu') = 1$, for any $\mu' \in \Delta\Omega$. The Sender’s expected payoff from inducing the posterior μ under her conjecture is then equal to

$$\widehat{V}(\mu) = \sum_{\omega \in \Omega, r \in \mathcal{R}} \left(\int_A v(a, \omega) d\xi_0(a|\mu^r) \right) \pi_0(r|\omega, \mu) \mu(\omega). \quad (4.1)$$

Provided that \widehat{V} is upper semi-continuous, all the results from Section 3 continue to hold. A special case is when the Sender conjectures that the Receiver will play favorably to her when indifferent, and that the additional information the Receiver has access to is invariant to the realization of the Sender’s signal. This is true, for example, when the Receiver observes the realization of such an additional signal *before* observing the realization of the Sender’s signal. Such conjectures are captured by $\pi_0(r|\omega, \mu)$ that do not depend on μ . In Section 5, we characterize robust solutions in an example from Guo and Shmaya (2019) featuring a privately-informed Receiver where the Sender’s conjecture has these precise properties.

4.2 Multiple Receivers

In the baseline model, the Sender faces a single Receiver. Our approach extends to the case of multiple Receivers under the assumption that the Sender is restricted to *public signals*. Under such an assumption, many persuasion problems can be characterized in terms of reduced-form payoffs satisfying the properties discussed above.

With multiple Receivers, however, robustness to strategy selection (corresponding to tie-breaking in the single-Receiver case) can be just as important as robustness to additional information. In the Bayesian-persuasion literature, it is customary to assume that the Sender is able to coordinate the Receivers on the strategy profile most favorable to her, among those consistent with the assumed solution concept.¹³ Under

¹²The above formulation implicitly assumes that such additional information does not depend on the specific signal q used by the Sender to generate the posterior μ . This assumption permits us to formulate the Sender’s problem as choosing a distribution of posterior beliefs rather than a signal.

¹³Notable exceptions include Inostroza and Pavan (2022), Mathevet et al. (2020), Morris et al. (2020), Ziegler (2020), and Li et al. (2021)

robust design, instead, the Sender may not trust that the Receivers will play favorably to her. Instead, she may seek a signal that yields the maximal payoff guarantee when Nature provides additional information to the Receivers *and* coordinates them on the strategy profile most adversarial to her (among those consistent with the assumed solution concept).

The case of public disclosures by Nature. Consider first the case in which Nature is expected to disclose the same information to all the Receivers. The Receivers are assumed to share a common prior μ_0 . Given the common posterior μ_0^s induced by the Sender's signal realization s , Nature reveals an additional public signal r to the Receivers drawn from a distribution $\pi(\cdot|\omega, \mu_0^s) \in \Delta\mathcal{R}$. Given the final (common) posterior $\mu_0^{s,r}$ induced by the combination of the realizations of the Sender's and Nature's signals, the Receivers play some Bayesian game. For any common posterior $\mu \in \Delta\Omega$, denote by $EQ^*(\mu)$ the set of strategy profiles that are consistent with the assumed solution concept and the common posterior μ . Finally, let $\xi(\cdot|\mu) \in \Delta EQ^*(\mu)$ denote a (possibly stochastic) rule describing the selection of a strategy profile from $EQ^*(\mu)$.

In this setting, $\underline{V}(\mu)$ represents the Sender's expected payoff when, given the common posterior μ , Nature induces the Receivers to play according to the selection $\xi(\cdot|\mu) \in \Delta EQ^*(\mu)$ that is least favorable to the Sender. Under regularity conditions, the function \underline{V} is lower semi-continuous. The function \underline{V} is then the Sender's expected payoff when, in addition to coordinating the Receivers to play adversarially, Nature also discloses additional (public) information to the Receivers so as to minimize the Sender's expected payoff. As in the baseline model, we then have $\underline{V} = \text{lco}(\underline{V})$.

The Sender's conjecture is that the Receivers observe exogenous public signals with distribution $\pi_0(\cdot|\omega, \mu)$, and that, for any final common posterior μ' , they play according to a selection rule $\xi_0(\cdot|\mu') \in \Delta EQ^*(\mu')$. The combination of π_0 and ξ_0 defines the Sender's conjecture. Given such a conjecture, the Sender's expected payoff from inducing the common posterior μ is equal to $\widehat{V}(\mu)$. Provided that this function is upper semi-continuous, all the results from the previous section continue to hold.

The case of private disclosures by Nature. Our approach can also accommodate discriminatory disclosures by Nature, whereby Nature sends different signals to different Receivers. This case can be relevant for settings in which the Sender is restricted to public disclosures (e.g., because of regulatory constraints) but

is nevertheless concerned about the possibility that the Receivers may be endowed with private signals and/or be able to acquire additional information in a decentralized fashion, after hearing the Sender’s public announcement.

With private signals, the distinction between strategy selection and additional information provided by Nature becomes blurred. This is best illustrated by the solution concept of Bayes Correlated Equilibrium (BCE) in which private recommendations that are potentially informative about the state are part of the solution concept (see [Bergemann and Morris, 2016](#)). If the worst-case scenario originates in Nature coordinating the Receivers on the BCE that minimizes the Sender’s expected payoff among all BCE consistent with the common posterior that she induces, then specifying the information provided by Nature becomes redundant. Thus, it is no longer helpful to derive the worst-case payoff for the Sender in two steps, by first looking at strategy profiles for given information, and then at different disclosures by Nature.

In such cases, we can bypass the function \underline{V} by assuming that $\underline{V} \equiv \underline{V}$. The function $\underline{V}(\mu)$ is then interpreted as the Sender’s payoff from inducing the common posterior belief μ when Nature responds by disclosing (possibly private) signals to the Receivers and coordinating them on a strategy profile that minimizes the Sender’s expected payoff given the assumed solution concept. This definition will guarantee that $\underline{V}(\mu)$ is convex (if it were not, Nature could disclose additional public information to further decrease the Sender’s payoff, contradicting the definition of \underline{V}); moreover, it will typically be lower semi-continuous. Hence, \underline{V} is trivially the lower convex closure of \underline{V} . The Sender’s payoff under the assumed conjecture, \widehat{V} , is then defined as above, with the exception that the Sender’s conjecture can now specify discriminatory disclosures by Nature. Provided that \widehat{V} is upper semi-continuous, all our results apply.

Bypassing \underline{V} might be seen as conceptually compelling: It means that equilibrium selection and information provision by Nature are put on an equal footing. However, under this symmetric approach, some of the assumptions of our results become more difficult to verify. For example, to identify the set \mathcal{F} in [Theorem 1](#), one would in principle need to compute \underline{V} , which can be challenging in some applications. For example, when the assumed solution concept is BCE, computing \underline{V} requires characterizing the Sender’s payoff in the worst BCE consistent with any given common posterior $\mu \in \Delta\Omega$. In certain cases, the set \mathcal{F} can be identified without computing the entire set of BCE: To show that states in the support of some belief μ must be separated, it suffices to construct a single BCE consistent with μ that yields a payoff

to the Sender below the full-disclosure level—see Subsection 5.4 for an illustration.

An alternative approach to incorporating private disclosures into our analysis is by applying the results of Mathevet et al. (2020) who propose a formal decomposition of any signal into a public and a (purely) private component. Relying on their characterization, we can define $\underline{V}(\mu)$ as the expected payoff of the Sender in the worst-case equilibrium when Nature complements μ with purely private signals. Then, relative to \underline{V} , $\underline{\mathcal{V}}$ captures the effects of additional *public* disclosure by Nature, implying that $\underline{\mathcal{V}}$ is the lower convex closure of \underline{V} . This approach is more tractable than the one discussed previously if computing such \underline{V} is easier than computing $\underline{\mathcal{V}}$ directly.

5 Applications

In this section, we present four applications, illustrating the four cases we have considered: the baseline model, the single-Receiver case under a general conjecture, and two models with multiple Receivers and public or private disclosure by Nature, respectively. The results follow as straightforward consequences of our general theory—we include the proofs for completeness in the Online Appendix.

5.1 Lemons problem

The Sender is a seller, and the Receiver is a buyer. The seller values an indivisible good at ω while the buyer values it at $\omega + D$, where $D > 0$ is a known constant. The value ω is observed by the seller but not by the buyer. To avoid confusion, we use a “tilde” ($\tilde{\omega}$) whenever we refer to ω as a random variable. The seller can commit to an information disclosure policy about the object quality, ω . We consider a simple trading protocol in which, after the information structure is determined, a random exogenous price p is drawn from a uniform distribution over $[0, 1]$ and trade happens if and only if both the buyer and the seller agree to trade at that price (the exogenous price can be interpreted as a benchmark price in the market, or can be seen as coming from an exogenous third party, e.g., a platform). That is, if the state is ω and the buyer’s belief about the state is μ , then trade happens if and only if $p \geq \omega$ and $\mathbb{E}_\mu[\tilde{\omega} | \tilde{\omega} \leq p] + D > p$.¹⁴ To avoid trivial cases, we assume that the support of the price distribution contains Ω , that is, $\Omega \subseteq (0, 1)$. The seller chooses the signal before observing ω (hence the choice of the signal by the seller reveals

¹⁴Because p is drawn from a continuous distribution, the way the buyer’s indifference is resolved plays no role in this example.

no additional information to the buyer). We are interested in finding the robustly optimal policy for the seller, under the conjecture that the buyer does not have any exogenous information other than the one contained in the prior.

The payoff to the seller under the conjecture is given by¹⁵

$$\widehat{V}(\mu) = \sum_{\omega \in \Omega} \left(\int_{\omega}^1 (p - \omega) \mathbf{1}_{\{\mathbb{E}_{\mu}[\tilde{\omega} | \tilde{\omega} \leq p] + D > p\}} dp \right) \mu(\omega).$$

In this example, $\underline{V} = \widehat{V}$ because the buyer's tie-breaking rule does not influence the Sender's payoff in expectation. The following lemma identifies a key property of robust solutions.

Lemma 2. *Any two states ω and ω' such that $|\omega - \omega'| > D$ must be separated under any robust solution.*

For intuition, suppose that only types ω' and ω are present in the market, and $\omega > \omega'$. If the buyer's posterior belief μ puts sufficient mass on the low state ω' , namely, $\mathbb{E}_{\mu}[\tilde{\omega}] + D < \omega$, then the high type ω does not trade. Indeed, any price below ω is rejected by the ω -type seller, and any price above ω is rejected by the buyer. In contrast, the high type ω would trade with positive probability if her type were disclosed to the buyer. At the same time, type ω' does not benefit from the presence of the higher type ω because of adverse selection: $\mathbb{E}_{\mu}[\tilde{\omega} | \tilde{\omega} \leq p] = \omega'$ for all prices $p \in [\omega' + D, \mathbb{E}_{\mu}[\tilde{\omega}] + D]$ that could be accepted by the buyer if she did not condition on the fact that $\tilde{\omega} \leq p$. In short, if ω and ω' are not separated, Nature can induce posterior beliefs that reduce the high type's probability of trade (relative to full disclosure) without improving the terms of trade for the low type. It follows that Nature can push the seller's expected payoff below what she could obtain by fully disclosing the state.

The preceding argument does not apply to types that are less than D apart because the adverse selection problem is mute for such types, as the next lemma shows.

Lemma 3. *Suppose that $\text{supp}(\mu) \subseteq [\underline{\omega}_{\mu}, \underline{\omega}_{\mu} + D]$, where $\underline{\omega}_{\mu}$ is the minimum of $\text{supp}(\mu)$. Then, $\mathbf{1}_{\{\mathbb{E}_{\mu}[\tilde{\omega} | \tilde{\omega} \leq p] + D > p\}} = \mathbf{1}_{\{\mathbb{E}_{\mu}[\tilde{\omega}] + D > p\}}$ for any $p \geq \underline{\omega}_{\mu}$.*

Intuitively, Lemma 3 states that when μ puts mass on types that are less than D apart, adverse selection has no bite – the buyer trades under the same prices as if

¹⁵Note that the seller's payoff is computed before the price p is realized and before the seller learns her value ω for the good.

the seller did not possess private information (that is, she does not need to condition on $p \geq \tilde{\omega}$). We can now use this observation to prove a result that helps characterize robust solutions. For any $B \subseteq \Omega$, we let $\text{diam}(B) := \max(B) - \min(B)$.

Lemma 4. *Fix any $B \subseteq \Omega$ such that $\text{diam}(B) \leq D$. Then, $\underline{V}|_{\Delta B}(\mu)$ is concave on ΔB (and non-affine if $|B| \geq 2$).*

Lemma 4 states that the seller does not benefit from splitting posterior beliefs with sufficiently small supports. The reason is that, once the possible detrimental effects of adverse selection are eliminated (which is the case when $\text{diam}(B) \leq D$), further informing the buyer of her value for the good only reduces the seller's ability to extract surplus from the buyer. The next result is then a simple corollary.

Lemma 5. $\mathcal{F} = \{B \subset \Omega : \text{diam}(B) \leq D\}$.

Indeed, we know that $\text{diam}(B) \leq D$ is necessary for $B \in \mathcal{F}$ by Lemma 2. Lemma 4 tells us that this condition is sufficient as well: Because $\underline{V}|_{\Delta B}(\mu)$ is concave when $\text{diam}(B) \leq D$, it lies everywhere above the full-disclosure payoff on that subspace.

Lemma 5 states that any worst-case optimal distribution must disclose enough information to make the adverse selection problem mute. Furthermore, there is no need to disclose any additional information. Because disclosing additional information is detrimental to the Sender, as implied by Lemma 4 combined with the fact that $\underline{V} = \widehat{V}$, any robust solution discloses just enough information to eliminate the adverse selection problem.

Proposition 2. *Under any robust solution ρ_{RS} , for any $\mu, \mu' \in \text{supp}(\rho_{RS})$, $\text{diam}(\text{supp}(\mu)) \leq D$; $\text{diam}(\text{supp}(\mu')) \leq D$; but $\text{diam}(\text{supp}(\mu) \cup \text{supp}(\mu')) > D$.*

The result says that robust solutions are minimally informative among those that remove the adverse selection problem. Indeed, since $\widehat{V}|_{\Delta B}(\mu)$ is concave but not affine on ΔB whenever $\text{diam}(B) \leq D$, if $\text{diam}(\text{supp}(\mu) \cup \text{supp}(\mu')) \leq D$, the Sender could merge μ and μ' into a single posterior, improve her expected payoff, while maintaining worst-case optimality. In particular, full disclosure is not a robust solution as long as there exist ω and ω' in Ω that are less than D apart.

A closed-form characterization of the optimal policy seems difficult (for the same reasons that make it difficult to solve for a Bayesian solution). However, one of the benefits of the proposed solution concept is that it permits one to identify important

properties that all robust solutions must satisfy. Here, that property is that robust solutions must disclose just enough information to neutralize the adverse selection problem. Note that such property need not extend to Bayesian solutions. We can verify that by looking at the tractable binary-state case: When the two states are more than D apart, the unique Bayesian solution pools these states with positive probability, whereas a robust solution separates them, by Lemma 2.

5.2 Informed Receiver: Guo and Shmaya (2019)

We now analyze another simple model of buyer-seller interactions along the lines of Guo and Shmaya (2019): The seller owns an indivisible good of quality ω and gets a payoff of 1 if and only if the buyer accepts to trade at an exogenously specified price p . The seller's conjecture is that the buyer has private information about the product's quality ω summarized by the realization r of a signal drawn from a finite set $\mathcal{R} \subset \mathbb{R}$, according to the distribution $\pi_0(r|\omega)$. The seller also conjectures that, in case of indifference, the buyer will play favorably to the seller, which amounts to accepting to trade. The seller can provide any information about ω to the buyer.

Guo and Shmaya (2019) show that, when $\pi_0(r|\omega)$ satisfies MLRP (formally, when $\pi_0(r|\omega)$ is log-supermodular), a Bayesian solution for the above conjecture has an *interval structure*: Each buyer's type r is induced to trade on an interval of states, and less optimistic types trade on an interval that is a subset of the interval over which more optimistic types trade. Here, we characterize the robust solution for the seller. To avoid uninteresting cases, we assume that π_0 is not fully revealing.¹⁶

Given any final posterior $\mu_0^{s,r} \in \Delta\Omega$ for the buyer, the seller's payoff under the least-favorable tie-breaking rule is

$$\underline{V}(\mu_0^{s,r}) = \mathbf{1}_{\{\sum_{\omega \in \Omega} \omega \mu_0^{s,r}(\omega) > p\}}.$$

The seller's payoff from inducing a posterior μ_0^s under her conjecture (where the posterior is obtained by conditioning only on the realization of the seller's signal s) is equal to

$$\widehat{V}(\mu_0^s) = \sum_{\omega \in \Omega} \sum_{r \in \mathcal{R}} \mathbf{1}_{\left\{ \frac{\sum_{\omega' \in \Omega} \omega' \pi_0(r|\omega') \mu_0^s(\omega')}{\sum_{\omega' \in \Omega} \pi_0(r|\omega') \mu_0^s(\omega')} \geq p \right\}} \pi_0(r|\omega) \mu_0^s(\omega).$$

The following result is then a simple implication of Corollary 2.

¹⁶That is, conditional on any state ω , there is positive conditional probability that the signal realization r from π_0 does not reveal that the state is ω .

Proposition 3. *Any robust solution separates any state $\omega \leq p$ from any state $\omega' > p$.*

A robust solution thus eliminates buyer’s uncertainty over whether or not to purchase the product. In other words, when the seller faces uncertainty about the buyer’s exogenous information, she cannot benefit from disclosing information strategically.

Intuitively, if a posterior belief pools together states that are both below and above p , Nature could send a signal that induces a sufficiently pessimistic belief about the quality of the good to induce the buyer not to trade, even when the good is of high quality. By fully disclosing the state, the seller guards herself against such a possibility and ensures that all high-quality goods ($\omega > p$) are bought with certainty.

5.3 Regime change

Next, we study an application featuring multiple Receivers in which Nature is restricted to disclosing information publicly and where the functions \underline{V} and \widehat{V} represent the Sender’s payoff under the lowest and the highest rationalizable profiles in the continuation game among the Receivers, respectively.

Consider the following stylized game of regime change. A continuum of agents of measure 1, uniformly distributed over $[0, 1]$, must choose between two actions, “attack” or “not attack.” Let $a_i = 1$ (respectively, $a_i = 0$) denote the decision by agent i to attack (respectively, not attack), and A the aggregate size of the attack. Regime change happens if and only if $A \geq \omega$, where $\omega \in \Omega \subset \mathbb{R}$ parametrizes the strength of the regime (the underlying fundamentals) and is commonly believed to be drawn from a distribution μ_0 whose support intersects each of the following three sets: $(-\infty, 0)$, $[0, 1]$, and $(1, \infty)$. Each agent’s payoff from not attacking is normalized to zero, whereas his payoff from attacking is equal to g in case of regime change and b otherwise, with $b < 0 < g$. Hence, under complete information, for $\omega \leq 0$ (alternatively, $\omega > 1$), it is dominant for each agent to attack (alternatively, not to attack), whereas for $\omega \in (0, 1]$ both attacking and not attacking are rationalizable actions (see, among others, [Inostroza and Pavan, 2022](#) and [Morris et al., 2020](#) for similar games of regime change). The Sender’s payoff is equal to $1 - A$ (that is, she seeks to minimize the size of the aggregate attack). The Sender is constrained to disclose the same information to all agents, as in the case of stress testing. Contrary to what is typically assumed in the literature, the Sender is uncertain about the exogenous information the agents are endowed with.

The Sender’s conjecture is that the agents do not have access to any information

other than that contained in the common prior μ_0 and that, in case of multiple rationalizable profiles, the agents play the profile most favorable to the Sender. The Bayesian solution for the above conjecture is similar to the one in the judge example of [Kamenica and Gentzkow \(2011\)](#). To see this, note that for the Receivers to abstain from attacking, it must be that their common posterior assigns probability at least $\alpha := g/(g+|b|)$ to the event that $\omega > 0$.¹⁷ Let $\mu_0^+ := \mu_0(\omega > 0)$ denote the probability assigned by the prior μ_0 to the event that $\omega > 0$ and (to make the problem interesting) assume that $\mu_0^+ < \alpha$, so that, in the absence of any disclosure, all agents attack under the unique rationalizable profile. Under the assumed conjecture, the Sender then maximizes her payoff through a policy that, when $\omega > 0$, sends the “null” signal $s = \emptyset$ with certainty, whereas, when $\omega \leq 0$, fully discloses the state with probability $\phi_{BP} \in (0, 1)$ and sends the signal $s = \emptyset$ with the complementary probability, where ϕ_{BP} is defined by $\mu_0^+ / [\mu_0^+ + (1 - \mu_0^+)(1 - \phi_{BP})] = \alpha$.

The above Bayesian solution, however, is not robust. First, when the agents assign sufficiently high probability to the event $\omega \in (0, 1]$, while it is rationalizable for each of them to abstain from attacking, it is also rationalizable for them to attack. Hence, if the Sender does not trust that the agents will coordinate on the rationalizable profile most favorable to her, it is not enough to persuade them that $\omega > 0$; the Sender must persuade them that $\omega > 1$. Furthermore, if the agents may have access to information other than the one contained in the prior, then worst-case optimality requires that all states $\omega > 1$ be separated from all states $\omega \leq 1$. (For any induced posterior whose support contains both states $\omega > 1$ and states $\omega \leq 1$, Nature can construct another posterior under which it is rationalizable for all agents to attack also when $\omega > 1$, thus bringing the Sender’s payoff below her full-disclosure payoff.) One may then conjecture that full disclosure of the state is a robust solution under the conjecture described above. This is not the case. The reason is that, in case Nature (and the agents) play according to the Sender’s conjecture, fully disclosing the state triggers an aggregate attack of size $A = 1$ for all $\omega \leq 0$. The Sender can do better by pooling states below 0 with states in $[0, 1]$ and then hope that Nature (and the agents) play as conjectured. The next proposition summarizes the above results.

Proposition 4. *The following policy is a Bayesian solution. If $\omega \leq 0$, the state is fully revealed with probability $\phi_{BP} \in (0, 1)$ whereas, with the complementary probability, the Sender sends the “null” signal $s = \emptyset$. If $\omega > 0$, the signal $s = \emptyset$ is sent with*

¹⁷When, instead, $\mathbb{P}(\omega > 0) < \alpha$, the unique rationalizable profile is for each agent to attack.

certainty. Such a policy, however, is not robust. The following policy, instead, is a robust solution. If $\omega \leq 0$, the state is fully revealed with probability $\phi_{RS} \in (0, 1)$, with $\phi_{RS} > \phi_{BP}$, whereas, with the complementary probability, the signal $s = \emptyset$ is sent. If $\omega \in (0, 1]$, the signal $s = \emptyset$ is sent with certainty. Finally, if $\omega > 1$, the state is fully revealed with certainty.

While neither the Bayesian nor the robust solutions in the above proposition are unique, any robust solution must fully separate states $\omega > 1$ from states $\omega \leq 1$, whereas any Bayesian solution pools states $\omega > 1$ with states $\omega \leq 1$. The robust solution displayed in the proposition Blackwell dominates the Bayesian solution, consistently with Corollary 6.

5.4 Multiple Receivers and private disclosures by Nature

Our last application is a variant of the prosecutor-judge example of Section 1 in which the prosecutor faces two judges. Each judge has the same preferences as in the original example, but with the sentence of each judge now interpreted as the judge's recommendation.¹⁸ The defendant is convicted only if both judges vote to convict him. In this case, the sentence specifies a number of years equal to the minimum of the numbers asked by the two judges. Let $x_j \in [\underline{x}, \bar{x}]$, with $\underline{x} > 0$, denote the number of years asked by judge $j = 1, 2$. As in the original game, each judge feels morally obliged to convict if her posterior belief that the defendant is guilty exceeds $2/3$ and to acquit otherwise. When she recommends to convict, the number of years that the judge asks is linearly increasing in the conditional probability she assigns to state f , exactly as in the original example of Section 1. Denote by $A_j = \{0\} \cup [\underline{x}, \bar{x}]$ the judge's action set, with $a_j = 0$ denoting the recommendation to acquit, and by $\mu_j(\omega)$ the judge's posterior belief that the state is ω . Then,

$$a_j(\mu_j) = \mathbf{1}_{\{\mu_j(m) + \mu_j(f) > \frac{2}{3}\}} \min\{\bar{x}, \underline{x} + \frac{2\mu_j(f)}{\mu_j(f) + \mu_j(m)}(\bar{x} - \underline{x})\},$$

whereas the actual sentence is given by $x(\mu_1, \mu_2) = \min\{a_1(\mu_1), a_2(\mu_2)\}$.

As before, the prosecutor maximizes the expected number of years determined by the actual sentence. Her conjecture is that each judge's only information is that associated with the common prior μ_0 , given by $\mu_0(i) = 1/2$, and $\mu_0(m) = \mu_0(f) = 1/4$.

It is easy to see that the Bayesian solution is the same as in the original version

¹⁸That is, each judge's utility depends only on the recommendation she makes, not on the actual sentence—the judges are Kantianists rather than Consequentialists.

with a single judge. It is also easy to see that, when Nature is expected to disclose the same information to both judges, the unique robust solution is the same as in the single-judge case: separate the state $\omega = i$ and pool the other two states. Indeed, in this case, we have that $\underline{V}(\mu) = x(\mu, \mu) = a_1(\mu)$, and thus the objective function of the prosecutor is the same as in the single-judge case.

Suppose, instead, that the prosecutor does not exclude the possibility that Nature discloses different information to the two judges, perhaps because they can call different witnesses and question them independently. As explained in Section 4, in case of private disclosure by Nature, it is not helpful to define \underline{V} and \underline{V} separately. Instead, we set $\underline{V} = \underline{V}$ with $\underline{V}(\mu)$ defined as the Sender-worst BCE payoff consistent with the common posterior belief μ . Even though the game between the judges is simple, computing \underline{V} is difficult. Instead, we make use of Corollary 2: States ω, ω' must be separated by a robust solution whenever, for some $\lambda \in (0, 1)$, $\underline{V}(\lambda\delta_\omega + (1-\lambda)\delta_{\omega'}) < \lambda\underline{V}(\delta_\omega) + (1-\lambda)\underline{V}(\delta_{\omega'})$. The right-hand side of this condition does not depend on what the Sender expects Nature to do: when the state is disclosed, there is a unique BCE. Furthermore, because the left-hand side is never larger than the payoff that the Sender expects when Nature is restricted to public disclosures, we have that any worst-case optimal policy (and hence any robust solution) must separate the state $\omega = i$ from $\omega' \in \{m, f\}$, just like when Nature is restricted to public disclosures. Suppose the states $\omega = m$ and $\omega' = f$ are not separated. Then, starting from any posterior with support $\{m, f\}$ induced by the Sender, Nature can first generate the common posterior $(1/2)\delta_m + (1/2)\delta_f$ using a public signal, and then engineer an additional discriminatory disclosure that fully reveals the state to judge 1, and discloses a binary signal $r_2 \in \{m, f\}$ to judge 2 that matches the true state with conditional probability $2/3$ in each state. Under such a policy, when the state is m , the actual sentence is equal to \underline{x} because this is the sentence asked by the fully-informed judge 1. When, instead, the state is f , the fully-informed judge 1 recommends \bar{x} , whereas the less-informed judge 2 recommends \bar{x} with probability $2/3$ (after observing $r_2 = f$) and $(1/3)\underline{x} + (2/3)\bar{x}$ with probability $1/3$ (after observing $r_2 = m$). We thus have that

$$\underline{V}\left(\frac{1}{2}\delta_m + \frac{1}{2}\delta_f\right) < \frac{1}{2}\underline{x} + \frac{1}{2}\bar{x} = \frac{1}{2}\underline{V}(\delta_m) + \frac{1}{2}\underline{V}(\delta_f).$$

By Corollary 2, states m and f must also be separated by any robust solution. Full disclosure is therefore the unique robust solution. This application of Corollary 2

illustrates the force of Theorem 1: We are able to characterize the unique robust solution by constructing one BCE at a particular posterior belief (as opposed to computing all BCE at all possible beliefs).

Suppose that the two judges are obliged to share all their information before making the decision, and the Sender knows that. By Aumann’s theorem, this case is equivalent to assuming that Nature can only send public signals. An interesting conclusion obtains: If the Sender is sure that the judges share their information, she should reveal less information than if she thought that it is possible that the judges are asymmetrically informed.

6 Alternative approaches to robustness

6.1 Weighted objective function

Our solution concept assumes that the Sender follows a lexicographic approach: She first maximizes her objective in the worst-case scenario, and only in case of indifference chooses between worst-case optimal policies based on her conjecture. In this section, we examine an alternative objective function under which the Sender attaches a weight $\lambda \in [0, 1]$ to the worst-case scenario, and a weight $1 - \lambda$ to the base-case scenario.¹⁹ A possible interpretation is that the Sender is Bayesian, and the weights reflect the assessed probabilities of Nature being adversarial and behaving as conjectured by the Sender. We show that, under mild regularity conditions, robust solutions correspond exactly to solutions for the weighted objective function provided that the weight λ on the worst-case scenario is sufficiently large.²⁰

Throughout, we work with reduced-form payoff functions that satisfy the properties listed in Section 4. Formally, for some $\lambda \in [0, 1]$, the Sender’s problem is

$$\sup_{\rho \in \Delta \Delta \Omega} \left\{ \lambda \int \underline{V}(\mu) d\rho(\mu) + (1 - \lambda) \int_{\Delta \Omega} \widehat{V}(\mu) d\rho(\mu) \right\} \quad (6.1)$$

subject to (BP). Recall that \widehat{V} is assumed upper semi-continuous (we also assume that it is bounded), and \underline{V} is convex and continuous (see footnote 11). Therefore, the problem for a fixed λ is equivalent to a standard Bayesian persuasion problem

¹⁹When the Sender’s conjecture is that Nature behaves favorably to her (as in the analysis in Section 3), this approach shares some similarities with the literature on alpha-max-min preferences (Hurwicz, 1951, Gul and Pesendorfer, 2015, Grant et al., 2020).

²⁰We thank Emir Kamenica and Ron Siegel for suggesting we investigate the validity of this result.

with an upper semi-continuous objective function $\widehat{V}_\lambda(\mu) := \lambda \underline{V}(\mu) + (1 - \lambda) \widehat{V}(\mu)$, and a Bayes-plausible ρ is a solution if and only if it concavifies \widehat{V}_λ at the prior μ_0 . Our goal is to relate the solutions to the problem defined by (6.1) (which we denote by $S(\lambda)$ and refer to as λ -solutions) to robust solutions. Note that 0-solutions coincide with Bayesian solutions whereas 1-solutions are worst-case optimal solutions.

We first introduce some notation. Let d denote a metric on $\Delta\Omega$; since Ω is finite, all metrics are equivalent, and we can use the L^1 metric for convenience: $d(\mu, \eta) := \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|$. For any $\mu \in \Delta\Omega$ and $A \subseteq \text{supp}(\mu)$, let $\mu|_A$ be the belief obtained from μ by conditioning on the event A , that is, $\mu|_A(\omega) = \mu(\omega)/\mu(A)$ for $\omega \in A$.

Definition 4. The function \widehat{V} is regular if there exists a constant L such that for any non-degenerate $\mu \in \Delta\Omega$ and $A \subseteq \text{supp}(\mu)$, $\text{co}\widehat{V}(\mu|_A) \geq \widehat{V}(\mu) - Ld(\mu, \mu|_A)$.

Regularity requires that, for any μ and any subset A of its support, the designer can generate an expected payoff that is “not much worse” than $\widehat{V}(\mu)$ by only relying on beliefs supported on A , with the allowed loss scaling linearly with the distance between μ and ΔA . Since $d(\mu, \mu|_A) = 2\mu(A^c)$, where A^c denotes the complement of the set A , regularity only has bite when $\mu(A^c)$ is small; else the condition follows from boundedness of the function \widehat{V} . Lipschitz continuous functions are trivially regular since $\text{co}\widehat{V}(\mu|_A) \geq \widehat{V}(\mu|_A)$. However, the condition is significantly weaker because (i) the Lipschitz condition is required to hold only for pairs of beliefs of the form $(\mu, \mu|_A)$, (ii) only one direction of the Lipschitz inequality is required, and (iii) $\text{co}\widehat{V}(\mu|_A)$ is in general higher than $\widehat{V}(\mu|_A)$. The condition only rules out functions \widehat{V} that decrease at an infinite rate along a sequence of beliefs assigning a vanishing probability to some subset of their support. And, indeed, regularity allows for highly discontinuous objective functions. For example, when Ω is binary, regularity only requires that \widehat{V} has bounded steepness (as defined by Gale, 1967) at the two endpoints of the interval $\Delta\Omega$. In the mean-measurable case, in which $\widehat{V}(\mu) = v(\mathbb{E}_\mu[\omega])$ for some upper-semi continuous function v , it can be shown that \widehat{V} is regular if v has bounded steepness at the (finitely many) points $\omega \in \Omega$.

Theorem 2. *Suppose that \widehat{V} is regular. There exists $\bar{\lambda} < 1$ such that, for all $\lambda \in (\bar{\lambda}, 1)$, $S(\lambda)$ coincides with the set of robust solutions.*

It may seem puzzling that the equivalence between λ -solutions and robust solutions is achieved exactly at sufficiently high λ , rather than only in the limit as $\lambda \nearrow 1$.

This is a consequence of Proposition 1 which shows that worst-case optimality imposes restrictions only on the supports of the induced posteriors, and not on the weights assigned to any admissible posterior. Combined with regularity of \widehat{V} this property implies that, for high λ , the gain from inducing any non-admissible posterior under the base-case scenario is always dominated by the loss under the worst-case scenario.

In the Online Appendix, we show that, even without the regularity condition, a slightly weaker version of one direction of the equivalence holds: Any limit of λ -solutions as $\lambda \nearrow 1$ is a robust solution (and therefore some robust solution is a limit of λ -solutions). However, we also show, by means of an example, that there exist robust solutions that cannot be obtained as the limit of λ -solutions.

6.2 Dominance

In this section, we examine the relationship between robustness and the notion of undominated policies. When the Sender faces non-Bayesian uncertainty over the Receivers' information and strategy selection, it is natural for her to avoid signals that are dominated. Informally, we say that a signal dominates another one if it performs weakly better for *any* choice of Nature, and strictly better for some. Our next result shows that—under certain conditions—any robust solution is *undominated*.

To define dominance formally, we again bypass the distinction between information disclosure and strategy selection. We introduce a function \bar{V} interpreted as the Sender's payoff from inducing a common posterior μ , when Nature selects a signal and a strategy profile (consistent with the assumed solution concept) that *maximize* the Sender's payoff. Note that \bar{V} must be concave under this interpretation (otherwise, Nature could further increase the Sender's payoff by concavifying \bar{V} with an additional public signal). Formally, let \bar{V} be any concave continuous function such that $\bar{V} \geq \widehat{V} \geq \underline{V}$. If Nature is allowed to respond to any posterior μ induced by the Sender with an arbitrary signal and strategy profile (consistent with the assumed solution concept), then it can generate any payoff function V that lies between \underline{V} and \bar{V} . This motivates the following definition of dominance.

Definition 5. A Bayes-plausible distribution $\rho \in \Delta\Delta\Omega$ dominates a Bayes-plausible distribution $\rho' \in \Delta\Delta\Omega$ if, for *any* measurable $V : \Delta\Omega \rightarrow \mathbb{R}$ such that $V(\mu) \in [\underline{V}(\mu), \bar{V}(\mu)]$ for any $\mu \in \Delta\Omega$, we have that $\int V(\mu)d\rho(\mu) \geq \int V(\mu)d\rho'(\mu)$, with the inequality strict for at least one such function V . A Bayes-plausible distribution ρ is undominated if there exists no Bayes-plausible distribution ρ' that dominates it.

Theorem 3. (a) *At least one robust solution is undominated.* (b) *If $\text{co}\widehat{V} = \overline{V}$, then all robust solutions are undominated.*

The result in part (a) follows from the fact that a robust solution can be dominated only by another robust solution (by the definition of robustness). The result in part (b) is more convoluted. Heuristically, it follows from the fact that, given any pair of robust solutions ρ and ρ' , if, for some feasible response V by Nature, ρ performs strictly better than ρ' , then one can construct another feasible response V' under which ρ' performs strictly better than ρ . The construction of V' hinges on the fact that the two solutions perform equally well both under the worst-case scenario and under the Sender's conjecture, along with the fact that the Sender's payoff under the conjecture is linked to the maximal feasible payoff over all possible responses by Nature (by the condition $\text{co}\widehat{V} = \overline{V}$). Without the last property, the fact that the policies are both robust solutions does not impose enough structure on the way they may perform under alternative responses by Nature, and one may dominate the other. As an illustration, in the judge-prosecutor example of Section 1, when the Sender's conjecture is that Nature always fully reveals the state, then full disclosure is robust. However, such a policy is dominated by the one that separates $\{i\}$ from $\{f, m\}$.

One may wonder whether Bayesian solutions are also undominated. The answer is no, even when $\text{co}\widehat{V} = \overline{V}$. We provide an example in the Online Appendix.

6.3 Alternative extensive forms

It is common in the mechanism-design literature to model robustness as worst-case optimality of a designer's policy in a game against adversarial Nature. Under this modeling convention, the properties of worst-case optimal policies generally depend on the assumptions about the extensive form. It is also known that randomization may sometimes improve the designer's payoff guarantee when Nature does not observe the designer's policy choice.²¹ In this subsection, we discuss alternative extensive forms and randomization in the context of our model. We start by restricting attention to non-stochastic choices of a policy by the Sender, and consider three cases:

1. Baseline model: Nature chooses its signal after observing both the Sender's signal q and its realization s ;

²¹For a discussion of this point, see for example [Ke and Zhang \(2020\)](#).

2. Conditionally independent signals: Nature chooses its signal after observing the Sender’s signal q (but not its realization s);
3. Simultaneous-move game: The Sender and Nature choose their respective signals simultaneously, without observing each other’s choices.

Our baseline definition of worst-case optimality (case 1) is motivated by the Sender’s uncertainty about the order in which various sources of information are consulted by the Receiver. This assumption is appropriate whenever the Sender does not feel confident that she is the last one to speak. In particular, this case captures the possibility that the Receiver may acquire additional information after learning the Sender’s signal realization.

Case 2 of conditionally independent signals arises when the exogenous sources of information depend on the Sender’s choice of a signal but not on the Sender’s signal realization. This corresponds to a situation in which the Sender must publicly commit to her signal *ex-ante*, and she is concerned that the Receiver might receive some information before observing the Sender’s signal realization. Formally, having observed q , Nature selects a signal $\pi(r|\omega)$ that is independent of $q(s|\omega)$ conditional on the state ω .

Case 3 of a simultaneous-move game arises when the Sender believes that the exogenous sources of information, captured by Nature’s signal $\pi(r|\omega)$, have been already determined but are unknown to her. In this case, the Sender assumes that she is the last one to speak, and Nature’s choice is a way of capturing the Sender’s ambiguity about the fixed environment she is facing.

Cases 2 and 3 are equivalent in our model, in the sense that they lead to the same set of worst-case optimal policies for the Sender. We show this formally in the Online Appendix. Intuitively, a version of the minimax theorem holds in our setting, and the full-disclosure payoff is the value of the zero-sum game between the Sender and Nature. Given the equivalence, we refer to these two cases jointly as simultaneous-move (SM) robustness, and examine properties of SM-robust solutions in the Online Appendix. As we discuss there, some of the results are in common to those in the baseline case but there are also important differences (for example, existence of a SM-robust solution is not guaranteed and Bayesian solutions can be more informative than SM-robust solutions).

When Nature observes the Sender’s choice of a signal q (in cases 1 and 2), it clearly

does not benefit the Sender to randomize over her choice of a policy. Perhaps more surprisingly, randomization also does not help in case 3. Formally, the Sender cannot expand the set of worst-case optimal distributions of posterior beliefs by randomizing over signals. Indeed, in the Bayesian persuasion model, any randomization over signals is itself a signal; thus, if the Sender can induce some distribution of posteriors via a random choice of a signal, then she can also generate the same distribution by deterministically choosing a composite signal (Nature’s problem is the same in both cases, so the optimal response by Nature is unaffected). Thus, our restriction to deterministic choices of policies is without loss of generality.

We briefly comment on two additional models of robustness. First, one may consider a situation where the Sender is uncertain about the Receiver’s prior (see [Kosterina \(2021\)](#)). If the prior is generated by providing an additional experiment to the Receiver, then this corresponds to the case considered in our Online Appendix. If, instead, such a requirement is not imposed, as in [Kosterina \(2021\)](#), then Nature’s problem is no longer a Bayesian persuasion problem, and our techniques do not apply. Second, one may contemplate extensive forms in which Nature is able to obfuscate the information provided by the Sender (e.g., through signal jamming). This would correspond to a robust approach to the case of an inattentive Receiver (considered by [Lipnowski et al., 2020](#), and [Bloedel and Segal, 2021](#)), and is worth investigating in future research.²²

7 Conclusions

We introduce and analyze a novel solution concept for information design in settings in which the Sender faces uncertainty about the Receivers’ sources of information and strategy selection. The Sender first identifies all information structures that are “worst-case optimal,” i.e., that yield the highest payoff when Nature provides information and coordinates the Receivers’ play in an adversarial fashion. The Sender then picks an information structure that maximizes her expected payoff under her conjecture—much like in the standard Bayesian persuasion model—but among information structures that are worst-case optimal. Our main technical result identifies sets of states that can be present together in one of the induced posteriors and states

²²See also [Arieli et al. \(2022\)](#) for a model in which information provided by the Sender to the Receiver is mediated by a mediator observing the realization of the Sender’s signal but not the state.

that must be separated. We show that robust solutions exist and can be characterized using canonical tools; we qualify in what sense they lead to more information disclosure than Bayesian solutions; we argue that, under reasonable conditions, robustness guarantees that the solution is undominated; and we illustrate the results in the context of existing and novel applications.

Throughout the analysis, we restrict attention to the case of *public persuasion* in which the Sender discloses the same information to all the Receivers. In future work, it would be interesting to extend the analysis to private persuasion, whereby the Sender discloses different signals to different Receivers. Our analysis also relies on the assumption that Nature can engineer any signal. It is natural to ask how the properties of robust solutions change as we impose constraints on the set of signals that Nature can choose from. Finally, it would be interesting to see how existing results in the persuasion literature change once robustness is accounted for, and whether robust solutions can provide insights about problems that are inherently intractable in the Bayesian framework.

References

- Arieli, Itai, Yakov Babichenko, and Fedor Sandomirskiy**, “Bayesian Persuasion with Mediators,” *Working Paper*, 2022.
- Au, Pak Hung and Keiichi Kawai**, “Competitive information disclosure by multiple senders,” *Games and Economic Behavior*, 2020, *119*, 56–78.
- Aumann, Robert J and Michael Maschler**, *Repeated Games with Incomplete Information*, MIT press, 1995.
- Babichenko, Yakov, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi**, “Regret-Minimizing Bayesian Persuasion,” *Working Paper*, 2021.
- Bergemann, Dirk and Stephen Morris**, “Bayes correlated equilibrium and the comparison of information structures in games,” *Theoretical Economics*, 2016, *11* (2), 487–522.
- and –, “Information Design: A Unified Perspective,” *Journal of Economic Literature*, 2019, *57*, 44–95.
- , **Benjamin Brooks, and Stephen Morris**, “The Limits of Price Discrimination,” *American Economic Review*, March 2015, *105* (3), 921–57.
- Bloedel, Alexander W. and Ilya Segal**, “Persuading a Rationally Inattentive Agent,” *Working Paper, Stanford University*, 2021.
- Board, Simon and Jay Lu**, “Competitive Information Disclosure in Search Markets,” *Journal of Political Economy*, 2018, *126* (5), 1965–2010.

- Börger, Tilman**, “(No) Foundations of Dominant-Strategy Mechanisms: A Comment on Chung and Ely (2007),” *Review of Economic Design*, 2017, 21, 73–82.
- Brooks, Benjamin and Songzi Du**, “Optimal Auction Design With Common Values: An Informationally Robust Approach,” *Econometrica*, 2021, 89 (3), 1313–1360.
- Carroll, Gabriel**, “Information Games and Robust Trading Mechanisms,” *Working Paper*, 2019.
- Cui, Zhihan and Dilip Ravindran**, “Competing Persuaders in Zero-Sum Games,” *Working Paper, Columbia University*, 2020.
- Gale, David**, “A Geometric Duality Theorem with Economic Applications,” *Review of Economic Studies*, 1967, 34 (1), 19–24.
- , **Victor Klee**, and **R. T. Rockafellar**, “Convex Functions on Convex Polytopes,” *Proceedings of the American Mathematical Society*, 1968, 19 (4), 867–873.
- Gentzkow, Matthew and Emir Kamenica**, “Competition in Persuasion,” *The Review of Economic Studies*, 10 2016, 84 (1), 300–322.
- and – , “Bayesian persuasion with multiple senders and rich signal spaces,” *Games and Economic Behavior*, 2017, 104, 411 – 429.
- Grant, Simon, Patricia Rich, and Jack Stecher**, “Worst- and Best-Case Expected Utility and Ordinal Meta-Utility,” *Working Paper*, 2020.
- Gul, Faruk and Wolfgang Pesendorfer**, “Hurwicz expected utility and subjective sources,” *Journal of Economic Theory*, 2015, 59, 465–488.
- Guo, Yingni and Eran Shmaya**, “The Interval Structure of Optimal Disclosure,” *Econometrica*, 2019, 87 (2), 653–675.
- Hu, Ju and Xi Weng**, “Robust Persuasion of a Privately Informed Receiver,” *Economic Theory*, 2021, pp. 909–953.
- Hurwicz, Leonid**, “Optimality Criteria for Decision Making Under Ignorance,” *Cowles Commission Discussion Paper*, 1951.
- Inostroza, Nicolas and Alessandro Pavan**, “Adversarial Coordination and Public Information Design,” *Working Paper, Northwestern University*, 2022.
- Kamenica, Emir**, “Bayesian Persuasion and Information Design,” *Annual Review of Economics*, 2019, 11 (1), 249–272.
- and **Matthew Gentzkow**, “Bayesian Persuasion,” *American Economic Review*, 2011, 101, 2590–2615.
- Ke, Shaowei and Qi Zhang**, “Randomization and Ambiguity Aversion,” *Econometrica*, 2020, 88 (3), 1159–1195.
- Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li**, “Persuasion of a Privately Informed Receiver,” *Econometrica*, 2017, 85, 1949–1964.

- Kosterina, Svetlana**, “Persuasion with Unknown Beliefs,” *forthcoming at Theoretical Economics*, 2021.
- Laclau, Marie and Ludovic Renou**, “Public Persuasion,” *Working Paper, Paris School of Economics*, 2017.
- Li, Fei and Peter Norman**, “Sequential persuasion,” *Theoretical Economics*, 2021, 16 (2), 639–675.
- , **Yangbo Song, and Mofei Zhao**, “Global Manipulation by Local Obfuscation,” *Working Paper*, 2021.
- Lipnowski, Elliot, Laurent Mathevet, and Dong Wei**, “Attention Management,” *American Economic Review: Insights*, 2020, 2 (1), 17–32.
- Mathevet, Laurent, Jacopo Perego, and Ina Taneva**, “On Information Design in Games,” *Journal of Political Economy*, 2020, 128 (4), 1370–1404.
- Matysková, Ludmila and Alfonso Montes**, “Bayesian Persuasion With Costly Information Acquisition,” *Working Paper, Bonn University*, 2021.
- Morris, Stephen, Daisuke Oyama, and Satoru Takahashi**, “Implementation via Information Design in Binary-Action Supermodular Games,” *Working Paper, MIT*, 2020.
- Nikzad, Afshin**, “Persuading a pessimist: Simplicity and robustness,” *Games and Economic Behavior*, 2021, 129, 144–157.
- Wu, Wenhao**, “Sequential Bayesian Persuasion,” *Working Paper*, 2021.
- Ye, Lintao**, “Beneficial Persuasion,” *Working Paper, Washington University in St. Louis*, 2021.
- Ziegler, Gabriel**, “Adversarial Bilateral Information Design,” *Working Paper, Northwestern University*, 2020.

A Appendix

A.1 Proof of Lemma 1

Fix the Sender’s signal q . For any realized $s \in \text{supp}(q)$, Nature’s problem of minimizing the Sender’s payoff is

$$- \sup_{\pi: \Omega \times \{s\} \rightarrow \Delta \mathcal{R}} \sum_{\omega \in \Omega, r \in \mathcal{R}} -\underline{V}(\mu_0^{s,r}) \pi(r|\omega, s) \mu_0^s(\omega). \quad (\text{A.1})$$

The optimization problem (A.1) is a Bayesian-persuasion problem with a finite state space and an upper semi-continuous objective function (because \underline{V} is lower semi-continuous). By [Kamenica and Gentzkow \(2011\)](#), it is without loss of generality to

restrict attention to π with $|\text{supp}(\pi)| = |\Omega|$, the supremum is attained, and the value of the problem is given by the negative of the concave closure of $-\underline{V}$, evaluated at μ_0^s .

It is immediate (see also [Hu and Weng, 2021](#)) that worst-case optimality of q is equivalent to generating the full-disclosure payoff $\underline{V}_{\text{full}}(\mu_0)$ in the worst-case scenario. Indeed, $\underline{V}_{\text{full}}(\mu_0)$ is an upper bound because Nature can always disclose the state; but this upper bound can be achieved by the Sender if she fully discloses the state herself. Using this observation and the definition of \underline{V} , we have that a signal q is worst-case optimal if and only if

$$\sum_{\omega \in \Omega, s \in \mathcal{S}} \underline{V}(\mu_0^s) q(s|\omega) \mu_0(\omega) = \underline{V}_{\text{full}}(\mu_0), \quad (\text{A.2})$$

and, moreover, $\underline{V} = -\text{co}(-\underline{V})$. A distribution ρ of posterior beliefs can be induced by some signal $q : \Omega \rightarrow \Delta \mathcal{S}$ if and only if ρ satisfies [\(BP\)](#). We conclude that a signal q satisfies [\(A.2\)](#) if and only if the distribution of posterior beliefs ρ_q that it induces satisfies [\(WC\)](#) and [\(BP\)](#).

A.2 Proof of Proposition 1

Let $\mathcal{X} = \{\rho \in \Delta \Delta \Omega : \rho \text{ satisfies } \text{[\(BP\)](#) and } \text{supp}(\rho) \subseteq \Delta_{\mathcal{F}} \Omega\}$, where $\Delta_{\mathcal{F}} \Omega := \{\mu \in \Delta \Omega : \text{supp}(\mu) \in \mathcal{F}\}$. It is enough to prove that $\mathcal{W} = \mathcal{X}$.²³

Proof of $\mathcal{W} \subseteq \mathcal{X}$: Let $\rho \in \mathcal{W}$. By definition of \mathcal{W} , ρ satisfies [\(BP\)](#). We will show that $\text{supp}(\rho) \subseteq \Delta_{\mathcal{F}} \Omega$. Suppose not. Then, there exists $A \subseteq \text{supp}(\rho)$, with $\rho(A) > 0$, such that for any $\mu \in A$, $\text{supp}(\mu) \notin \mathcal{F}$. That is, given μ , there exists $\eta \in \Delta \Omega$ with $\text{supp}(\eta) \subseteq \text{supp}(\mu)$ such that $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$. Recall that $\text{lco}(\underline{V})$ denotes the lower convex closure of \underline{V} , and that $\underline{V} = \text{lco}(\underline{V})$. Because $\text{lco}(\underline{V}) \leq \underline{V}$, we have that $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$. Because $\text{supp}(\eta) \subseteq \text{supp}(\mu)$, there exists a small enough $\lambda > 0$ such that $\mu = \lambda \eta + (1 - \lambda) \eta'$, for some $\eta' \in \Delta \Omega$. We have

$$\begin{aligned} \underline{V}(\mu) &= \underline{V}(\lambda \eta + (1 - \lambda) \eta') \leq \lambda \underline{V}(\eta) + (1 - \lambda) \underline{V}(\eta') \\ &< \lambda \underline{V}_{\text{full}}(\eta) + (1 - \lambda) \underline{V}_{\text{full}}(\eta') = \underline{V}_{\text{full}}(\mu), \end{aligned} \quad (\text{A.3})$$

where the first inequality follows from the convexity of \underline{V} , the second (strict) inequality from the fact that $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$ and $\underline{V} \leq \underline{V}_{\text{full}}$, and the final equality from the fact that $\underline{V}_{\text{full}}$ is affine.

We are ready to obtain a contradiction. Recall from [Lemma 1](#) that since ρ is a

²³The proof below works for arbitrary distributions, even though it would suffice for our purposes to prove the equivalence for finite-support distributions, given the assumption of finite signal spaces.

worst-case optimal distribution, it must satisfy $\int \underline{V}(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0)$ which, by (BP) and the fact that $\underline{V}_{\text{full}}$ is affine, can also be written as

$$\int [\underline{V}(\mu) - \underline{V}_{\text{full}}(\mu)] d\rho(\mu) = 0. \quad (\text{A.4})$$

Because $\underline{V} \leq \underline{V}_{\text{full}}$, we must have $\underline{V}(\mu) = \underline{V}_{\text{full}}(\mu)$ for ρ -almost all μ , contradicting (A.3) which holds for a ρ -positive-measure set A of posteriors μ .

Proof of $\mathcal{W} \supseteq \mathcal{X}$: Suppose that $\rho \in \mathcal{X}$. It suffices to show that ρ satisfies (WC). Because $\text{supp}(\rho) \subseteq \Delta_{\mathcal{F}}\Omega$, we have that, for any $\mu \in \text{supp}(\rho)$, $\underline{V}|_{\Delta_{\text{supp}}(\mu)} \geq \underline{V}_{\text{full}}|_{\Delta_{\text{supp}}(\mu)}$. Because \underline{V} dominates an affine function $\underline{V}_{\text{full}}$ on $\Delta_{\text{supp}}(\mu)$, so does its lower convex closure \underline{V} . We conclude that $\underline{V}(\mu) \geq \underline{V}_{\text{full}}(\mu)$ for all $\mu \in \text{supp}(\rho)$. Because disclosing the state is always possible for Nature, $\underline{V}(\mu) = \underline{V}_{\text{full}}(\mu)$ for all $\mu \in \text{supp}(\rho)$. Together with the fact that $\underline{V}_{\text{full}}$ is affine, this implies that ρ satisfies (WC).

A.3 Proof of Theorem 2

We state three lemmas that imply Theorem 2, and then prove them. We will use the observation that, for any $\mu \in \Delta\Omega$, and any $A \subseteq \text{supp}(\mu)$,

$$d(\mu, \mu|_A) = \sum_{\omega \in \text{supp}(\mu)} |\mu(\omega) - \mu|_A(\omega)| = \sum_{\omega \in A} \left(\frac{\mu(\omega)}{\mu(A)} - \mu(\omega) \right) + \sum_{\omega \notin A} \mu(\omega) = 2\mu(A^c), \quad (\text{A.5})$$

where A^c denotes the complement of the set A (in Ω). Abusing notation slightly, we define, for any $\mu \in \Delta\Omega$, the distance between μ and the set of admissible beliefs according to \mathcal{F} as

$$d(\mu, \Delta_{\mathcal{F}}\Omega) := \inf_{\eta \in \Delta_{\mathcal{F}}\Omega} d(\mu, \eta).$$

We first observe that if the Sender decides to induce a belief μ that does not belong to $\Delta_{\mathcal{F}}\Omega$, then the loss that she incurs in the worst-case scenario (relative to a worst-case optimal policy) is increasing in the distance between μ and beliefs in the admissible set $\Delta_{\mathcal{F}}\Omega$.

Lemma 6. *There exists a constant $\delta > 0$ such that, for any $\mu \notin \Delta_{\mathcal{F}}\Omega$,*

$$\underline{V}_{\text{full}}(\mu) - \underline{V}(\mu) \geq \delta \cdot d(\mu, \Delta_{\mathcal{F}}\Omega).$$

For regular functions, we can correspondingly bound from above the gains from inducing a belief $\mu \notin \Delta_{\mathcal{F}}\Omega$ in the base-case scenario. The Sender can always achieve $\text{co}(\widehat{V}_{\mathcal{F}})(\mu)$ without sacrificing worst-case optimality, by Corollary 7. For $\mu \notin \Delta_{\mathcal{F}}\Omega$, it

is possible that $\widehat{V}(\mu) > \text{co}(\widehat{V}_{\mathcal{F}})(\mu)$ but the difference can be upper bounded, and is small when μ is close to $\Delta_{\mathcal{F}}\Omega$.

Lemma 7. *For a regular function \widehat{V} , there exists $\Delta > 0$ such that for any $\mu \notin \Delta_{\mathcal{F}}\Omega$,*

$$\widehat{V}(\mu) - \text{co}(\widehat{V}_{\mathcal{F}})(\mu) \leq \Delta \cdot d(\mu, \Delta_{\mathcal{F}}\Omega).$$

Together, the above two lemmas imply the following result:

Lemma 8. *Suppose that \widehat{V} is regular. There exists $\bar{\lambda} < 1$ such that, for all $\lambda \in (\bar{\lambda}, 1]$, if ρ solves problem (6.1), then ρ must assign probability one to $\Delta_{\mathcal{F}}\Omega$.*

Theorem 2 follows from Lemma 8. Indeed, because, for high λ , any λ -solution assigns probability one to beliefs in $\Delta_{\mathcal{F}}\Omega$, any λ -solution delivers the same expected payoff to the Sender in the worst-case scenario (namely, the full-disclosure payoff). As long as the weight $1 - \lambda$ on the base-case scenario is strictly positive, a λ -solution must then maximize the Sender's payoff in the base-case scenario, conditional on being worst-case optimal, that is, it must be a robust solution.²⁴

A.3.1 Proof of Lemma 6

For any $B \subseteq \Omega$, with $B \notin \mathcal{F}$, fix an arbitrary $\mu_B \in \Delta\Omega$ with $\text{supp}(\mu_B) \subseteq B$ such that $\underline{V}(\mu_B) < \underline{V}_{\text{full}}(\mu_B)$, and hence $\underline{V}(\mu_B) = \text{lco}(\underline{V})(\mu_B) < \underline{V}_{\text{full}}(\mu_B)$. Then, let $\delta_B := \underline{V}_{\text{full}}(\mu_B) - \underline{V}(\mu_B)$ and $\delta^* := \min_{B \notin \mathcal{F}} \delta_B > 0$.

Consider any $\mu \notin \Delta_{\mathcal{F}}\Omega$. Let $B \subseteq \text{supp}(\mu)$ be such that $B \notin \mathcal{F}$. We can write $\mu = \kappa\mu_B + (1 - \kappa)\mu'$ for some μ' and κ , as long as $\mu(\omega) \geq \kappa\mu_B(\omega)$ for all $\omega \in \text{supp}(\mu)$ —this equality can be written in particular for $\kappa = \min_{\omega \in B} \{\mu(\omega)\}$. Because $\underline{V}_{\text{full}} - \underline{V}$ is a non-negative and concave function (concavity follows from the fact that it is the difference between an affine function and a convex function), we have that

$$\begin{aligned} (\underline{V}_{\text{full}} - \underline{V})(\mu) &= (\underline{V}_{\text{full}} - \underline{V})(\kappa\mu_B + (1 - \kappa)\mu') \geq \\ \kappa(\underline{V}_{\text{full}} - \underline{V})(\mu_B) + (1 - \kappa)(\underline{V}_{\text{full}} - \underline{V})(\mu') &\geq \delta_B \cdot \min_{\omega \in B} \{\mu(\omega)\} \geq \delta^* \cdot \min_{\omega \in B} \{\mu(\omega)\}. \end{aligned}$$

We can order that states in $\text{supp}(\mu) = \{\omega_1, \dots, \omega_k\}$ so that $\mu(\omega_1) \geq \mu(\omega_2) \geq \dots \geq \mu(\omega_k)$, and then find the highest $i < k$ such that $A = \{\omega_1, \omega_2, \dots, \omega_i\} \in \mathcal{F}$ (such i exists since $\{\omega_1\} \in \mathcal{F}$ but $\{\omega_1, \dots, \omega_k\} \notin \mathcal{F}$). Then, setting $B = A \cup \{\omega_{i+1}\} \notin \mathcal{F}$, we

²⁴Formally, for $\lambda \in (\bar{\lambda}, 1)$, ρ concavifies $\lambda\underline{V} + (1 - \lambda)\widehat{V}$ at μ_0 if and only if it concavifies \widehat{V} at μ_0 on $\Delta_{\mathcal{F}}\Omega$. This, however, is equivalent to concavifying $\widehat{V}_{\mathcal{F}}$ at μ_0 . By virtue of Corollary 7, ρ is thus a robust solution.

have that

$$\delta^* \cdot \min_{\omega \in A \cup \{\omega_{i+1}\}} \{\mu(\omega)\} = \delta^* \cdot \mu(\omega_{i+1}) \geq \frac{\delta^*}{|\Omega|} \cdot \mu(A^c) = \frac{\delta^*}{|\Omega|} \frac{1}{2} \cdot d(\mu, \mu|_A),$$

where the last step follows from equation (A.5). Putting the above inequalities together, and using the fact that $\mu|_A \in \Delta_{\mathcal{F}}\Omega$, we obtain

$$(\underline{V}_{\text{full}} - \underline{V})(\mu) \geq \frac{\delta^*}{|\Omega|} \frac{1}{2} \cdot d(\mu, \mu|_A) \geq \frac{\delta^*}{|\Omega|} \frac{1}{2} \cdot \inf_{\eta \in \Delta_{\mathcal{F}}\Omega} d(\mu, \eta) = \frac{\delta^*}{|\Omega|} \frac{1}{2} \cdot d(\mu, \Delta_{\mathcal{F}}\Omega),$$

which finishes the proof if we set $\delta := \frac{\delta^*}{|\Omega|} \frac{1}{2}$.

A.3.2 Proof of Lemma 7

Fix $\mu \notin \Delta_{\mathcal{F}}\Omega$. We first claim that for any $A \subseteq \text{supp}(\mu)$, we can write

$$\mu = \kappa \mu|_A + (1 - \kappa) \eta$$

for some $\eta \in \Delta\Omega$, and $\kappa = 1 - \frac{1}{2}d(\mu, \mu|_A)$. Indeed, since $\mu|_A = \mu(\omega)/\mu(A)$ for every $\omega \in A$, we can take $\kappa = \mu(A)$; but then equation (A.5) implies that

$$\kappa = 1 - \mu(A^c) = 1 - \frac{1}{2}d(\mu, \mu|_A).$$

We will use this fact for a particular set A . Namely, let $A \subseteq \text{supp}(\mu)$ be such that $A \in \mathcal{F}$ and $d(\mu, \Delta_{\mathcal{F}}\Omega) = d(\mu, \Delta A)$.²⁵ Using the fact that a concave closure is a concave function, we have that

$$\begin{aligned} \text{co}(\widehat{V}_{\mathcal{F}})(\mu) &= \text{co}(\widehat{V}_{\mathcal{F}})(\kappa \mu|_A + (1 - \kappa) \eta) \geq \kappa \text{co}(\widehat{V}_{\mathcal{F}})(\mu|_A) + (1 - \kappa) \text{co}(\widehat{V}_{\mathcal{F}})(\eta) \\ &= \text{co}(\widehat{V}_{\mathcal{F}})(\mu|_A) + (1 - \kappa) [\text{co}(\widehat{V}_{\mathcal{F}})(\eta) - \text{co}(\widehat{V}_{\mathcal{F}})(\mu|_A)] \geq \text{co}(\widehat{V}_{\mathcal{F}})(\mu|_A) - (1 - \kappa) 2M, \end{aligned}$$

where $M < \infty$ is an upper bound on the norm of $\text{co}(\widehat{V}_{\mathcal{F}})$ (which exists since \widehat{V} was assumed to be bounded). Moreover, since $\kappa = 1 - \frac{1}{2}d(\mu, \mu|_A)$, we obtain that

$$\text{co}(\widehat{V}_{\mathcal{F}})(\mu) \geq \text{co}(\widehat{V}_{\mathcal{F}})(\mu|_A) - M d(\mu, \mu|_A).$$

Since $\mu|_A \in \Delta_{\mathcal{F}}\Omega$, we have that $\text{co}(\widehat{V}_{\mathcal{F}})(\mu|_A) = \text{co}(\widehat{V})(\mu|_A)$. Using the assumption that \widehat{V} is regular, we can conclude that

$$\text{co}(\widehat{V}_{\mathcal{F}})(\mu) \geq \text{co}(\widehat{V})(\mu|_A) - M d(\mu, \mu|_A) \geq \widehat{V}(\mu) - (M + L) d(\mu, \mu|_A).$$

²⁵To see why such a set A exists, note that the infimum in the definition $d(\mu, \Delta_{\mathcal{F}}\Omega) := \inf_{\eta \in \Delta_{\mathcal{F}}\Omega} d(\mu, \eta)$ is attained at some feasible η^* ($\Delta_{\mathcal{F}}\Omega$ is a compact set, and d is a continuous function), and we can take $A := \text{supp}(\eta^*)$. Moreover, $\text{supp}(\eta^*) \subset \text{supp}(\mu)$ because otherwise we could decrease the distance between η^* and μ by shifting mass under η^* from some $\omega \in \text{supp}(\eta^*) \setminus \text{supp}(\mu)$ to some ω' such that $\mu(\omega') > \eta^*(\omega')$, contradicting the definition of η^* .

Finally, using equation (A.5) and our definition of the set A , we obtain that

$$d(\mu, \mu|_A) = 2\mu(A^c) \leq 2d(\mu, \Delta A) = 2d(\mu, \Delta_{\mathcal{F}}\Omega),$$

which establishes the desired conclusion if we set $\Delta := 2(M + L)$.

A.3.3 Proof of Lemma 8

It is enough to prove that, for high enough λ , if $\text{supp}(\rho) \not\subseteq \Delta_{\mathcal{F}}\Omega$, then the Sender's objective $\int \left[\lambda \underline{V}(\mu) + (1 - \lambda) \widehat{V}(\mu) \right] d\rho(\mu)$ increases strictly by splitting any $\mu \in \text{supp}(\rho)$ such that $\mu \notin \Delta_{\mathcal{F}}\Omega$ into beliefs that yield $\text{co}(\widehat{V}_{\mathcal{F}})(\mu)$ —such a split is always available to the Sender and, by definition of $\text{co}(\widehat{V}_{\mathcal{F}})$, yields the payoff $\underline{V}_{\text{full}}(\mu)$ in the worst-case scenario. By Lemma 6 and 7, we have that, for some $\Delta > 0$ and $\delta > 0$,

$$\begin{aligned} & \left[\lambda \underline{V}_{\text{full}}(\mu) + (1 - \lambda) \text{co}(\widehat{V}_{\mathcal{F}})(\mu) \right] - \left[\lambda \underline{V}(\mu) + (1 - \lambda) \widehat{V}(\mu) \right] \\ &= \lambda \left[\underline{V}_{\text{full}}(\mu) - \underline{V}(\mu) \right] + (1 - \lambda) \left[\text{co}(\widehat{V}_{\mathcal{F}})(\mu) - \widehat{V}(\mu) \right] \\ &\geq (\lambda\delta - (1 - \lambda)\Delta) \cdot d(\mu, \Delta_{\mathcal{F}}\Omega) > 0 \end{aligned}$$

if $\lambda > \bar{\lambda}$ where $\bar{\lambda} := \frac{\Delta}{\Delta + \delta} < 1$.

A.4 Proof of Theorem 3

Part (a). Let S^* be the set of robust solutions, represented as distributions of posterior beliefs. This set is non-empty and closed (by Berge's theorem), hence compact in the weak* topology. Note that if an element ρ^* of S^* is dominated, it must be dominated by another element of S^* . Indeed, a policy that is not a robust solution cannot dominate ρ^* because, by definition, it either yields a strictly lower payoff when Nature responds to each μ with $\underline{V}(\mu)$, or, it yields a strictly lower payoff when Nature responds to each μ with $\widehat{V}(\mu)$ (by assumption, $V = \underline{V}$ and $V = \widehat{V}$ are both feasible choices by Nature).

Let \mathcal{P} be the set of all feasible functions V that are additionally upper semi-continuous. By Zermelo's theorem, every set can be well-ordered. We thus introduce a well-order \sqsubset on \mathcal{P} . For any $V \in \mathcal{P}$, let $B^*(V) \subset S^*$ be the subset of S^* constructed inductively as follows. Let V_0 be the lowest element of \mathcal{P} according to the order \sqsubset . Then, let

$$B^*(V_0) := \operatorname{argmax}_{\rho \in S^*} \left\{ \int V_0(\mu) d\rho(\mu) \right\},$$

that is, the subset of robust solutions that are optimal for the Sender against V_0 .

The set $B^*(V_0)$ is non-empty and closed (and hence compact in the weak* topology) because V_0 is upper semi-continuous and S^* is non-empty and compact. For any $V \in \mathcal{P}$, then let

$$B(V) := \bigcap_{V' \sqsubset V} B^*(V'),$$

$$B^*(V) := \operatorname{argmax}_{\rho \in B(V)} \left\{ \int V(\mu) d\rho(\mu) \right\}.$$

The sets $B^*(V)$ are nested, in the sense that $B^*(V') \subseteq B^*(V)$ if $V \sqsubset V'$. There are also non-empty and compact (again by Berge's theorem). By an application of the Finite Intersection Axiom, we can conclude that $\bigcap_{V \in \mathcal{P}} B^*(V) \neq \emptyset$ and any $\rho^* \in \bigcap_{V \in \mathcal{P}} B^*(V)$ is an undominated robust solution when we restrict attention to functions V that are upper semi-continuous.

To finish the proof, suppose that such a ρ^* is dominated. Then, it must yield the Sender a payoff strictly lower than the one achieved by another robust solution ρ' when Nature responds with a feasible V that is not upper semi-continuous. However, any measurable V can be approximated point-wise by a sequence V_n of upper semi-continuous functions. By Lebesgue's dominated convergence theorem, the Sender's expected payoff differential between ρ^* and ρ' under V_n must converge to her expected payoff differential under the limit function V . If the expected payoff differential under the limit function V is strictly negative, the expected payoff differential must also be negative under V_n , for large n , contradicting the fact that ρ^* is undominated when Nature responds with upper semi-continuous functions, as shown above.

Part (b). We now establish that, when $\operatorname{co}\widehat{V} = \overline{V}$, any robust solution is undominated. Pick any robust solution ρ' . Again, it suffices to show that ρ' is not dominated by any other robust solution ρ . By Corollary 7, any robust solution achieves $\operatorname{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$ under the conjecture, which corresponds to Nature selecting $V = \widehat{V}$. Suppose first that there exists $\mu \in \Delta\Omega$ such that $\operatorname{co}(\widehat{V}_{\mathcal{F}})(\mu) > \underline{V}(\mu)$ and $\rho'(\mu) \neq \rho(\mu)$. There are two subcases: Either (a) $\rho'(\mu) > \rho(\mu)$ or (b) $\rho'(\mu) < \rho(\mu)$.

In case (a), consider the feasible response by Nature V that responds to μ according to the Sender's conjecture, and that responds adversarially to any other posterior: $V(\mu) = \widehat{V}(\mu)$, and $V(\mu') = \underline{V}(\mu')$ for all $\mu' \neq \mu$. Because μ is induced under some robust solution (that is, $\mu \in \operatorname{supp}(\rho') \cup \operatorname{supp}(\rho)$), by Corollary 7, it must be that $\widehat{V}(\mu) = \operatorname{co}(\widehat{V}_{\mathcal{F}})(\mu)$. Thus, under the specified response by Nature, the Sender's ex-

pected payoff under a robust solution $\rho^* \in \{\rho, \rho'\}$ is given by

$$\rho^*(\mu)\text{co}(\widehat{V}_{\mathcal{F}})(\mu) + \left(\int \underline{V}(\mu')d\rho^*(\mu') - \underline{V}(\mu)\rho^*(\mu) \right).$$

Under a robust solution, by Lemma 1, we have that $\int \underline{V}(\mu')d\rho^*(\mu') = \underline{V}_{\text{full}}(\mu_0)$, and thus the difference in expected payoffs between ρ' and ρ when Nature responds with V is given by

$$[\rho'(\mu) - \rho(\mu)] \left[\text{co}(\widehat{V}_{\mathcal{F}})(\mu) - \underline{V}(\mu) \right] > 0,$$

where the inequality follows from the fact that $\rho'(\mu) > \rho(\mu)$. Thus, ρ does not dominate ρ' .

In case (b), consider the following response by Nature: $V(\mu) = \underline{V}(\mu)$ and $V(\mu') = \widehat{V}(\mu')$ for all $\mu' \neq \mu$. Under this response by Nature, the expected payoff under a robust solution $\rho^* \in \{\rho', \rho\}$ is equal to

$$\text{co}(\widehat{V}_{\mathcal{F}})(\mu_0) - \rho^*(\mu)[\text{co}(\widehat{V}_{\mathcal{F}})(\mu) - \underline{V}(\mu)].$$

To see this, recall that when Nature responds to any induced posterior with \widehat{V} , then ρ^* generates an expected payoff equal to $\text{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$ —this follows directly from the fact that ρ^* is a robust solution.²⁶ Conditional on inducing μ (which has probability $\rho^*(\mu)$), instead of $\widehat{V}(\mu) = \text{co}(\widehat{V}_{\mathcal{F}})(\mu)$, the Sender gets $\underline{V}(\mu)$.

Thus, the difference in expected payoffs between ρ' and ρ is given by

$$[\rho(\mu) - \rho'(\mu)] \left[\text{co}(\widehat{V}_{\mathcal{F}})(\mu) - \underline{V}(\mu) \right] > 0,$$

because $\rho(\mu) > \rho'(\mu)$. Hence, ρ does not dominate ρ' also in this case.

The final case to consider is when there exists no $\mu \in \Delta\Omega$ such that $\text{co}(\widehat{V}_{\mathcal{F}})(\mu) > \underline{V}(\mu)$ and $\rho'(\mu) \neq \rho(\mu)$. Put differently, for any μ such that $\rho'(\mu) \neq \rho(\mu)$ (such a μ must exist because otherwise the two solutions would coincide), we must have $\text{co}(\widehat{V}_{\mathcal{F}})(\mu) = \underline{V}(\mu)$ (since $\text{co}(\widehat{V}_{\mathcal{F}}) \geq \underline{V}$). Note, however, that $\text{co}(\widehat{V}_{\mathcal{F}})$ is a concave function while \underline{V} is a convex function, and thus they can be equal at μ if and only if they are both affine functions on $\Delta\text{supp}(\mu)$: In fact, we must have $\widehat{V} = \underline{V} = \underline{V}_{\text{full}}$ on $\Delta\text{supp}(\mu)$. Moreover, because \widehat{V} is affine on $\Delta\text{supp}(\mu)$, we have that $\text{co}\widehat{V}(\mu) = \widehat{V}(\mu)$ for any such μ . Finally, using the assumption of Theorem 3 that $\text{co}\widehat{V} = \bar{V}$, we conclude that $\bar{V} = \underline{V}$ on $\Delta\text{supp}(\mu)$. But this means that *any* V that Nature can select is affine on $\Delta\text{supp}(\mu)$. This implies that Nature's response conditional on any

²⁶In fact, from Corollary 7, $\int \widehat{V}_{\mathcal{F}}(\mu)d\rho^*(\mu) = \text{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$. The property then follows from the fact that, for any $\mu' \in \text{supp}(\rho^*)$, $\widehat{V}(\mu') = \widehat{V}_{\mathcal{F}}(\mu')$.

such μ is payoff-equivalent for the Sender. Because this is true for any μ at which ρ' and ρ differ, and because both distributions are robust solutions, it follows that these two signals are payoff-equivalent, and hence ρ does not dominate ρ' .

Online Appendix

OA.1 Proofs for Section 5

OA.1.1 Proof of Lemma 2

Pick any two states ω and ω' such that $\omega > \omega' + D$ and let $B = \{\omega', \omega\}$. To simplify the notation, for any $\lambda \in [0, 1]$, let $v(\lambda) := \underline{V}(\lambda\delta_\omega + (1 - \lambda)\delta_{\omega'})$. It is enough to prove that $v'(0) < v(1) - v(0)$ as this implies that $v(\lambda)$ is strictly below the payoff from full disclosure $\lambda v(1) + (1 - \lambda)v(0)$ for small enough $\lambda > 0$. Indeed, this means that $\underline{V}|_{\Delta B}(\mu)$ is below the full-disclosure payoff $\underline{V}_{\text{full}}|_{\Delta B}(\mu)$ for posterior beliefs μ supported on B that put sufficiently small mass on ω ; the conclusion then follows from Corollary 2. For low enough λ , using the fact that $\omega > \omega' + D$, we have $v(\lambda) = (1 - \lambda) \left(\int_{\omega'}^{\omega'+D} (p - \omega') dp \right)$. That is, only the low type ω' trades if the buyer believes the seller's type to be ω' with high probability. We thus have $v'(0) = - \int_{\omega'}^{\omega'+D} (p - \omega') dp$, so that $v'(0) - v(1) + v(0) = - \int_{\omega}^{\min\{\omega+D, 1\}} (p - \omega) dp < 0$ by the assumption that $\max \Omega < 1$.

OA.1.2 Proof of Lemma 3

Clearly, $\mathbf{1}_{\{\mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega} \leq p] + D > p\}} \leq \mathbf{1}_{\{\mathbb{E}_\mu[\tilde{\omega}] + D > p\}}$. Suppose that the inequality is strict for some $p \geq \underline{\omega}_\mu : \mathbb{E}_\mu[\tilde{\omega}] + D > p$ but $\mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega} \leq p] + D \leq p$. This is only possible when $p < \bar{\omega}_\mu$, where $\bar{\omega}_\mu$ is the maximum of $\text{supp}(\mu)$. But then

$$p \geq \mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega} \leq p] + D \geq \underline{\omega}_\mu + D \geq (\bar{\omega}_\mu - D) + D = \bar{\omega}_\mu > p,$$

a contradiction.

OA.1.3 Proof of Lemma 4

By Lemma 3, we can write

$$\underline{V}(\mu) = \sum_{\omega \in \text{supp}(\mu)} \left(\int_{\omega}^{\mathbb{E}_\mu[\tilde{\omega}] + D} (p - \omega) dp \right) \mu(\omega) = \frac{1}{2} \sum_{\omega \in \text{supp}(\mu)} (\mathbb{E}_\mu[\tilde{\omega}] + D - \omega)^2 \mu(\omega).$$

Let $B = \{\omega_1, \dots, \omega_n\}$ with $\omega_1 < \omega_2 < \dots < \omega_n$, and let $\mu_i = \mu(\omega_i)$. Then, \underline{V} can be treated as a function defined on a unit simplex in \mathbb{R}^n :

$$\underline{V}(\mu) = \frac{1}{2} \sum_{i=1}^n \mu_i \left(\sum_{j=1}^n \mu_j \omega_j + D - \omega_i \right)^2.$$

To prove the lemma, it is enough to prove that the function \tilde{V} defined by $\tilde{V}(\mu_2, \dots, \mu_n) = V(1 - \mu_2 - \dots - \mu_n, \mu_2, \dots, \mu_n)$ has a negative semi-definite hessian. By a direct calculation, denoting $\omega_{-1} = [\omega_2, \dots, \omega_n]$, we obtain that $\text{Hessian}(\tilde{V}) = -(\omega_{-1} - \omega_1)^T \cdot (\omega_{-1} - \omega_1)$, which is a negative semi-definite matrix (of rank 1).

OA.1.4 Proof of Proposition 4

Given any $\mu \in \Delta\Omega$, let $\mu^+ := \mu(\omega > 0)$ denote the probability that μ assigns to the event that $\omega > 0$. In this application, the Sender's conjecture is that the Receivers do not have any exogenous information other than the one contained in the prior μ_0 . Furthermore, for any common posterior μ , all agents attack if $\mu^+ < \alpha$, and refrain from attacking if $\mu^+ \geq \alpha$, where $\alpha := g/(g + |b|)$, implying that $\hat{V}(\mu) = 0$ if $\mu^+ < \alpha$ and $\hat{V}(\mu) = 1$ if $\mu^+ \geq \alpha$.

Let $\mu_0^+ < \alpha$, as assumed in the main text. The following policy is then a Bayesian solution. The Sender randomizes over two announcements, $s = 0$ and $s = 1$. She announces $s = 0$ with certainty when $\omega > 0$, and with probability $(1 - \phi_{BP}) \in (0, 1)$ when $\omega \leq 0$, with ϕ_{BP} satisfying $\mathbb{P}(\omega > 0 | s = 0) = \mu_0^+ / [\mu_0^+ + (1 - \mu_0^+)(1 - \phi_{BP})] = \alpha$. To see that this is a Bayesian solution, first note that, without loss of optimality, the Sender can confine attention to policies with two signal realizations, $s = 0$ and $s = 1$, such that, when $s = 0$ is announced, $\mathbb{P}(\omega > 0 | s = 0) \geq \alpha$ and all agents refrain from attacking, whereas when $s = 1$ is announced, $\mathbb{P}(\omega > 0 | s = 1) < \alpha$ and all agents attack.²⁷ Next, note that, starting from any binary policy announcing $s = 1$ with positive probability over a positive measure subset of \mathbb{R}_+ , one can construct another binary policy that announces $s = 0$ (thus inducing all agents to refrain from attacking) with a higher ex-ante probability, contradicting the optimality of the original policy. Hence, any binary Bayesian solution must announce $s = 0$ with certainty for all $\omega > 0$. Furthermore, under any Bayesian solution, the ex-ante probability $\sum_{\omega \in \Omega: \omega < 0} \pi(0 | \omega) \mu_0(\omega)$ is uniquely pinned down by the condition $\mathbb{P}(\omega > 0 | s = 0) = \mu_0^+ / [\mu_0^+ + \sum_{\omega \in \Omega: \omega < 0} \pi(0 | \omega) \mu_0(\omega)] = \alpha$. Because the Sender's preferences depend only on $1 - A$, the specific way the policy announces $s = 0$ when $\omega < 0$ is irrelevant, thus implying that the binary policy described above is indeed a Bayesian

²⁷The arguments for this result are the usual ones. Starting from any policy with more than two signal realizations, one can pool into $s = 0$ all signal realizations leading to a posterior assigning probability at least α to the event that $\omega > 0$ and into $s = 1$ all signal realizations leading to a posterior assigning probability less than α to $\omega > 0$. The binary policy so constructed is payoff-equivalent to the original one.

solution. By the same token, the above binary policy is payoff-equivalent to one that announces $s = 0$ with certainty when $\omega > 0$, whereas, when $\omega < 0$, fully reveals the state with probability ϕ_{BP} , and announces $s = 0$ with the complementary probability. The signal realization $s = 0$ can then be interpreted as the decision not to disclose any information (equivalently, as the “null signal” $s = \emptyset$), as claimed in the proposition.

To see that the above Bayesian policy is not robust, let $\mu^{(0,1]} := \mu(\omega \in (0, 1])$ denote the probability that μ assigns to the interval $(0, 1]$. Recall that, given any posterior μ , if $\mu^+ := \mu(\omega > 0) < \alpha$, the unique rationalizable action is to attack. If $\mu^+ \in [\alpha, \alpha + \mu^{(0,1)}]$, both attacking and not attacking are rationalizable. Finally, if $\mu^+ > \alpha + \mu^{(0,1]}$, the unique rationalizable action is to refrain from attacking. Hence, under the most adversarial selection, $\underline{V}(\mu) = 0$ if $\mu^+ \leq \alpha + \mu^{(0,1]}$, and $\underline{V}(\mu) = 1$ if $\mu^+ > \alpha + \mu^{(0,1]}$. Next, observe that worst-case optimality requires that all states $\omega > 1$ be separated from all states $\omega \leq 1$. Indeed, $\underline{V}_{\text{full}}(\mu) = \mu(\omega > 1) = \mu^+ - \mu^{(0,1]}$ and, given any common posterior μ induced by the Sender, Nature always minimizes the Sender’s payoff by using a signal that discloses the same information to all agents. Arguments similar to those in the judge’s example in Section 3 imply that any worst-case optimal distribution (and hence any robust solution) must separate states $\omega > 1$ from states $\omega \leq 1$.

Because the above restriction is the only one imposed by worst-case optimality, on the restricted domain $\bar{\Omega} := \{\omega \in \Omega : \omega \leq 1\}$, any robust solution must coincide with a Bayesian solution. Let $\phi_{RS} \in (0, 1)$ be implicitly defined by $\mu_0^{(0,1]}/[\mu_0^{(0,1]} + (1 - \mu_0^+)(1 - \phi_{RS})] = \alpha$. Arguments similar to the ones above then imply that the following policy is a Bayesian solution on the restricted domain. When $\omega \in (0, 1]$, the Sender announces $s = 0$ with certainty. When, instead, $\omega \leq 0$, with probability $\phi_{RS} > \phi_{BP}$, the Sender fully reveals the state, and with the complementary probability $1 - \phi_{RS}$, announces $s = 0$. Lastly, observe that, given any posterior μ with $\text{supp}(\mu) \subset (1, \infty)$, the unique rationalizable profile features all agents refraining from attacking. This means that, once the Sender fully separates the states $\omega \leq 1$ from the states $\omega > 1$, she may as well fully reveal the state when the latter is strictly above 1.

Combining all the arguments above together, it is then easy to see that the following policy is a robust solution. When $\omega \leq 0$, with probability $\phi_{RS} \in (0, 1)$, the Sender fully reveals the state, whereas, with the complementary probability $1 - \phi_{RS}$, she announces $s = \emptyset$. When $\omega \in (0, 1]$, the Sender announces $s = \emptyset$ with certainty. Finally, when $\omega > 1$, the Sender fully reveals the state, as claimed in the proposition.

OA.2 Auxiliary results for Section 6

OA.2.1 Relaxing the regularity assumption in Theorem 2

In this appendix, we examine the consequences of relaxing the regularity condition in Theorem 2. One direction of Theorem 2 continues to hold in a slightly weaker form.

Theorem OA.1. *If $\lambda_n \nearrow 1$, and $\rho_n \in S(\lambda_n)$ converges to ρ in the weak* topology as $n \rightarrow \infty$, then ρ is a robust solution.*

Proof. Take ρ_n as in the statement of the theorem. By optimality of ρ_n , the value of the Sender's objective (with weight λ_n) cannot be increased strictly by switching to a robust solution. That is,

$$\int_{\Delta\Omega} \left[(1 - \lambda_n) \widehat{V}(\mu) + \lambda_n \underline{V}(\mu) \right] d\rho_n(\mu) \geq (1 - \lambda_n) \text{co}(\widehat{V}_{\mathcal{F}})(\mu_0) + \lambda_n \underline{V}_{\text{full}}(\mu_0).$$

Lemma 6 and the above inequality jointly imply that there exists $\delta > 0$ such that

$$\int_{\Delta\Omega} \widehat{V}(\mu) d\rho_n(\mu) - \text{co}(\widehat{V}_{\mathcal{F}})(\mu_0) \geq \frac{\lambda_n}{1 - \lambda_n} \cdot \delta \cdot \int_{\Delta_{\mathcal{F}}^c\Omega} d(\mu, \Delta_{\mathcal{F}}\Omega) d\rho_n(\mu). \quad (\text{OA.1})$$

Because the left hand side of the above inequality is bounded, and $\lambda_n/(1 - \lambda_n)$ diverges to infinity, we must have that

$$\int_{\Delta_{\mathcal{F}}^c\Omega} d(\mu, \Delta_{\mathcal{F}}\Omega) d\rho_n(\mu) \rightarrow 0.$$

The function $d(\mu, \Delta_{\mathcal{F}}\Omega)$ is continuous and bounded. By definition of convergence in the weak* topology, we have,

$$\int_{\Delta_{\mathcal{F}}^c\Omega} d(\mu, \Delta_{\mathcal{F}}\Omega) d\rho(\mu) = 0.$$

Because the integrand is strictly positive, we must have that $\text{supp}(\rho) \subseteq \Delta_{\mathcal{F}}\Omega$, and thus ρ is worst-case optimal.

Since the right hand side of inequality (OA.1) is non-negative, we have that

$$\text{co}(\widehat{V}_{\mathcal{F}})(\mu_0) \leq \limsup_n \int_{\Delta\Omega} \widehat{V}(\mu) d\rho_n(\mu) \leq \int_{\Delta\Omega} \widehat{V}(\mu) d\rho(\mu) \leq \text{co}(\widehat{V}_{\mathcal{F}})(\mu_0),$$

where the second inequality comes from upper-semi continuity of \widehat{V} , and the last inequality follows from the fact that ρ is worst-case optimal, while $\text{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$ is the upper bound on the conjectured payoff that a worst-case optimal distribution can yield. This, however, means that $\int_{\Delta\Omega} \widehat{V}(\mu) d\rho(\mu) = \text{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$, and thus ρ is a robust solution, by Corollary 7. \square

Next, we show that, without the regularity condition (Definition 4), there exist robust solutions that cannot be approximated by λ -solutions.

Example OA.1. Let $\Omega = \{1, 2, 3\}$, and $\mu_0 = (1/3, 1/3, 1/3)$. Let \underline{V} be equal to 0 everywhere except at $\mu = \mu_0$ where $\underline{V}(\mu_0) = -1$. Let \widehat{V} be such that

$$\widehat{V}(1, 0, 0) = \widehat{V}(0, 1, 0) = \widehat{V}(0, 0, 1) = \widehat{V}(1/2, 1/2, 0) = \widehat{V}(1/2, 0, 1/2) = 0,$$

and

$$\widehat{V}(1 - 2x, x, x) = \sqrt{x}, \forall x \leq 1/3,$$

and $\widehat{V}(\mu) = -1$ anywhere else. Notice that \widehat{V} violates regularity because along the line segment $(1 - 2x, x, x)$, as $x \rightarrow 0$, \widehat{V} decreases at an infinite rate to 0, while $\widehat{V}(\mu) \leq 0$ for all μ that do not have full support.

By definition of \underline{V} , and Proposition 1, any worst-case optimal solution puts no mass on beliefs with full-support. Thus, a robust solution is any Bayes-plausible convex combination of beliefs μ at which $\widehat{V}(\mu) = 0$. However, we will show that in the limit as $\lambda \nearrow 1$, all λ -solutions must put positive (bounded away from zero) mass on the belief $\mu = (1, 0, 0)$. Therefore, the distribution ρ_{RS} that puts mass 1/3 on $\mu = (1/2, 1/2, 0)$, mass 1/3 on $\mu = (1/2, 0, 1/2)$, mass 1/6 on $\mu = (0, 1, 0)$, and mass 1/6 on $\mu = (0, 0, 1)$ is a robust solution but is not a limit of λ -solutions.

Note first that $\underline{V}(\mu) := \text{lco}(\underline{V})(\mu) = -3 \min_{\omega} \mu(\omega)$. Consider a distribution ρ that attaches weight m (potentially $m = 0$) to beliefs of the form $(1 - 2x, x, x)$ for $x \in (0, 1/3]$. Because the objective function $\widehat{V}_{\lambda}(\mu) := \lambda \underline{V}(\mu) + (1 - \lambda) \widehat{V}(\mu)$ is strictly concave on that line segment, a λ -solution attaches the entire weight m to a single x^* . For a fixed λ , the optimal choice of x^* is

$$x^* = \left(\frac{1 - \lambda}{6\lambda} \right)^2.$$

The remaining mass $1 - m$ must be distributed over the beliefs $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1/2, 1/2, 0)$, and $(1/2, 0, 1/2)$, with weights satisfying the Bayes-plausibility constraint. Because the Sender's payoff is equal to 0 on any such belief, a λ -solution is characterized by the level of m that maximizes

$$(1 - m)[0] + m[-3\lambda x^* + (1 - \lambda)\sqrt{x^*}] = m \frac{(1 - \lambda)^2}{12\lambda}$$

subject to the Bayes-plausibility constraint. Because the above function is increasing in m , any λ -solution, for $\lambda < 1$, attaches probability m^* to the belief $(1 - 2x^*, x^*, x^*)$, where $m^* \geq 1/3$ is the largest value of m consistent with Bayes plausibility. Next

observe that $(1 - 2x^*, x^*, x^*)$ converges to $(1, 0, 0)$ as $\lambda \nearrow 1$. Hence, all limits of λ -solutions put at least $1/3$ mass on $(1, 0, 0)$ which is what we wanted to prove.

OA.2.2 Example showing that Bayesian solutions can be dominated

In this subsection, we construct an example showing that a Bayesian solution can be dominated even under the assumption made in case (b) of Theorem 3.

Consider the following conjecture \widehat{V} (equal to \underline{V}) defined over the set $[0, 1]$ of posteriors over a binary state, with prior $\mu_0 = 1/2$: $\widehat{V}(\mu) = (|\mu - \frac{1}{2}| - \frac{1}{4})^2$. That is, $\widehat{V}(\mu) \leq 1/16$ and $\widehat{V}(\mu) = 1/16$ exactly at $\mu \in \{0, 1/2, 1\}$. Then let $\overline{V} = \text{co}\widehat{V}$, and $\underline{V} = \text{lco}\widehat{V}$ in Definition 5 of dominance.

No disclosure is a Bayesian solution, yielding a payoff of $1/16$. However, no disclosure is dominated by full disclosure: Full disclosure yields $1/16$ always, that is, regardless of what Nature does. On the other hand, there are signals for Nature (corresponding to some selection of the function V) under which no disclosure by the Sender generates strictly less than $1/16$; for example, Nature can induce the beliefs $1/4$ and $3/4$ with probability $1/2$ each, yielding a zero payoff for the Sender.

It is instructive to see which step of the proof of Theorem 3(b) fails for Bayesian solutions: In case (a) of that proof, we relied on Lemma 1 to argue that for a robust solution ρ , $\int \underline{V}(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0)$, which is a property equivalent to worst-case optimality. This is not true for no disclosure in the above example, because no disclosure is a Bayesian solution that is not worst-case optimal.

OA.3 Simultaneous-move-robust solutions

In our baseline model, we did not impose any restrictions on the signal chosen by Nature. In particular, Nature's choice of the signal could depend on the Sender's signal *realization*. In this appendix, we study a solution concept under which Nature chooses a signal simultaneously with the Sender. The assumption might be appropriate for settings in which Nature's move reflects the Sender's ambiguity over the information the Receivers possess prior to receiving the Sender's information, and acquiring additional information after receiving the Sender's information is too costly or otherwise infeasible for the Receivers.

To simplify exposition, we work with the baseline model of Section 2, except that

we allow for general conjectures. Unless specified otherwise, we maintain all the assumptions imposed in the main text.

The Sender continues to choose an information structure $q : \Omega \rightarrow \Delta\mathcal{S}$ which maps states ω into probability distributions of signal realizations $s \in \mathcal{S}$, but we do not assume that \mathcal{S} is finite (this would be with loss of generality). We also modify Nature’s strategy space: Nature selects a signal $\pi : \Omega \rightarrow \Delta\mathcal{R}$ that is independent of the Sender’s signal conditional on the state, with a signal space \mathcal{R} that is potentially infinite. Let Π_{CI} be the set of signals available to Nature, where “CI” stands for “conditionally independent.”²⁸

The base-case payoff $\widehat{v}(q)$ obtained when the Sender selects a signal q is computed under the conjecture that Nature selects some fixed (conditionally independent) signal $\pi_0 : \Omega \rightarrow \Delta\mathcal{R}$:

$$\widehat{v}(q) := \sum_{\omega \in \Omega} \int_{\mathcal{S}} \int_{\mathcal{R}} \left(\int_A v(a, \omega) d\xi_0(a | \mu_0^{s,r}) \right) d\pi_0(r | \omega) dq(s | \omega) \mu_0(\omega),$$

where ξ_0 is the conjectured tie-breaking rule, with $\xi_0(A^*(\mu) | \mu) = 1$ for all μ .²⁹ We can similarly define \widehat{V} as in formula (4.1) in Section 4, except that the conjecture about Nature is that it uses a signal $\pi_0 \in \Pi_{CI}$ (π_0 is not a function of the posterior belief generated by the Sender). Throughout, we assume that \widehat{V} is upper semi-continuous.

Let

$$\underline{v}(q, \pi) := \sum_{\omega \in \Omega} \int_{\mathcal{S}} \int_{\mathcal{R}} \underline{V}(\mu_0^{s,r}) d\pi(r | \omega) dq(s | \omega) \mu_0(\omega),$$

denote the Sender’s payoff from choosing q when Nature chooses π , under the adversarial selection \underline{V} (defined as in the main text). We define two notions of worst-case optimality, corresponding to cases 2 and 3 introduced in Section 6.3.

Definition OA.1. A signal $q \in Q$ is CI-worst-case optimal if it maximizes the worst-case payoff:

$$q \in \operatorname{argmax}_{q' \in Q} \left\{ \inf_{\pi \in \Pi_{CI}} \underline{v}(q', \pi) \right\}.$$

Definition OA.2. A signal $q \in Q$ is SM-worst-case optimal if it is part of a Bayes-Nash equilibrium of a simultaneous-move game against Nature: There exists $\pi \in \Pi_{CI}$

²⁸As in the main text, we assume that \mathcal{R} and \mathcal{S} are subsets of some sufficiently rich but fixed space.

²⁹We continue to denote by $A^*(\mu) := \operatorname{argmax}_{a \in A} \sum_{\Omega} u(a, \omega) \mu(\omega)$ the set of actions that maximize the Receiver’s expected payoff when her posterior belief is μ .

such that

$$\begin{aligned} q &\in \operatorname{argmax}_{q' \in Q} \underline{v}(q', \pi), \\ \pi &\in \operatorname{argmin}_{\pi' \in \Pi_{CI}} \underline{v}(q, \pi'). \end{aligned}$$

CI-worst-case optimality captures the idea that Nature can best respond to the Sender's choice of a signal but cannot condition on the Sender's signal realization. SM-worst-case optimality corresponds to a simultaneous-move game, in which Nature does not observe the Sender's choice of a signal. As foreshadowed in Section 6.3, we can prove that these two definitions are equivalent in our problem.

Lemma OA.1. *The following statements are equivalent:*

1. q is CI-worst-case optimal;
2. q is SM-worst-case optimal;
3. q generates the full-disclosure payoff in the worst-case scenario:

$$\inf_{\pi \in \Pi_{CI}} \underline{v}(q, \pi) = \underline{V}_{full}(\mu_0).$$

Proof. (1) \implies (2). Suppose that $q \in \operatorname{argmax}_{q' \in Q} \{\inf_{\pi \in \Pi_{CI}} \underline{v}(q', \pi)\}$. We argue that (q, π_{full}) is a Bayes-Nash equilibrium of the simultaneous-move game between Nature and the Sender, where π_{full} is the full-disclosure signal. Optimality of q for the Sender is trivial since any policy q' leads to the full-disclosure payoff $\underline{V}_{full}(\mu_0)$ against π_{full} . Optimality of π_{full} for Nature follows from the fact that q maximizes $\inf_{\pi \in \Pi_{CI}} \underline{v}(q', \pi)$ over all $q' \in Q$, which implies that, given q , Nature cannot bring the Sender's payoff below \underline{V}_{full} .³⁰

(2) \implies (3). Suppose that (q, π) is a Bayes-Nash equilibrium of the simultaneous-move game between Nature and the Sender. Since the Sender can always fully disclose the state, we have that $\underline{v}(q, \pi) \geq \underline{V}_{full}(\mu_0)$; but since Nature can also choose to fully disclose the state, we have that $\underline{v}(q, \pi) \leq \underline{V}_{full}(\mu_0)$. It follows that $\min_{\pi \in \Pi_{CI}} \underline{v}(q, \pi) = \underline{V}_{full}(\mu_0)$ which gives us (3).

(3) \implies (1). Because $\inf_{\pi \in \Pi_{CI}} \underline{v}(q', \pi) \leq \underline{V}_{full}(\mu_0)$ for all $q' \in Q$, (3) implies that q maximizes $\inf_{\pi \in \Pi_{CI}} \underline{v}(q', \pi)$, and hence q is CI-worst-case optimal. \square

³⁰Else, the Sender could improve upon q by fully disclosing the state, making Nature's move irrelevant, which contradicts the assumption that $q \in \operatorname{argmax}_{q' \in Q} \{\inf_{\pi \in \Pi_{CI}} \underline{v}(q', \pi)\}$.

Lemma [OA.1](#) has the flavor of the minimax theorem, with the full-disclosure payoff playing the role of the value of the zero-sum game between the Sender and Nature. Our minimax theorem does not require any continuity assumptions because full disclosure can always be obtained by either player, regardless of the sequence of moves. Given Lemma [OA.1](#), we can use any of the three equivalent definitions of worst-case optimality. The lemma reveals that the key difference to the baseline case is that Nature must select a signal that is conditionally independent of the Sender’s signal; the sequence of moves is not important.

We let W_{SM} denote the set of SM-worst-case optimal signals. Then, we define a *SM-robust solution* analogously to Definition [2](#): A signal q is a SM-robust solution if it maximizes $\widehat{v}(q)$ over W_{SM} .

OA.3.1 Summary of results

We start by summarizing the relationship between robust and SM-robust solutions. The summary serves as a road map for the next subsections where the results foreshadowed here are formally developed.

Characterizing SM-robust solutions turns out to be significantly more complicated than characterizing robust solutions. In particular, the restrictions imposed by SM-worst-case optimality do not take the tractable form described in Theorem [1](#). Therefore, the results that we obtain for this case are more limited in scope:

- Corollary [1](#) fails for SM-robust solutions, i.e., a SM-robust solution may fail to exist. We show in Subsection [OA.3.3](#) (Theorem [OA.2](#)) that a SM-robust solution exists under a stronger assumption of continuity of \underline{V} . Moreover, we introduce a notion of weak SM-robust solutions (that relaxes the condition of SM-worst-case optimality), and show that a weak SM-robust solution exists under no further assumptions on \underline{V} .
- In Subsection [OA.3.5](#), we provide a sufficient condition (Theorem [OA.3](#)) for state separation under a SM-robust solution. This condition is weaker than the one in Corollary [2](#); that is, whenever two states must be separated under a SM-robust solution, they also must be separated under a robust solution.
- Corollary [4](#) does not extend to SM-robust solutions because we do not have a characterization similar to the one in Theorem [1](#). In Subsection [OA.3.2](#) and Subsection [OA.3.5](#), we obtain various (weaker) sufficient conditions for either

full-disclosure to be the unique SM-robust solution, or for all distributions to be SM-worst-case optimal.

- In Subsection [OA.3.4](#), we analyze the binary-state case. Unlike robust solutions, as described by [Corollary 3](#), SM-robust solutions for binary-state problems may coincide with neither Bayesian solutions nor full disclosure. However, we give sufficient conditions for Bayesian solutions and full disclosure, respectively, to constitute SM-robust solutions.
- In Subsection [OA.3.6](#), we show that [Corollary 5](#) and [Corollary 6](#) fail for SM-robust solutions. That is, it is possible that a Bayesian solution is strictly more informative than all SM-robust solutions.
- [Corollaries 7](#) and [8](#) also fail: In fact, a SM-robust solution may require infinitely many signal realizations even when the state space is finite.

OA.3.2 Preliminary observations

We first make a couple of observations to simplify the problem of finding a SM-robust solution.

Lemma OA.2. *The set of SM-robust solutions when the signal space used by Nature is equal to Ω is the same as when it is equal to \mathcal{R} , for any $\mathcal{R} \supset \Omega$.*

Proof. Observe that, for any $\pi : \Omega \rightarrow \Delta\mathcal{R}$,

$$\begin{aligned} \underline{v}(q, \pi) &= \sum_{\omega \in \Omega} \int_{\mathcal{R}} \int_{\mathcal{S}} \underline{V}(\mu_0^{s,r}) d\pi(r|\omega) dq(s|\omega) \mu_0(\omega) \\ &= \int_{\mathcal{R}} \underbrace{\left(\sum_{\omega \in \Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu_0^{s,r}) dq(s|\omega) \right] \mu_0^r(\omega) \right)}_{\underline{V}_q(\mu_0^r)} \left(\sum_{\omega \in \Omega} d\pi(r|\omega) \mu_0(\omega) \right), \end{aligned}$$

where

$$\underline{V}_q(\mu) := \sum_{\omega \in \Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu^s) dq(s|\omega) \right] \mu(\omega).$$

Therefore,

$$\underline{v}(q, \pi) = \int_{\mathcal{R}} \underline{V}_q(\mu_0^r) d\Pi_{\mu_0, \pi}(r),$$

where $\Pi_{\mu_0, \pi} \in \Delta\mathcal{R}$ denotes the unconditional distribution over \mathcal{R} induced by μ_0 and π . From this observation, it is easy to see that, without loss of generality, we can

assume that Nature chooses a distribution $\nu \in \Delta\Delta\Omega$ over posterior beliefs over Ω , subject to Bayes plausibility. In particular, to minimize the Sender's payoff, Nature solves the following problem: $\inf_{\nu \in \Delta\Delta\Omega} \int \underline{V}_q(\mu) d\nu(\mu)$ subject to Bayes-plausibility $\int \mu d\nu(\mu) = \mu_0$. When $\underline{V}(\mu)$ is lower semi-continuous, so is $\underline{V}_q(\mu)$, for any q . Formally, for any sequence $\{\mu_n\}$ of posterior beliefs over Ω converging to $\mu \in \Delta\Omega$, we have that

$$\begin{aligned}
\liminf_n \underline{V}_q(\mu_n) &= \liminf_n \sum_{\Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu_n^s) dq(s|\omega) \right] \mu_n(\omega) \\
&= \liminf_n \left\{ \sum_{\Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu_n^s) dq(s|\omega) \right] \mu(\omega) + \sum_{\Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu_n^s) dq(s|\omega) \right] [\mu_n(\omega) - \mu(\omega)] \right\} \\
&\geq \sum_{\Omega} \left[\int_{\mathcal{S}} \liminf_n \underline{V}(\mu_n^s) dq(s|\omega) \right] \mu(\omega) + \liminf_n \sum_{\Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu_n^s) dq(s|\omega) \right] [\mu_n(\omega) - \mu(\omega)] \\
&\geq \sum_{\Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu^s) dq(s|\omega) \right] \mu(\omega) - \|\underline{V}\| \cdot \liminf_n \sum_{\Omega} |\mu_n(\omega) - \mu(\omega)| \\
&= \sum_{\Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu^s) dq(s|\omega) \right] \mu(\omega) = \underline{V}_q(\mu),
\end{aligned}$$

where the first inequality follows from Fatou's lemma, whereas the second inequality follows from the fact that \underline{V} is bounded, along with the continuity of posterior beliefs in the prior.

Therefore, Nature's problem has a solution. Furthermore, minimizing the Sender's payoff requires at most $|\Omega|$ signals (by the same argument as in [Kamenica and Gentzkow, 2011](#)). Thus, it is without loss of generality to set $\mathcal{R} = \Omega$ to characterize SM-worst-case optimal signals. \square

From now on we assume that $\mathcal{R} = \Omega$ (unless stated otherwise) and abuse notation slightly by letting $\pi(r|\omega)$ denote the probability Nature sends signal r in state ω (using the fact that the signal space is finite).

We apply a similar transformation to the Sender's problem next. By the law of total probability,

$$\sum_{\omega, r \in \Omega} \int_{\mathcal{S}} \underline{V}(\mu_0^{s,r}) \pi(r|\omega) dq(s|\omega) \mu_0(\omega) = \int_{\mathcal{S}} \underbrace{\left(\sum_{\omega, r \in \Omega} \underline{V}(\mu_0^{s,r}) \pi(r|\omega) \mu_0^s(\omega) \right)}_{\underline{V}_{\pi}(\mu_0^s)} \left(\sum_{\omega \in \Omega} dq(s|\omega) \mu_0(\omega) \right),$$

where

$$\underline{V}_{\pi}(\mu) := \sum_{\omega, r \in \Omega} \underline{V}(\mu^r) \pi(r|\omega) \mu(\omega),$$

and hence

$$\sum_{\omega, r \in \Omega} \int_{\mathcal{S}} \underline{V}(\mu_0^{s,r}) \pi(r|\omega) dq(s|\omega) \mu_0(\omega) = \int_{\mathcal{S}} \underline{V}_{\pi}(\mu^s) dQ_{\mu_0, q}(s),$$

where $Q_{\mu_0, q} \in \Delta \mathcal{S}$ is the unconditional distribution over \mathcal{S} induced by μ_0 and q . Therefore, the problem of finding a SM-robust solution is equivalent to the problem of finding a Bayes-plausible $\rho \in \Delta \Delta \Omega$ that maximizes $\int \widehat{V}(\mu) d\rho(\mu)$ among all SM-worst-case optimal distributions. By an argument analogous to the one used to prove Lemma 1, SM-worst-case optimality is equivalent to

$$\inf_{\pi: \Omega \rightarrow \Delta \mathcal{R}} \int \underline{V}_{\pi}(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0). \quad (\text{SM-WC})$$

As before, we will abuse terminology slightly by calling ρ a SM-robust solution. We also introduce the set \mathcal{W}_{SM} of worst-case optimal distributions of posterior beliefs (induced by the set W_{SM} of worst-case optimal signals).

Condition (SM-WC), contrasted with condition (WC) from Lemma 1, highlights the difference between worst-case optimality and SM-worst-case optimality. In Lemma 1, the infimum operator (embedded in the definition of \underline{V}) is inside the integral, i.e., the infimum over Nature's signals is computed posterior by posterior. For SM-worst-case optimality, instead, the infimum operator is outside the integral because Nature cannot respond differently to each realized posterior induced by the Sender's signal.

OA.3.3 Existence

Unlike in the baseline model, without additional restrictions on \underline{V} , existence of a SM-robust solution cannot be guaranteed. Example OA.2 illustrates the difficulty.

Example OA.2 (Non-existence of SM-robust solutions). Suppose the state ω is binary, $\Omega = \{0, 1\}$, $\Delta \Omega = [0, 1]$, $\mu \in [0, 1]$ is the probability that $\omega = 1$, and $\mu_0 = 1/2$. Define the correspondence

$$\mathcal{V}(\mu) := \begin{cases} \{2\mu\} & \mu < 1/2, \\ [-1, 1] & \mu = 1/2, \\ \{2 - 2\mu\} & \mu > 1/2, \end{cases}$$

and let \widehat{V} and \underline{V} be, respectively, the point-wise highest and lowest selection from the correspondence \mathcal{V} . Then, \widehat{V} is continuous, whereas \underline{V} has a discontinuity at $\mu = 1/2$. A distribution ρ is SM-worst-case optimal if and only if it guarantees the Sender a payoff of 0 (this is the payoff from full disclosure of the binary state). Any Bayes-

plausible continuous distribution of posterior beliefs (for example, $\rho \in \Delta\Delta\Omega$ that is uniform on $[0, 1]$) yields a payoff guarantee of 0 because Nature cannot induce a posterior belief of $1/2$ with positive probability. This conclusion relies crucially on the assumption that Nature’s signal is conditionally independent of the Sender’s signal.

To see why a SM-robust solution does not exist, note that the set \mathcal{W}_{SM} is not closed. For example, consider any sequence of Bayes-plausible distributions of posterior beliefs such that (i) each distribution in the sequence is atomless, and (ii) the sequence converges (in the weak* topology) to a Dirac delta at $1/2$ (induced by the uninformative signal). Then, each distribution in the sequence belongs to \mathcal{W}_{SM} but the limit does not. Moreover, the sequence yields expected base-case payoffs that converge to the upper bound of 1. The supremum of 1 cannot be achieved by any SM-worst-case optimal distribution because the only candidate—a Dirac delta at $1/2$ —is not SM-worst-case optimal.

Note that a Dirac delta at $1/2$ (which corresponds to no disclosure) can be approximated by a sequence of distributions that are SM-worst-case optimal. ■

The observations in the example above motivate a weaker definition of robustness for which existence is guaranteed.

Definition OA.3. A Bayes-plausible distribution of posterior beliefs $\rho \in \Delta\Delta\Omega$ is a *weak* SM-robust solution if it maximizes $\int \widehat{V}(\mu)d\rho(\mu)$ over $cl(\mathcal{W}_{SM})$, where $cl(\mathcal{W}_{SM})$ denotes the closure (in the weak* topology) of the set of SM-worst-case optimal distributions of posterior beliefs.

A weak solution thus relaxes the requirement that the distribution ρ is SM-worst-case optimal. Instead, it requires that it can be approximated by distributions that are SM-worst-case optimal. We establish the following existence result.

Theorem OA.2. *A weak SM-robust solution exists. If \underline{V} is continuous, then a SM-robust solution also exists.*

Proof. Define

$$v(\rho) := \inf_{\pi: \Omega \rightarrow \Delta\mathcal{R}} \int \underline{V}_\pi(\mu)d\rho(\mu)$$

as the SM-worst-case value for the Sender when she chooses the distribution ρ . We will prove that this function is continuous in ρ when \underline{V} is continuous. Throughout, we use the weak* topology on the space of distributions.

First, by a result in [Kamenica and Gentzkow \(2011\)](#), for any Bayes-plausible distribution of posterior beliefs $\rho \in \Delta\Delta\Omega$ there exists a signal $q_\rho : \Omega \rightarrow \Delta\mathcal{S}$ that induces this distribution (the subsequent results do not depend on which particular q_ρ we select). From the proof of [Lemma OA.2](#), we then have that $v(\rho)$ is equal to the value of the following minimization problem by Nature: $\inf_{\nu \in \Delta\Delta\Omega} \int \underline{V}_{q_\rho}(\mu) d\nu(\mu)$ subject to $\int \mu d\nu(\mu) = \mu_0$, where, for any signal q , \underline{V}_q is defined as in the proof of [Lemma OA.2](#).

Second, note that, under the assumption that \underline{V} is continuous, $\int \underline{V}_{q_\rho}(\mu) d\nu(\mu)$ is continuous in (ν, ρ) (this amounts to saying that, under a continuous objective function, the payoff from any pair of signals is continuous in their distribution).

Third, because the set of distributions $\nu \in \Delta\Delta\Omega$ satisfying the Bayes plausibility constraint $\int \mu d\nu(\mu) = \mu_0$ is compact, and because the objective function \underline{V} is continuous, it follows from Berge's theorem of maximum that the value function $v(\rho)$ is continuous in ρ . Hence, the problem of finding a distribution $\rho \in \Delta\Delta\Omega$ that maximizes $v(\rho)$ over the set of Bayes-plausible distributions has a solution, and the set of solutions, \mathcal{W}_{SM} , is non-empty and compact.

When, instead, \underline{V} is not continuous, what remains true is that the set $cl(\mathcal{W}_{SM})$ is non-empty (because full disclosure belongs to it, by [Lemma OA.1](#)) and compact because it is a closed subset of a compact space (the space of all Bayes-plausible distributions).

We can now finish the proof of both parts of [Theorem OA.2](#) with a single argument by observing that in the case when \underline{V} is continuous, we have $\mathcal{W}_{SM} = cl(\mathcal{W}_{SM})$. Thus, the problem of finding a (weak) SM-robust solution is equivalent to the problem of finding a distribution $\rho \in \Delta\Delta\Omega$ that maximizes $\int \widehat{V}(\mu) d\rho(\mu)$ over $cl(\mathcal{W}_{SM})$. Because the objective function is upper semi-continuous in ρ (this follows from the fact that, by assumption, \widehat{V} is upper semi-continuous), and the domain $cl(\mathcal{W}_{SM})$ is compact, a solution to the above problem exists, thus establishing existence of (weak) SM-robust solutions. \square

When Nature can send arbitrary signals, including signals that are correlated with the Sender's signal, existence of robust solutions does not require the additional assumption that \underline{V} is continuous (see [Corollary 1](#) in the main text). This is because, in that case, given any induced posterior μ , adversarial Nature always brings the conditional expected payoff of the Sender down to $\text{lco}(\underline{V})(\mu)$ —the lower convex closure of \underline{V} evaluated at μ . The lower convex closure is a convex function, and convex

functions are continuous on the interior of the domain. This guarantees that the set \mathcal{W} of worst-case optimal distributions is closed, while, in general, the set of SM-worst-case optimal distributions \mathcal{W}_{SM} need not be closed.

OA.3.4 SM-robustness for binary state

In this subsection, we consider the case of a binary Ω . Unlike in the baseline model, considering this case first is useful because our general characterization of state separation in the next subsection relies on the analysis of the binary case. Let $\Omega = \{0, 1\}$, and, with a slight abuse of notation, let $\underline{V}(\mu)$ denote the payoff to the Sender when the posterior belief μ puts probability μ on state 1. Let $s := \underline{V}(1) - \underline{V}(0)$ denote the slope of the (affine) function describing the full-disclosure payoff.

Proposition OA.1. *If either (i) \underline{V} is right-differentiable at 0 and $\underline{V}'(0) < s$, or (ii) \underline{V} is left-differentiable at 1 and $\underline{V}'(1) > s$, then full disclosure is the unique SM-robust solution.*

Proof. We only prove the result for case (i)—the proof for case (ii) is analogous. We do so by showing that full disclosure is the unique signal that is SM-worst-case optimal. Without loss of generality, normalize $\underline{V}(0) = 0$ so that $s = \underline{V}(1)$. Full disclosure yields the payoff of $\mu_0 \underline{V}(1)$ regardless of what Nature does. We will prove that the only way to guarantee a payoff of $\mu_0 \underline{V}(1)$ is to disclose all information. To show this, it suffices to show that for all Bayes-plausible $\rho \in \Delta\Delta\Omega$ with support other than $\{0, 1\}$ (where $\mu = 0$ and $\mu = 1$ are the two Dirac distributions assigning measure one to $\omega = 0$ and $\omega = 1$, respectively), there exists a (binary) signal π for Nature such that the Sender's payoff given ρ and π is strictly below $\mu_0 \underline{V}(1)$.

Let π be the binary signal given by $\pi(1|1) = \bar{\pi}$, $\pi(0|1) = 1 - \bar{\pi}$, and $\pi(0|0) = 1$, where $\bar{\pi} \in [0, 1]$. Under this signal, given any posterior belief μ induced by the Sender, Nature splits μ into $p = 1$ with probability $\mu\bar{\pi}$, and $p = \frac{(1-\bar{\pi})\mu}{1-\mu\bar{\pi}}$ with probability $1 - \mu\bar{\pi}$. Let $U_\rho(\bar{\pi})$ denote the conditional expected payoff to the Sender when the latter chooses the distribution $\rho \in \Delta\Delta\Omega$ and Nature chooses the signal π with parameter $\bar{\pi}$:

$$\begin{aligned} U_\rho(\bar{\pi}) &= \int_0^1 \left[\mu\bar{\pi}\underline{V}(1) + (1 - \mu\bar{\pi})\underline{V}\left(\frac{(1-\bar{\pi})\mu}{1-\mu\bar{\pi}}\right) \right] d\rho(\mu) \\ &= \mu_0\bar{\pi}\underline{V}(1) + \int_0^1 (1 - \mu\bar{\pi})\underline{V}\left(\frac{(1-\bar{\pi})\mu}{1-\mu\bar{\pi}}\right) d\rho(\mu). \end{aligned}$$

In particular, we have that $U_\rho(1) = \mu_0 \underline{V}(1)$ because $\bar{\pi} = 1$ corresponds to a signal by Nature that fully discloses the state. Let $U'_\rho(1)$ denote the left derivative of $U_\rho(\bar{\pi})$

with respect to $\bar{\pi}$, evaluated at $\bar{\pi} = 1$ (let $\Delta\rho(1)$ be the probability mass that ρ puts on the belief $\mu = 1$). We then have that

$$\begin{aligned}
U'_\rho(1) &= \lim_{\epsilon \rightarrow 0} \frac{U_\rho(1) - U_\rho(1 - \epsilon)}{\epsilon} = \mu_0 \underline{V}(1) - \lim_{\epsilon \rightarrow 0} \frac{\int_0^1 (1 - \mu(1 - \epsilon)) \underline{V}\left(\frac{\epsilon\mu}{1 - \mu(1 - \epsilon)}\right) d\rho(\mu)}{\epsilon} \\
&\stackrel{(1)}{=} \mu_0 \underline{V}(1) - \int_{[0, 1)} \left(\lim_{\epsilon \rightarrow 0} \frac{\underline{V}\left(\frac{\epsilon\mu}{1 - \mu(1 - \epsilon)}\right)}{\frac{\epsilon\mu}{1 - \mu(1 - \epsilon)}} \frac{\mu - \mu^2 + \mu^2\epsilon}{1 - \mu + \mu\epsilon} \right) d\rho(\mu) - \underline{V}(1)\Delta\rho(1) \\
&= \mu_0 \underline{V}(1) - \underline{V}'(0) [\mu_0 - \Delta\rho(1)] - \underline{V}(1)\Delta\rho(1) = [\mu_0 - \Delta\rho(1)] [s - \underline{V}'(0)] > 0,
\end{aligned} \tag{OA.1}$$

as long as $\mu_0 > \Delta\rho(1)$ —which is true except when ρ is full disclosure. In step (1) above, we have used the Lebesgue dominated convergence theorem (using the fact that \underline{V} is bounded, and has a right derivative at $\mu = 0$). The reason why we separated the integral over $[0, 1]$ into an integral over $[0, 1)$ and its value at 1 is that, for all $\mu < 1$, we have that $\lim_{\epsilon \rightarrow 0} \frac{\epsilon\mu}{1 - \mu(1 - \epsilon)} = 0$, but for $\mu = 1$, $\frac{\epsilon\mu}{1 - \mu(1 - \epsilon)} = 1$.

Summarizing, unless $\rho = \rho_{\text{full}}$, where ρ_{full} denotes the distribution induced by full disclosure, we have $U'_\rho(1) > 0$, and hence $\mu_0 \underline{V}(1) = U_\rho(1) > U_\rho(1 - \epsilon)$ for small enough ϵ . This means that, when $\rho \neq \rho_{\text{full}}$, Nature can bring the Sender's payoff strictly below the full-disclosure payoff $\underline{V}_{\text{full}}(\mu_0)$ by selecting a binary signal π that is almost fully revealing. Therefore, full disclosure is the unique SM-worst-case optimal distribution, and hence the unique SM-robust solution. \square

The judge example of [Kamenica and Gentzkow \(2011\)](#) satisfies assumption (i) of Proposition [OA.1](#) because the derivative of \underline{V} at 0 is 0, while the slope $s = \underline{V}(1) - \underline{V}(0)$ is strictly positive. Therefore, the unique SM-robust solution is full disclosure of the state.

The proof of Proposition [OA.1](#) shows that, through an appropriate binary signal, Nature can decompose any non-degenerate posterior belief μ induced by the Sender into a Dirac delta at $\omega = 1$ and a posterior arbitrarily close to a Dirac at $\omega = 0$. The condition $s > \underline{V}'(0)$ implies that any posterior belief close to (but different from) a Dirac at $\omega = 0$ gives the Sender a payoff strictly less than a Dirac at $\omega = 0$. In turn, this implies that, unless the Sender fully reveals the state herself, Nature can bring the Sender's expected payoff strictly below the full-disclosure payoff. In such cases, full disclosure is the unique SM-robust solution.

Loosely speaking, the Sender fully reveals the state not because she is worried

that, else, Nature will do it, but because she is worried that Nature will *almost* fully reveal the state. Under the conditions in Proposition OA.1, almost full revelation is strictly worse than full revelation.

The above intuition can also be used to compare SM-worst-case optimality to worst-case optimality (and hence SM-robustness to robustness). As explained in the main text, a sufficient condition for full disclosure to be the unique robust solution is that the payoff $\underline{V}(\mu)$ lies below the full-disclosure payoff $(1 - \mu)\underline{V}(0) + \mu\underline{V}(1)$ at *some* interior $\hat{\mu}$. A sufficient condition for full disclosure to be the unique SM-robust solution is that $\underline{V}(\mu)$ is below the full-disclosure payoff $(1 - \mu)\underline{V}(0) + \mu\underline{V}(1)$ for μ sufficiently close to one of the two bounds, $\mu = 0$ or $\mu = 1$. When Nature can condition her disclosure on the *realization* of the Sender’s signal (equivalently, on the posterior μ induced by the Sender), for any interior μ , Nature can induce the “final” posterior belief $\hat{\mu}$ with positive probability, without restricting its own ability to influence the Receivers’ beliefs conditional on other realizations of the Sender’s signal. In contrast, when Nature’s signal is conditionally independent, and Nature chooses to induce the posterior belief $\hat{\mu}$ with positive probability conditional on the Sender inducing μ , it can no longer independently choose what posterior beliefs the Receivers will have conditional on other realizations of the Sender’s signal. In particular, even if Nature’s signal realization shifts μ to a $\hat{\mu}$ that yields a low payoff to the Sender, the same signal realization could shift some other η induced by the Sender to a $\hat{\eta}$ that has a high payoff to the Sender. In short, Nature cannot target the same posterior belief $\hat{\mu}$ regardless of the realization of the Sender’s signal.

There is an important exception though: By “almost” fully disclosing the state, Nature can ensure that, no matter the posterior belief induced by the Sender, the final posterior is in an arbitrary small neighborhood of a Dirac belief δ_ω , with a probability arbitrarily close to 1 conditional on ω (effectively, in this case, although Nature cannot always target a particular $\hat{\mu}$, it can target an arbitrarily small region). If the Sender’s payoff $\underline{V}(\mu)$ is below the full-disclosure payoff for μ in a neighborhood of δ_ω , Nature can exploit any discretion left by the Sender to push the Sender’s payoff strictly below $\underline{V}_{\text{full}}$. This is what makes the neighborhoods of Dirac measures special in the analysis of SM-worst-case optimality.

As a partial converse to Proposition OA.1, we have the following result:

Proposition OA.2. *If $\underline{V}(\mu) \geq \underline{V}_{\text{full}}(\mu)$ for all μ , then all Bayes-plausible distributions $\rho \in \Delta\Delta\Omega$ are SM-worst-case optimal. In this case, a distribution $\rho \in \Delta\Delta\Omega$ is*

a SM-robust solution if and only if it is a Bayesian solution.

Proof. By Proposition 1 in the main text, all Bayes-plausible distributions are worst-case optimal under the assumptions of Proposition OA.2; hence they are also SM-worst-case optimal. For $\rho \in \Delta\Delta\Omega$ to be a SM-robust solution, ρ must maximize \widehat{V} over the entire set of Bayes-plausible distributions, which means that ρ must be a Bayesian solution. Likewise, if ρ is a Bayesian solution, it maximizes \widehat{V} over the entire set of SM-worst-case optimal solutions and hence it is SM-robust. \square

We can summarize the results for the binary-state case as follows. If $\underline{V}(\mu) \geq \underline{V}_{\text{full}}(\mu)$ for all μ , then, neither worst-case nor SM-worst-case optimality have any bite. In this case, the set of SM-robust solutions coincides with the set of robust solutions, which coincides with the set of Bayesian solutions. If, instead, $\underline{V}(\mu) < \underline{V}_{\text{full}}(\mu)$ for *some* μ , then full disclosure is the unique robust solution but not necessarily the unique SM-robust solution. However, full disclosure is the unique SM-robust solution if $\underline{V}(\mu) < \underline{V}_{\text{full}}(\mu)$ for μ in some neighborhood of either 0 or 1. When $\underline{V}(\mu) < \underline{V}_{\text{full}}(\mu)$ for some interior μ but not in any neighborhood of either 0 or 1, the set of SM-robust solutions may be difficult to characterize.

OA.3.5 State separation under SM-robustness

In this subsection, we characterize properties of SM-robust solutions for the general case with an arbitrary number of states. The analysis parallels the one leading to Theorem 1 in the main text, but the results are not as sharp as in the case of robust solutions.

Given a function $V : \Delta\Omega \rightarrow \mathbb{R}$, let $dV(\mu; \mu')$ denote the Gateaux derivative of V at μ in the direction of μ' . The latter is defined by

$$dV(\mu; \mu') := \lim_{\epsilon \rightarrow 0} \frac{V((1 - \epsilon)\mu + \epsilon\mu') - V(\mu)}{\epsilon},$$

whenever the limit exists. Recall that $\underline{V}_{\text{full}}(\mu) = \sum_{\Omega} \underline{V}(\delta_{\omega})\mu(\omega)$ is the expected payoff from full disclosure. We then have that, starting from the Dirac measure $\mu = \delta_{\omega}$, the Gateaux derivative of $\underline{V}_{\text{full}}(\mu)$ in the direction of the Dirac measure $\delta_{\omega'}$ is equal to

$$d\underline{V}_{\text{full}}(\delta_{\omega}; \delta_{\omega'}) = \lim_{\epsilon \rightarrow 0} \frac{\underline{V}_{\text{full}}((1 - \epsilon)\delta_{\omega} + \epsilon\delta_{\omega'}) - \underline{V}_{\text{full}}(\delta_{\omega})}{\epsilon} = \underline{V}(\delta_{\omega'}) - \underline{V}(\delta_{\omega}).$$

Theorem OA.3. *Suppose that for some $\omega, \omega' \in \Omega$, $d\underline{V}(\delta_{\omega}; \delta_{\omega'}) < \underline{V}(\delta_{\omega'}) - \underline{V}(\delta_{\omega})$. Then, any SM-worst-case optimal distribution ρ must separate states ω and ω' with probability one.*

Proof. The proof relies on insights developed for the binary-state case (see Proposition OA.1). Nature can always fully reveal the states $\Omega \setminus \{\omega, \omega'\}$, so that, conditional on the state belonging to $\{\omega, \omega'\}$, the results for the binary-state case apply.

Suppose that some SM-worst-case optimal distribution ρ does not separate ω and ω' . That is, there exists a non-zero-measure set of $\mu \in \text{supp}(\rho)$ such that $\mu(\omega)\mu(\omega') > 0$. Consider a signal π by Nature that reveals all states other than ω and ω' perfectly, and, conditional on the state belonging to $\{\omega, \omega'\}$, sends signals as in the proof of Proposition OA.1. The condition $d\underline{V}(\delta_\omega; \delta_{\omega'}) < \underline{V}(\delta_{\omega'}) - \underline{V}(\delta_\omega)$ implies that the assumptions of Proposition OA.1 hold. Given π , the Sender's expected payoff is strictly below her full-disclosure payoff, and hence ρ is not a SM-worst-case optimal distribution. \square

We can also identify a simple sufficient condition under which no states need to be separated, and hence SM-robust solutions coincide with Bayesian solutions.

Corollary OA.1. *If $\underline{V} \geq \underline{V}_{full}$, then all Bayes-plausible distributions are SM-worst-case optimal.*

This is the same condition as the one identified by Corollary 4 in the main text. Moreover, Corollary 4 actually implies Corollary OA.1 because if a distribution is worst-case optimal when Nature can choose any signal, then it is also worst-case optimal when Nature is restricted to choosing conditionally independent signals.

Theorem OA.3 takes a more tractable form in the case when $\Omega \subset \mathbb{R}$, and the Sender's payoff depends only on the expected state.

Corollary OA.2. *Suppose that $\underline{V}(\mu) = u(\mathbb{E}_\mu[\omega])$ for some differentiable function u . If $u'(\omega) < \frac{u(\omega') - u(\omega)}{\omega' - \omega}$, then any SM-worst-case optimal distribution must separate the states ω and ω' with probability one.*

OA.3.6 A Bayesian solution can Blackwell dominate a SM-robust solution

Corollary 6 in the main text states that, for any Bayesian solution ρ_{BP} , one can find a robust solution ρ_{RS} that is either incomparable to, or more informative than, ρ_{BP} in the Blackwell sense. In this subsection, we show that this conclusion does not extend to SM-robust solutions. We do this by means of a counterexample. Our counterexample is rather contrived and has no immediate economic interpretation.

The example exploits the fact that Corollary 5 in the main text does not extend to SM-robust solutions: A mean-preserving spread of a SM-worst-case optimal distribution need not be SM-worst-case optimal. For intuition, think of a mean-preserving spread as an additional signal, on top of the original signal selected by the Sender. When Nature can condition her signal on the realization of the Sender’s signal, she can entertain mean-preserving spreads that provide additional information to the Receivers for some realizations of the Sender’s signals but not for others. This means that any mean-preserving spread engineered by the Sender can also be engineered by Nature. The result that mean-preserving spreads of worst-case optimal policies are worst-case optimal then follows from the fact that Nature can always engineer such spreads herself starting from the original distribution selected by the Sender. Hence, for the original distribution to be worst-case optimal, it must be that any mean-preserving spread of such distribution is also worst-case optimal.

This conclusion does not extend to the case of conditionally independent signals. The reason is that, when Nature is not allowed to condition her signal on the realization of the Sender’s signal, any mean-preserving spread of the Sender’s signal that Nature can choose provides more information to the Receivers than the original signal for *all* non-degenerate μ in the support of the Sender’s original distribution. This means that certain mean-preserving spreads by the Sender cannot be replicated by Nature. As a result, there is no guarantee that a mean-preserving spread designed by the Sender preserves SM-worst-case optimality. In turn, this implies that the Sender can strictly benefit from withholding information.

Counterexample. The state is binary, $\Omega = \{0, 1\}$, and the prior is uniform. Letting μ denote the probability assigned to the state $\omega = 1$, the Sender’s base-case payoff is given by $\widehat{V}(\mu) = 2$ if $\mu \notin G$ and $\widehat{V}(\mu) = 3$ if $\mu \in G$, where $G := \{1/3, 7/12, 2/3, 3/4\}$. Clearly, given \widehat{V} , there are many Bayesian solutions—any Bayes-plausible distribution of posteriors with support in G is optimal. Consider the solution ρ_{BP} that puts mass 1/2 on 1/3, mass 1/4 on 7/12, and mass 1/4 on 3/4. This solution is Blackwell more informative than the Bayesian solution ρ_R that puts mass 1/2 on 1/3, and mass 1/2 on 2/3. Indeed, the distribution ρ_{BP} can be obtained from the distribution ρ_R by sending an additional signal whenever the posterior induced by ρ_R is 2/3 (the additional signal then decomposes 2/3 into the posteriors 7/12 and 3/4). Figure OA.3.1 illustrates the base-case payoff function \widehat{V} (the black solid line) and the fact that ρ_{BP} is a mean-preserving spread of ρ_R (this fact is indicated by the

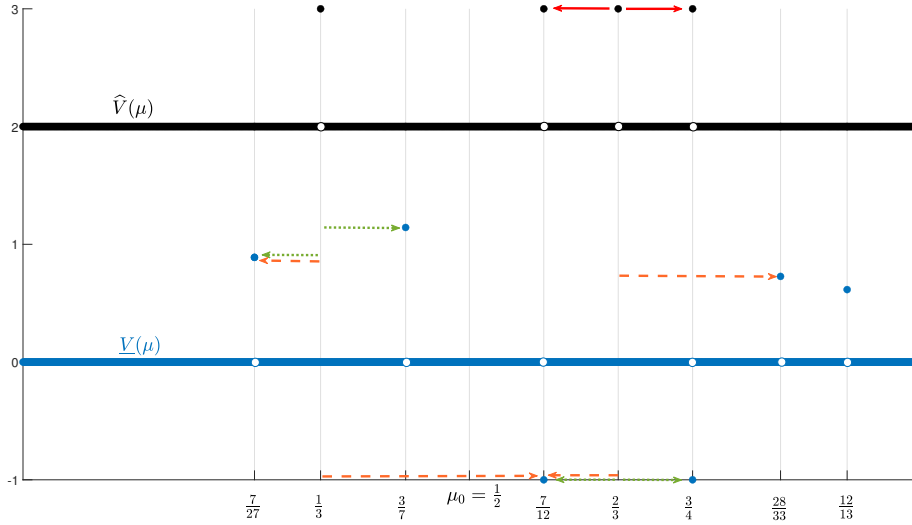


Figure OA.3.1: The functions \underline{V} and \widehat{V}

red solid arrows).

We complete the construction of the counterexample by selecting the Sender's payoff under the adversarial tie-breaking \underline{V} so that ρ_R is the unique SM-robust solution. We first give an intuitive description of how we derive \underline{V} from the properties required for the counterexample to work, and then provide a formal definition of \underline{V} and prove the result.

The idea is to construct a function \underline{V} under which the Sender gets a low payoff from beliefs $7/12$ and $3/4$, so that ρ_{BP} is not SM-worst-case optimal. Suppose that $\underline{V}(\mu) = 0$ except over a finite set of points, and that $\underline{V}(7/12) = \underline{V}(3/4) = -1$. Then, ρ_{BP} is clearly not SM-worst-case optimal, because, by not disclosing any information, Nature guarantees that the Sender's expected payoff under ρ_{BP} is strictly below her full-disclosure payoff, which is equal to zero. Note, however, that this is not enough, because under such \underline{V} , ρ_R is also not SM-worst-case optimal. Indeed, by choosing π appropriately, Nature can induce the posterior $\mu = 7/12$ and/or the posterior $\mu = 3/4$ with positive probability, thus bringing the Sender's payoff strictly below the full-disclosure payoff. Therefore, we construct \underline{V} so that, whenever Nature's response results in a low payoff for the Sender conditional on one of her posterior beliefs under ρ_R , it must result in a sufficiently high payoff for the Sender conditional on the other posterior belief. For example, Nature can split $2/3$ into $7/12$ and $3/4$;

however, the unique signal that achieves this split—precisely because the signal must be conditionally independent—must also split the other posterior $1/3$ into $7/27$ and $3/7$ (as illustrated by the green dotted arrows in Figure OA.3.1). Thus, we choose the values of \underline{V} to be sufficiently high at $7/27$ and $3/7$.

Figure OA.3.1 depicts one more possible response by Nature—indicated by the orange dashed arrows—that constrains the values of \underline{V} . To take into account all the relevant responses by Nature, we can use Lemma OA.2 which says that, to minimize the Sender’s expected payoff, Nature can restrict attention to binary signals. If $\underline{V}(7/12) = \underline{V}(3/4) = -1$, and $\underline{V}(\mu) \geq 0$ for all $\mu \notin \{7/12, 3/4\}$, it suffices to consider binary signals that, given ρ_R , induce a final posterior of either $7/12$ or $3/4$ with strictly positive probability. We also know from the proof of Lemma OA.2 that Nature’s problem can be thought of as choosing a distribution over $[0, 1]$ that minimizes the expectation of $\underline{V}_q(\mu)$ over all Bayes-plausible distributions, where q is any Sender’s signal that induces the distribution ρ_R . One such signal is given by $\mathcal{S} = \{l, h\}$, $q(l|0) = 2/3$, and $q(l|1) = 1/3$. Given this q , the Sender’s expected payoff when Nature induces the posterior μ is equal to

$$\underline{V}_q(\mu) = \sum_{\omega \in \Omega} \left[\int_{\mathcal{S}} \underline{V}(\mu^s) dq(s|\omega) \right] \mu(\omega) = \left(\frac{2}{3} - \frac{1}{3}\mu \right) \underline{V} \left(\frac{\mu}{2-\mu} \right) + \left(\frac{1}{3} + \frac{1}{3}\mu \right) \underline{V} \left(\frac{2\mu}{1+\mu} \right).$$

To guarantee that ρ_R is a SM-worst-case optimal distribution, it then suffices to choose a \underline{V} that (i) takes value 0 almost everywhere (including at $\mu = 0$ and at $\mu = 1$), (ii) is such that $\underline{V}(\mu) < 0$ only for $\mu \in \{7/12, 3/4\}$, at which it takes value $\underline{V}(7/12) = \underline{V}(3/4) = -1$, and (iii) induces $\underline{V}_q(\mu) \geq 0$ for all μ . There are only four values of μ at which $\underline{V}_q(\mu)$ could be potentially negative: $\mu \in \{7/17, 3/5, 14/19, 6/7\}$. Indeed, only for these four posteriors, given the Sender’s signal q , the final posterior takes a value equal to either $7/12$ or $3/4$. These four posteriors are given by the solutions to $\mu/(2-\mu) = 7/12$, $\mu/(2-\mu) = 3/4$, $(2\mu)/(1+\mu) = 7/12$, and $(2\mu)/(1+\mu) = 3/4$. At each such μ , we want $\underline{V}_q(\mu) = 0$. This gives us four equations in four unknowns—the values of \underline{V} at the posterior beliefs $2\mu/(1+\mu)$ and $\mu/(2-\mu)$ when the latter beliefs, computed for $\mu \in \{7/17, 3/5, 14/19, 6/7\}$, differ from either $7/12$ or $3/4$. Solving this system, we obtain that

$$\underline{V} \left(\frac{7}{27} \right) = \frac{8}{9}, \quad \underline{V} \left(\frac{3}{7} \right) = \frac{8}{7}, \quad \underline{V} \left(\frac{28}{33} \right) = \frac{8}{11}, \quad \underline{V} \left(\frac{12}{23} \right) = \frac{8}{13}, \quad (\text{OA.2})$$

as illustrated in Figure OA.3.1. This completes the construction of the function \underline{V} .

The following claim is then true.³¹

Claim OA.1. *Let $\Omega = \{0, 1\}$, the prior be uniform, $\underline{V}(\mu) = 0$ except that $\underline{V}(7/12) = \underline{V}(3/4) = -1$ and (OA.2) holds, and $\widehat{V}(\mu) = 2$ except that $\widehat{V}(1/3) = \widehat{V}(7/12) = \widehat{V}(2/3) = \widehat{V}(3/4) = 3$. Then, there exists a Bayesian solution ρ_{BP} that strictly Blackwell dominates the unique SM-robust solution ρ_R .*

By the construction of \underline{V} , ρ_R is SM-worst-case optimal, and because it yields the maximal payoff of 3 under \widehat{V} , it is a SM-robust solution. It only remains to show that ρ_R is the *unique* SM-robust solution. To see this, note that any other distribution ρ' that yields a payoff of 3 under \widehat{V} must assign strictly positive probability to either 7/12 or 3/4 and no mass outside of $\{1/3, 7/12, 2/3, 3/4\}$ (since this is the only way to guarantee an expected payoff of 3 which is required for being a SM-robust solution). Furthermore, for ρ' to be SM-worst-case optimal, it must yield a non-negative expected payoff under \underline{V} when Nature discloses no information which is impossible if ρ' assigns positive probability to $\{7/12, 3/4\}$.

Summarizing, we have constructed an example of a Bayesian solution ρ_{BP} that strictly dominates the unique SM-robust solution ρ_R in the Blackwell order.

³¹Note that, contrary to what we assumed throughout the analysis, the function \underline{V} considered in this example is not lower semi-continuous. However, this is not essential for the result. The specific function \underline{V} considered here simplifies the calculations but the result remains true also for certain lower semi-continuous functions.