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

Piotr Dworczak and Alessandro Pavan†

October 5, 2020

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 equilibrium selection. Thus, 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.

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

JEL codes: D83, G28, G33

*For comments and useful suggestions, we thank Emir Kamenica, 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 sources the Receiver may attain to 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 actually 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 an alternative formulation of 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. We first present our main ideas in a simple case inspired by the judge example from Kamenica and Gentzkow (2011).

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 probability of the state f. The prosecutor tries to maximize the expected sentence (with acquitting modeled as a sentence of $x = 0$). Formally, if μ is the induced posterior, with $\mu(\omega)$ denoting the probability of state ω, the Sender's payoff given μ is equal to

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

The Bayesian-persuasion solution (henceforth, Bayesian solution) is as follows: The prosecutor induces the posterior belief $(\mu(i),\, \mu(m),\, \mu(f)) = (1, 0, 0)$ with probability 1/4 (by saying "innocent" with probability 1/2 conditional on the state being i) and (1/3, 1/3, 1/3) with probability 3/4 (by saying "guilty" in all other cases). The expected payoff is $(3/4)\bar{x}$. (This can be verified using the concavification technique.)

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 just fully revealing the state.[2] 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 $(1/4)\underline{x} + (1/4)\bar{x}$.

The key observation of our paper is that the prosecutor—even if she is primarily concerned with the worst-case scenario—may prefer to *not fully disclose* the state. Consider the following alternative partitional signal: separate the state "innocent" from the other two states and then pool together the two remaining states. Suppose that the witness is adversarial. When it is already revealed that the defendant is innocent, the witness has no information 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 with this policy is exactly the same as under full disclosure. At the same time, it is strictly better if the prosecutor's conjecture turns out to be true—a

[1]The 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.

[2]It is known that full disclosure is a worst-case optimal signal in cases like the one we study here; see for example Hu and Weng (2019).

situation we will refer to as the *best case*: For example, the witness does not turn up, is uninformed about the state, or decides to stay quiet. In that case, the alternative policy yields $(1/2)\bar{x}$ which is strictly more than the full-disclosure payoff.

By the above reasoning, it is tempting to conclude that the prosecutor might as well stick to the original Bayesian solution. After all, if the witness discloses the state in case she is adversarial, shouldn't the prosecutor focus on maximizing her payoff under her conjecture? The problem with that argument is that the most adversarial scenario need *not* involve the witness fully disclosing the state. As a result, the Bayesian solution may yield a significantly lower payoff than full disclosure in the worst case. In the Bayesian solution, when the prosecutor induces the posterior $(1/3,\ 1/3,\ 1/3)$, the witness may reveal the state f with some small probability $\epsilon > 0$. This shifts the judge's posterior belief that the defendant is guilty conditional on the prosecutor saying "guilty" just below the threshold of $2/3$. The judge then acquits the defendant with probability arbitrarily close to one, not just when the latter is innocent but also when they are guilty. Thus, the worst-case expected payoff for the prosecutor from adopting the Bayesian solution is arbitrarily close to 0—showing that the Bayesian solution need not be robust to mis-specifications in the conjecture.

As is typically the case with non-Bayesian uncertainty, any policy chosen by the prosecutor results in a range of possible 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 is concerned about protecting herself against the worst possible case, and hence dismisses any policy that does not yield the maximal payoff guarantee in the worst-case scenario. Second, when the 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 in the best case, understood to be the one in which her conjecture is exactly right. The combination of these two properties defines a *robust solution*: a best-case optimal policy among those that are worst-case optimal.[3] It is straightforward to observe that the alternative policy described above is in fact a robust solution for the prosecutor. ■

[3]In 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 worst- and best-case payoffs, then, under permissive regularity conditions, the solutions coincide with robust solutions as long as the weight on the worst-case scenario is sufficiently large.

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 assume that the Sender's conjecture is that the Receiver does not have any exogenous information other than the one contained in the common prior, as in the canonical Bayesian persuasion model—we relax this assumption later in the analysis. We introduce a third player, called Nature, that may send an additional signal to the Receiver, potentially conditioning on the realization of the Sender's signal. We define a worst-case solution to be any information policy for the Sender that maximizes her expected payoff under the assumption that Nature sends information to minimize the Sender's payoff. We define a robust solution to be an information policy that maximizes the Sender's expected payoff under the conjecture, among all policies that are worst-case optimal.

Despite the fact that robust solutions involve worst-case optimality, we prove that they exist under standard conditions, and can be characterized by applying techniques similar to those used to identify Bayesian solutions (e.g. concavification). However, the economic properties of robust solutions can be quite different from those of Bayesian solutions. Our main technical 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 also 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 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 exactly the rationale behind the Sender always revealing the state "innocent" in the roust 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 there is always a way to specify the conjecture so that 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. The judge example above 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 equilibrium most favorable to her, in case there are multiple strategy 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

[4] Of 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.

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 (the best case). However, she may first impose a "robustness test" (worst-case optimality) to rule out policies that deliver a suboptimal payoff in the worst Bayes correlated equilibrium. For any given specification of the worst- and best-case Sender payoffs, our separation theorem characterizes the resulting robust solutions.

We hasten to clarify that our definitions of the worst and best case are not symmetric, and are defined *relative* to the Sender's conjecture. The worst-case scenario is that the conjecture is false in the worst possible way for the Sender. In contrast, our definition of the best-case scenario is that the conjecture is true, even when this conjecture does not coincide with the most favorable scenario for the Sender.[5]

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 in Section 3, we derive the main properties of robust solutions. Section 4 extends the model and the results to general persuasion problems, and Section 5 illustrates the results with a few applications. Finally, in Section 6, we discuss the relationship between robust solutions and the notion of dominance, as well as the version of the problem in which the Sender maximizes a weighted sum of her worst- and best-case payoffs. Most proofs are collected in Appendix A. The Online Appendix contains supplementary results, most notably a discussion of the case when Nature is constrained to send signals that are conditionally independent of the Sender's signal.

Related literature. The paper is related to the fast-growing literature on Bayesian persuasion and information design (see, among others, Bergemann and Morris, 2019, and Kamenica, 2019 for surveys). Most closely related are papers that adopt an adversarial approach to the design of the optimal information structure. Inostroza and Pavan (2018), Li et al. (2019), Mathevet et al. (2020), Morris et al. (2019), and

[5]In the single-Receiver case, the conjecture that the Receiver is uninformed happens to be the most favorable possible case for the Sender; we allow for arbitrary conjectures and hence this need not be the case in general.

Ziegler (2019) focus on the adversarial selection of the continuation strategy profile (rather than of the Receivers' exogenous information sources). Kosterina (2019) and Hu and Weng (2019), instead, study signals that maximize the Sender's payoff in the worst-case scenario, when the Sender faces uncertainty over the Receivers' exogenous private information (as in this paper). Kosterina (2019) assumes that Nature cannot condition the information she provides to the Receiver on the realization of the Sender's signal. This situation corresponds to the case of conditionally independent robust solutions examined in the Online Appendix of this paper. In a two-action-many-state model, she then studies how the Sender's worst-case optimal signal depends on the Sender's ignorance of the Receiver's prior beliefs and shows that when such ignorance is pronounced, the optimal policy recommends the Sender's preferred action with positive probability on all states (in contrast, the Bayesian solution is a cut-off policy recommending the Sender's preferred action with certainty on all states above a cut-off and with probability zero on the remaining states). Hu and Weng (2019) consider a fairly general model of Bayesian persuasion and 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.

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 (2018) study simultaneous-games where multiple Senders disclose information about the quality of their products. 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 (2019), and Wu (2018), instead, analyze games in which Senders move in a sequence and identify conditions for (a) full information revelation and (b) silent equilibria (that is, equilibria in which all

[6]As in Kosterina (2019), in the Stackelberg version of the zero-sum game between the competing designers, Cui and Ravindran (2020) assume that follower cannot condition its information on the realization of the leader's signal. As anticipated above, this scenario corresponds to the case of conditionally independent signals examined in the Online Appendix.

Senders but one remain silent) to span all equilibrium outcomes.

Kolotilin et al. (2017) and Laclau and Renou (2017), instead, consider persuasion of privately informed Receivers. In the first paper, the Receiver's private information is about a payoff component different from the one corresponding to the Sender's signal. In the second paper, the Receiver has multiple priors and max-min preferences.[7]

Our results are different from those in any of the above papers and reflect a different approach to the design of the optimal signals. Once she identifies all signals that are worst-case optimal, the Sender looks at the performance of any such signal under the best-case scenario (that is, under the assumed conjecture, 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 strictly benefit from withholding some information under the optimistic scenario while still guaranteeing the same worst-case payoff. Our lexicographic approach to the assessment of different information structures is in the same spirit of the one proposed by Börgers (2017) in the context of robust mechanism design.

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 $A^\star(\mu) = \text{argmax}_{a\in A} \sum_\Omega u(a, \omega)\mu(\omega)$ denote the set of actions that maximize the Receiver's expected payoff when her posterior belief over the state ω is $\mu \in \Delta\Omega$. The Sender has a continuous utility function $v(a, \omega)$. She chooses an information structure $q : \Omega \to \Delta\mathcal{S}$ that maps states into probability distributions over 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 cor-

[7] See also the literature on Bayesian Persuasion with rationally inattentive Receivers (Bloedel and Segal, 2018, Lipnowski et al., 2019, Matysková, 2019, and Ye, 2019). Contrary to the present paper, in that literature, the Receivers' attention costs are known to the designer.

related 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 $\pi : \Omega \times \mathcal{S} \to \Delta\mathcal{R}$ 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 even 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 $\mathcal{X}$. 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 forms a Bayesian belief about the distribution of the Receiver's exogenous information and the way the Receiver plays in case of indifference. We will 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 exposition, we assume in this section that the Sender's conjecture is that the Receiver only knows 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 can formally define

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

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 will sometimes refer to the function $\widehat{v}(\cdot)$ as the *best-case payoff*. As explained in the Introduction, the "best case" refers to the scenario in which the Sender's conjecture turns out to be correct.

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

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

where

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

We call $\underset{\sim}{v}(\cdot)$ 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 disclosure policy π 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 induces him to break 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 $\underset{\sim}{v}$ over the set of all signals Q.

We let $W \subset Q$ denote the set of worst-case optimal signals for the Sender.

Since Nature can always disclose the state, the Sender's payoff in the worst-case scenario is upper bounded by the full-information payoff $\underline{V}_{\text{full}}(\mu_0)$, defined as

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

where δ_ω is the Dirac distribution assigning measure one to the state ω. Clearly, this upper bound can be achieved if the Sender discloses the state herself.

Observation 1. A signal q is worst-case optimal (i.e., $q \in W$) if and only if $\underset{\sim}{v}(q) = \underline{V}_{\text{full}}(\mu_0)$. The set W of worst-case optimal signals is non-empty because full disclosure of the state is always worst-case optimal.

With this observation, we can now formally define the notion of a robust solution.

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

As anticipated in the Introduction, the definition of a robust solution reflects the Sender's lexicographic attitude towards the uncertainty she faces. First, the Sender seeks an information structure that is worst-case optimal, i.e., that is not outperformed by any other information structure, 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 Nature behaves according to her conjecture. In short, a robust solution is best-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 over posterior beliefs (rather than signals). The next lemma ensures that this is indeed possible in our setting. Define, for any $\mu \in \Delta\Omega$,

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

That is, $\underset{\sim}{V}(\mu)$ is the expected payoff for 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 $\underset{\sim}{V}(\mu)$ is defined at the interim stage, conditional on the Sender inducing some belief μ with her signal realization.

Lemma 1. *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 $\mathcal{W}$, where $\mathcal{W} \subset \Delta\Delta\Omega$ is the set of distributions over posterior beliefs satisfying*

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

and Bayes plausibility

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

Lemma 1 is intuitive. Any distribution over posterior beliefs ρ induced by some signal must satisfy Bayes plausibility (BP) (we refer to any distribution satisfying (BP) as *feasible*). Given any feasible distribution $\rho \in \Delta\Delta\Omega$, the Sender expects

Nature to respond to any posterior belief μ in the support of ρ by choosing a signal $\pi : \Omega \to \Delta\mathcal{R}$ that minimizes the Sender's expected payoff. Condition (WC) then states that ρ maximizes the Sender's payoff in the worst-case scenario. Thus, a signal q is in W if and only if the distribution over posterior beliefs ρ_q induced by q is in $\mathcal{W}$. Hereafter, we will abuse terminology and call ρ_{RS} a robust solution if it maximizes $\int \widehat{V}(\mu) d\rho(\mu)$ over all distributions $\rho \in \Delta\Delta\Omega$ satisfying (BP) and (WC), with no further reference to the underlying signal.

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 best-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 satisfy constraint (WC). In order to understand how this constraint changes the solution, we can further characterize the function $\underset{\smile}{V}(\mu)$. For any belief μ induced by the Sender's signal realization, an adversarial Nature solves a standard Bayesian persuasion problem with μ as a prior, trying to minimize the expected payoff of the Sender. Thus, we can express Nature's problem as an optimization over Bayes-plausible distributions over posterior beliefs (averaging out to μ), with a posterior belief $\eta \in \Delta\Omega$ leading to the expected payoff $\underline{V}(\eta)$ for the Sender. Let $\mathrm{lco}(\underline{V})$ denote the *lower convex closure* of $\underline{V}$, that is, $\mathrm{lco}(\underline{V}) = -\mathrm{co}(-\underline{V})$, where the concave closure $\mathrm{co}(\cdot)$ of a function is defined as in Kamenica and Gentzkow (2011). Then, we conclude that $\underset{\smile}{V}(\mu) = \mathrm{lco}(\underline{V})(\mu)$. The key consequence is that $\underset{\smile}{V}(\mu)$ is a convex function that coincides with $\underline{V}(\mu)$ at Dirac deltas $\mu = \delta_\omega$.

With these observations in mind, we are ready to state our main characterization result. For any function $V : X \to \mathbb{R}$, and $Y \subseteq X$, let $V|_Y$ denote a function defined on the domain Y that coincides with V on Y. Given any measurable set X and any probability measure $\gamma \in \Delta X$ over X, let $\mathrm{supp}(\gamma)$ denote the support of γ, i.e., the smallest closed subset of X whose complement has zero measure under γ.

Theorem 1 (Separation Theorem)**.** *Let*

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

Then,

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

Therefore, $\rho_{RS} \in \Delta\Delta\Omega$ 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

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

Theorem 1 states 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 "allowed" subsets of the state space, as described by the collection $\mathcal{F}$. Moreover, the theorem describes exactly what the allowed subsets are: the subset $B \subseteq \Omega$ is allowed if (and only if) *any* posterior supported on B yields the Sender an expected payoff no smaller than the one the Sender could obtain, starting from μ, by fully disclosing the state. Intuitively, the Sender engages in non-trivial information design only on subsets of the state space on which she "prefers obfuscation to transparency." Importantly, this condition is expressed effectively in terms of the primitives of the model (apart from solving for the best-response correspondence of the Receiver), and in particular 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 support $\text{supp}(\eta) \subseteq \text{supp}(\mu)$, starting from μ, Nature can induce the belief η with positive probability, while respecting the Bayes plausibility constraint (for example, by disclosing the state when not inducing η). 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 Observation 1), any such μ can be part of a worst-case optimal distribution.

For further illustration, consider Example 1 from the Introduction. If the designer induces the belief $\mu = (1/3, 1/3, 1/3)$ in the support of the Bayesian solution, Nature can "split" μ into $\eta = (1/(3-\epsilon), 1/(3-\epsilon), (1-\epsilon)/(3-\epsilon))$ with conditional probability $1-\epsilon/3$ and $\eta' = (0, 0, 1)$ with conditional probability $\epsilon/3$. When the final belief of the judge is η, the defendant is acquitted (since the posterior probability of the defendant being guilty is below 2/3). As a result, the Sender's conditional expected payoff is arbitrarily close to 0. In contrast, the Sender would have received a conditional expected payoff of $\underline{V}_{\text{full}}(\mu) = (1/3)\underline{x} + (1/3)\overline{x}$ by fully disclosing the state. Thus, the posterior belief μ cannot be part of a robust solution. By a similar reasoning, no posterior belief that mixes the state i with some other state from $\{m, f\}$ can be part of a robust solution either, as Nature would always find a way to disclose information to push the Sender's expected payoff strictly below the full-disclosure payoff required for worst-case optimality. A simple calculation confirms that $\mathcal{F} = \{\{i\}, \{m\}, \{f\}, \{m, f\}\}$ for Example 1. Theorem 1 then predicts that a robust solution must reveal the state i, and maximizes the Sender's expected payoff $\widehat{V}$ conditional on state $\{m, f\}$. 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 reveling i and saying nothing in all other states 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 always exists.*

Indeed, the set $\mathcal{W}$ of worst-case optimal distributions is closed, and thus compact (this is 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 conditionally independent signal. When, instead, 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 only two states, Corollary 2 exhausts all possibilities.

Corollary 3 (Complete characterization for binary-state case)**.** *When $\Omega = \{\omega_L, \omega_H\}$, 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).

We can easily extend the conditions for full disclosure to be the unique robust solution, or for robust solutions to coincide with Bayesian solutions, to arbitrary problems.

Corollary 4 (Full disclosure)**.** *Full disclosure is the unique robust solution if $\mathcal{F} = \Omega$, meaning that any pair of states must be separated under any worst-case optimal distribution.*

Corollary 5 (No restrictions)**.** *All feasible distributions are worst-case optimal if, and only if, $\Omega \in \mathcal{F}$, meaning that no pair of states must be separated under any worst-case optimal distribution. Then, the set of robust solutions coincides with the set of Bayesian solutions.*

For an illustration of Corollaries 4-5, 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. If the Sender maximizes the seller's profit, then Corollary 4 applies: If any two states ω and ω' are not separated by the Sender, Nature can ensure that the seller does not extract all the surplus. Thus, $\mathcal{F}$ only contains singletons (i.e., $\mathcal{F} = \Omega$), and full disclosure is the unique robust solution. In contrast, when the Sender maximizes the buyer's surplus, Corollary 5 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.—although quite complicated—is in fact robust.

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 say that $\rho \in \Delta\Delta\Omega$ Blackwell dominates $\rho' \in \Delta\Delta\Omega$ if there exist signals $q : \Omega \to \Delta(\mathcal{S}' \times \mathcal{S})$ inducing ρ and $q' : \Omega \to \Delta\mathcal{S}'$ inducing ρ' such that the marginal distribution of q on $\mathcal{S}'$ conditional on any $\omega \in \Omega$ coincides with that of q'.

Corollary 6 (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 distributions out of the set $\mathcal{W}$.

Suppose that there exists a Bayesian solution that Blackwell dominates a robust solution. Then, by Corollary 6, that Bayesian solution must be worst-case optimal, and hence it is also a robust solution. Therefore, we obtain the following conclusion:

Corollary 7 (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} dominates ρ_{BP}.*

Corollary 7 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 Blackwell order. However, it can never happen that a Bayesian solution strictly Blackwell dominates a maximally informative robust solution.

While the result in Corollary 7 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 7 says that the Sender *herself* will provide more (or at least not less) information than she would in the Bayesian-persuasion model. Second, we show in the Online Appendix that the conclusion of Corollary 7 actually fails in the version of the model where Nature can only send signals that are conditionally independent of the Sender's signal (conditional on the state). The counterexample we provide is based on failure of the property identified in Corollary 6. What makes Corollary 6 true in the baseline model is that Nature can induce *any* mean-preserving spread of the Sender's signal, and thus the Sender cannot improve her worst-case payoff by withholding information.[9]

Corollary 8 (Additional state separation under robust solutions)**.** *If a Bayesian solution ρ_{BP} is not robust and is strictly Blackwell dominated by a robust solution ρ_{RS}, then ρ_{RS} separates states that are not separated under ρ_{BP}.*

The result follows directly from the structure of the set $\mathcal{W}$. If a robust solution ρ_{RS} discloses more information than a Bayesian solution ρ_{BP}, and the latter is not robust, it cannot be that any posterior μ generated by ρ_{RS} has the same support as one of the posteriors generated by ρ_{BP}. It must be that ρ_{RS} separates states that ρ_{BP} does not separate.

Next, 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

[8]By 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.

[9]When, instead, Nature is constrained to conditionally independent signals, there exist mean-preserving spreads of the Sender's signal that Nature cannot induce. Thus, the Sender may choose to withhold information (relative to the Bayesian solution) to limit Nature's ability to induce certain distributions of posterior beliefs with low expected payoffs.

incorporate the constraints into the objective function $\widehat{V}$ by modifying its value on $\Delta^c_{\mathcal{F}}\Omega \equiv \Delta\Omega \setminus \Delta_{\mathcal{F}}\Omega$ (that is, on the set of posterior 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 define

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

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. By the above reasoning, we obtain the following corollary.

Corollary 9 (Concavification)**.** *A feasible distribution $\rho \in \Delta\Delta\Omega$ is a robust solution if and only if $\int \widehat{V}_{\mathcal{F}}(\mu)d\rho(\mu) = co(\widehat{V}_{\mathcal{F}})(\mu_0)$.*

Corollary 9 implies that the problem of finding a robust solution can always 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 sends at most as many signals as there are states, implying in particular that the restriction to finite signal spaces is without loss of optimality for the Sender.

Corollary 10 (Support)**.** *There always exists a robust solution ρ with $|supp(\rho)| \leq |\Omega|$.*

4 Extensions

By direct inspection of the proofs, it is easy to see that *all* the results of the previous section rely only on the following properties of the reduced-form payoffs:

- $\underline{V} : \Delta\Omega \to \mathbb{R}$ is lower semi-continuous;
- $\underset{\smile}{V} : \Delta\Omega \to \mathbb{R}$ is the lower convex closure of $\underline{V}$.[10]

[10]That $\underset{\smile}{V}$ is the lower convex closure of a lower semi-continuous function in turn implies that it is continuous. We prove this property in the Online Appendix (see Appendix OA.2.2).

- $\widehat{V} : \Delta\Omega \to \mathbb{R}$ 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 same 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 the one contained in the common prior. Moreover, she conjectures that, in case of indifference, the Receiver will resolve the indifference in favor of the Sender. Suppose, instead, that the Sender conjectures that Nature will respond to her disclosure with some signal $\pi_0 : \Omega \times \Delta\Omega \to \Delta\mathcal{R}$. That is, when the Sender's signal realization induces a posterior belief μ, the Sender conjectures that the Receiver will observe an additional signal realization r drawn from $\mathcal{R}$ with probability $\pi_0(r|\,\omega,\, \mu)$. The dependence of the distribution $\pi_0(\cdot|\,\omega,\, \mu)$ on μ captures the possibility that the additional information collected by the Receiver may depend on the posterior induced by the Sender's signal realization.[11] Moreover, the Sender conjectures that the Receiver will use a (potentially) stochastic and belief-dependent tie-breaking rule $\xi_0 : \Delta\Omega \to \Delta A$, where $\xi_0(\cdot|\mu')$ is the probability distribution over the Receiver's actions when the final posterior belief is μ', with the property that $\xi_0(A^\star(\mu')|\mu') = 1$, for any $\mu' \in \Delta\Omega$. The Sender's expected payoff from inducing the posterior μ under her conjecture is then equal to

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

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 extra information the Receiver has access to is invariant to the realization of the Sender's signal. This is the case, for example, when the Receiver observes the realization of such extra signal *before*

[11] The above formulation presumes that such an 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 in terms of a distribution over posterior beliefs instead of over signal structures.

observing the realization of the Sender's signal. This conjecture corresponds to the case where $\pi_0(r|\,\omega,\,\mu)$ does not depend on μ. In Section 5, we apply our analysis to 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.[12] Under robust design, instead, the Sender may not trust that the Receivers will play favorably to her. Instead, she may seek a signal structure 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 $\pi(r|\omega,\,\mu_0^s)$. Given the final (common) posterior $\mu_0^{s,r}$ induced by the combination of the Sender's and Nature's signal, the Receivers play some normal- or extensive-form game. For any common posterior $\mu \in \Delta\Omega$, denote by $EQ^\star(\mu)$ the set of strategy profiles that are consistent with the assumed solution concept and the common posterior μ. Finally, let $\xi(\cdot|\mu) \in \Delta EQ^\star(\mu)$ denote the possibly stochastic rule describing the selection of a strategy profile from $EQ^\star(\mu)$.

[12]Notable extensions include Inostroza and Pavan (2018), Li et al. (2020), Mathevet et al. (2020), Morris et al. (2019), and Ziegler (2019).

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 $\xi(\cdot|\mu) \in \Delta EQ^\star(\mu)$ that is least favorable to the Sender. Under regularity conditions, the function $\underline{V}$ is lower semi-continuous. The function $\underset{\smile}{V}$ 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 that $\underset{\smile}{V} = \mathrm{lco}(\underline{V})$.

The Sender's conjecture is that the Receivers will be able to collect public information according to the policy $\pi_0(\cdot|\omega,\, \mu)$, and that, for any final common posterior μ', they will play according to the selection $\xi_0(\cdot|\mu') \in \Delta EQ^\star(\mu')$. The combination of π_0 and ξ_0 is what 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. As in the single-Receiver case, the special case in which π_0 does not depend on μ may be interpreted as reflecting the Sender's belief that the Receivers are endowed with some information (not directly observed by the Sender) prior to observing the realization of the Sender's signal, but are unable to collect any further information after observing the realization of the Sender's signal. It may also be natural for the Sender to conjecture that the selection from $EQ^\star(\mu)$ be governed by a belief-invariant rule, e.g., the one that always picks the Sender's preferred strategy profile, as often assumed in the persuasion literature, or that, for any μ, $\xi_0(\cdot|\mu)$ be uniform over $EQ^\star(\mu)$, reflecting agnostic beliefs over strategy selection.

The case of private disclosures by Nature. Our approach can also accommodate for 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 the additional information provided by Nature becomes blurred. For example, when the solution concept is Bayes Correlated Equilibrium (BCE), 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 additional 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 the strategy profiles generating the lowest payoff for given information, and then looking at different disclosures by Nature.

However, we can bypass the function $\underline{V}$ by formally assuming that $\underline{V} \equiv \underset{\smile}{V}$. The function $\underset{\smile}{V}(\mu)$ is 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 inducing them to play according to the strategy profile that, given the assumed solution concept, minimizes the Sender's expected payoff. Then, $\underset{\smile}{V}$ is trivially the lower convex closure of $\underline{V}$, and it will typically be lower semi-continuous. The Sender's payoff under the assumed conjecture, $\widehat{V}$, is then defined as above, with the exception that the Sender conjecture is now allowed to specify discriminatory disclosures by Nature. Provided that $\widehat{V}$ is upper semi-continuous, then all our results apply. A negative consequence of bypassing $\underline{V}$ (which explains why we have not done it throughout) is that some of the assumptions of the results are more difficult to verify and/or satisfy. For example, to identify the set $\mathcal{F}$ in Theorem 1, one needs to compute $\underset{\smile}{V}$ which can be challenging in some applications (for example, when the assumed solution concept is BCE, this requires characterizing the Sender's payoff in the worst BCE consistent with any given common posterior $\mu \in \Delta\Omega$).[13] However, in certain applications, the set $\mathcal{F}$ can be identified even without computing the entire set of BCE, for any posterior μ. For an illustration, see the application in Section 5.

5 Applications

In this section, we present four applications, illustrating the four cases we have considered: the baseline model, a single Receiver under a general conjecture, and two

[13]In some cases, these challenges might be reduced by defining $\underline{V}(\mu)$ to be the expected payoff to the Sender of inducing a common posterior μ, in the worst equilibrium, when Nature complements μ with purely private signals in the sense formalized by Mathevet et al. (2020). Then, relative to $\underline{V}$, $\underset{\smile}{V}$ captures the effect of additional *public* disclosures by Nature, and thus $\underset{\smile}{V}$ is the lower convex closure of $\underline{V}$. In some applications, this approach can be more tractable, to the extent that computing $\underline{V}$—as defined above—is easier than computing $\underset{\smile}{V}$.

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 trading 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, $\Omega \cap (0, 1) = \Omega$. We are interested in finding the robustly optimal disclosure policy for the seller, under the conjecture that the buyer is completely uninformed about the quality of the good.

The payoff to the seller under the conjecture is given by[15]

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

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.*

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

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

For intuition, notice that when only types ω' and ω are present in the market, 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. At the same time, type ω' does not benefit from the presence of the higher type ω because of adverse selection: $\mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega} \geq p] = \omega'$ 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 $\tilde{\omega} \leq p$. Therefore, Nature can induce posterior beliefs that push the seller's expected payoff below what she could receive by fully disclosing the state. The above reasoning does not apply to types that are less than D apart. This is because the adverse selection problem is mute for such types, as the next lemma shows.

Lemma 3. *Suppose that supp*$(\mu) \subseteq [\underline{\omega}_\mu, \underline{\omega}_\mu + D]$*, where* $\underline{\omega}_\mu$ *is the minimum of supp*(μ)*. 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 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 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 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 1. *Under any robust solution ρ_{RS}, for any $\mu, \mu' \in supp(\rho_{RS})$, $diam(supp(\mu)) \leq D$; $diam(supp(\mu')) \leq D$; but $diam(supp(\mu) \cup supp(\mu')) > D$.*

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 solution must disclose just enough information to neutralize the adverse selection problem.

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 in this example amounts to accepting to trade. The seller can provide any information of her choice 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.

Consider now the situation faced by a seller with the above conjecture who is concerned about the validity of her conjecture and who seeks to protect herself against the worst-case scenario. To avoid uninteresting cases, assume that π_0 is not fully revealing.[16]

In such an environment, given any final posterior $\mu_0^{s,r} \in \Delta\Omega$ for the buyer (obtained by combining the seller's disclosure with the information provided by Nature), we have that the seller's payoff under the least-favorable tie-breaking rule is

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

where $\mathbf{1}(a)$ is the indicator function taking value 1 if the statement a is true and 0 otherwise. Instead, 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.

Proposition 2. *Any robust solution separates any state $\omega \leq p$ from any state $\omega' > p$.*

A robust solution thus essentially removes any 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 pulls 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

In this subsection, we consider 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

[16] That 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.

profiles in the continuation game among the Receivers, respectively.[17]

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" the regime and "not attack" it. Let $a_i = 1$ (respectively, $a_i = 0$) denote the decision by agent i to attack (respectively, not attack) the regime 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 $(-\infty, 0)$, $[0, 1]$, 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, 2018, and Morris et al., 2019 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 the one 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 $\alpha \equiv g/(g + |b|)$ to the event that $\omega > 0$.[18] Now let $\mu_0^+ \equiv \mu_0(\omega > 0)$ 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 in 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

[17] The results, however, do not hinge on public disclosures by nature. The same conclusions obtain when the Sender conjectures that the Receivers are commonly informed but does not rule out the possibility that Nature discloses information privately to the agents.

[18] When, instead, $Pr(\omega > 0) < \alpha$, the unique rationalizable profile is for each agent to attack.

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 that $\omega \in (0,1]$, 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-information 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) plays favorably to the Sender, 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 favorably. The next proposition summarizes the above results.

Proposition 3. *Suppose that the Sender's conjecture is that the agents possess no information other than the one contained in the common prior and that, given any common posterior induced by the Sender, they play the rationalizable profile most favorable to the Sender. 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, instead, $\omega > 0$, the signal $s = \emptyset$ is sent with 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 the results in Corollaries 7 and 8.

5.4 Multiple Receivers and private disclosures by Nature

Consider the following variant of the prosecutor-judge example of Section 1. 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.[19] The defendant is convicted only if both judges deliberate against him (that is, each votes to convict). 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 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,

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

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

As in the original version, the prosecutor maximizes the expected number of years determined by the actual sentence. Her conjecture is that each judge's only information is the one associated with the common prior which, as in the original example, is given by $\mu_0(i) = 1/2$, and $\mu_0(m) = \mu_0(f) = 1/4$. The prosecutor, however, is concerned that her conjecture could be wrong and seeks to protect herself against the worst-case scenario.

It is easy to see that, in this version of the game, the Bayesian solution is the same as in the original version 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

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

the objective function of the prosecutors is exactly 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 $\underset{\smile}{V}$ separately. Instead, we set $\underline{V} = \underset{\smile}{V}$ with $\underset{\smile}{V}(\mu)$ defined as the Sender-inferior BCE payoff consistent with the common posterior belief induced by the Sender being μ. Even though the game between the judges is simple (there is no strategic interaction), computing $\underset{\smile}{V}(\mu)$ for any μ is difficult. Instead, we will make use of Corollary 2: States ω, ω' must be separated by a robust solution whenever, for some $\lambda \in (0,\, 1)$,

$$\underset{\smile}{V}(\lambda\delta_\omega + (1-\lambda)\delta_{\omega'}) < \lambda\, \underset{\smile}{V}(\delta_\omega) + (1-\lambda)\, \underset{\smile}{V}(\delta_{\omega'}).$$

The right-hand side of the above 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. Now 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 the "correct" signal to judge 2 with probability 2/3. Formally, when the state is $\omega = m$ (alternatively, $\omega = f$), with probability 2/3, Nature discloses $r_2 = m$ to judge 2 (alternatively, $r_2 = f$) whereas, with probability 1/3, she discloses $r_2 = f$ (alternatively, $r_2 = m$). 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 (when observing $r_2 = f$, with associated posterior $\mu_2^f = (1/3)\delta_m + (2/3)\delta_f$) and $(1/3)\underline{x} + (2/3)\bar{x}$ with probability 1/3 (when observing $r_2 = m$, with associated posterior $\mu_2^m = (2/3)\delta_m + (1/3)\delta_f$). Clearly, the same outcome can be induced through a BCE.[20] We thus have that

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

Thus, by Corollary 2, states m and f must also be separated by any robust solution. By Corollary 4, full disclosure is then the unique robust solution. This application of Corollary 2 illustrates the force of Theorem 1: We were 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 Relationship to alternative approaches

In this section we explore how robust solutions relate to alternative solution concepts that also account for the Sender's concern for robustness. First, we show how robust solutions can be alternatively thought of as maximizing a convex combination between the Sender's payoff in the worst-case scenario and under her conjecture, with the weight to the worst-case scenario sufficiently large. Next, we establish the relationship between the set of our robust solutions and the set of undominated solutions.

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 policies based on her conjecture. In this section, we examine a more flexible objective function under which the designer attaches a weight $\lambda \in [0, 1]$ to the worst-case scenario, and a weight $1-\lambda$ to the best-case scenario. This approach is

[20] It suffices that, when the state is f, Nature recommends $(a_1, a_2) = (\bar{x}, \bar{x})$ with probability 2/3 and $(a_1, a_2) = (\bar{x}, (1/3)\underline{x} + (2/3)\bar{x})$ with probability 1/3, whereas, when the state is m, she recommends $(a_1, a_2) = (\underline{\bar{x}}, \bar{x})$
with probability 1/3 and $(a_1, a_2) = (\underline{x}, (1/3)\underline{x} + (2/3)\bar{x})$ with probability 2/3.

reminiscent of what is assumed in the literature on alpha-max-min preferences (Hurwicz, 1951, Gul and Pesendorfer, 2015, Grant et al., 2020). A possible interpretation is that the designer is Bayesian, and the weights reflect the assessed probabilities of Nature being adversarial and "favorable," respectively, with favorable interpreted as "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. The result uses the special structure of the persuasion model, and provides a Bayesian foundation for the less standard lexicographic approach.[21] Throughout, we work with reduced-form payoff functions with the properties listed in Section 4.

Formally, for some $\lambda \in [0, 1]$, the designer's problem is stated as follows

$$\sup_{q \in Q} \left\{ \lambda \underset{\smile}{v}(q) + (1-\lambda)\widehat{v}(q) \right\}.$$

Our previous results imply that this is equivalent to

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

subject to (BP). Recall that $\widehat{V}$ is assumed upper semi-continuous, and $\underset{\smile}{V}$ is convex and continuous (as a lower convex closure of a lower semi-continuous function $\underline{V}$). Therefore, the problem for a fixed λ is equivalent to a standard Bayesian persuasion problem with an upper semi-continuous objective function $\widehat{V}_\lambda(\mu) \equiv \lambda \underset{\smile}{V}(\mu) + (1-\lambda)\widehat{V}(\mu)$, and a feasible ρ is a solution if and only if it concavifies $\widehat{V}_\lambda$ at the prior μ_0.

Our goal is to relate the solutions to the problem defined by (6.1) (which we will denote by $S(\lambda)$ and refer to them as λ-solutions) to robust solutions. Note that 0-solutions coincide with Bayesian solutions while 1-solutions are worst-case optimal solutions. Under a regularity condition introduced below, we show that robust solutions are a subset of the worst-case optimal solutions that are also λ-solutions for sufficiently high λ. Let d denote the Chebyshev metric on $\Delta\Omega$: $d(\mu, \eta) = \max_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|$.

Definition 4. The function $\widehat{V}$ is regular if there exist positive constants K and L such that for every non-degenerate $\mu \in \Delta\Omega$ and every $\omega \in \text{supp}(\mu)$, there exists $\eta \in \Delta\Omega$ with $\text{supp}(\eta) \subseteq \text{supp}(\mu) \backslash \{\omega\}$ such that $d(\mu, \eta) \leq K\mu(\omega)$ and $\widehat{V}(\mu) - \widehat{V}(\eta) \leq L d(\mu, \eta)$.

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

Regularity requires that, for any μ and any $\omega \in \text{supp}(\mu)$, there exists a nearby belief supported on $\text{supp}(\mu) \setminus \{\omega\}$ that is not much worse for the designer under the best-case payoff $\widehat{V}$. This only has bite for beliefs μ for which $\mu(\omega)$ is small for some ω; else the condition follows from boundedness of the function $\widehat{V}$. Obviously, Lipschitz continuous functions are regular. However, the condition is much weaker because the Lipschitz condition is required to hold *(i)* only for beliefs μ that attach vanishing probability to some state ω, *(ii)* only for *some* belief η in the neighborhood of a given μ, and *(iii)* only in one direction (the condition rules out functions $\widehat{V}(\mu)$ that *decrease* at an infinite rate as $\mu(\omega)$ approaches 0). And, indeed, regularity allows for highly discontinuous objective functions (we maintain though that $\widehat{V}$ is upper semi-continuous). For example, $\widehat{V}(\mu) = v(\mathbb{E}_\mu[\omega])$ is regular, for any real-valued function v. This is because, when $\mu(\omega)$ is small, one can always find a belief η supported on $\text{supp}(\mu) \setminus \{\omega\}$ with exactly the same mean as μ. A different example is $\widehat{V}(\mu) = \sum_{i=1}^{k} a_i \mathbf{1}_{\{\mu \in A_i\}}$ for some finite partition $(A_1, \ldots, A_k)$ of $\Delta\Omega$; such an objective arises when the Receiver has finitely many actions, and the Sender's preferences are state-independent.

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

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

In the remainder of this section, we describe the key lemmas leading to Theorem 2 (the proof of these lemmas is in the Appendix).

First, we observe that if the designer decides to induce a belief $\mu \in \Delta^c_{\mathcal{F}}\Omega \equiv \Delta\Omega \setminus \Delta_{\mathcal{F}}\Omega$, then we can bound from below the loss that is incurred in the worst-case scenario relative to a worst-case optimal policy.

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

$$\underline{V}_{full}(\mu) - \underset{\sim}{V}(\mu) \geq \delta \cdot \max_{B \subseteq supp(\mu),\, B \notin \mathcal{F}} \min_{\omega \in B} \{\mu(\omega)\}.$$

For regular functions, we can correspondingly bound from above the gains from inducing a belief $\mu \in \Delta^c_{\mathcal{F}}\Omega$ in the best-case scenario. The Sender can always achieve

$\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu)$ without sacrificing worst-case optimality, by Corollary 9. For $\mu \in \Delta^c_{\mathcal{F}}\Omega$, it is possible that $\widehat{V}(\mu) > \mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu)$ but the difference can be upper bounded.

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

$$\widehat{V}(\mu) - co(\widehat{V}_{\mathcal{F}})(\mu) \leq \Delta \max_{B \subseteq supp(\mu),\, B \notin \mathcal{F}} \min_{\omega \in B} \{\mu(\omega)\}.$$

Together, the above two lemmas imply the following result:

Lemma 8. *Suppose that $\widehat{V}$ is regular. There exists $\overline{\lambda} < 1$ such that, for all $\lambda \in (\overline{\lambda}, 1]$, if ρ solves the problem defined by (6.1), then ρ cannot assign positive probability to $\Delta^c_{\mathcal{F}}\Omega$.*

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 (by the definition of $\mathcal{F}$, $\underset{\sim}{V}$ is affine on $\Delta_{\mathcal{F}}\Omega$). As long as the weight $1 - \lambda$ on the best-case scenario is strictly positive, a λ-solution must thus maximize the Sender's payoff in the best-case scenario, conditional on being worst-case optimal, that is, it must be a robust solution.[22]

6.2 Dominance

We finish by examining 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 one policy dominates another if it performs weakly better for *any* choice of Nature's signal and strategy selection, and strictly better for some. Our final result shows that—under certain conditions—any robust solution is guaranteed to be *undominated*.

To define dominance formally, we will again bypass the distinction between information disclosure and strategy selection. We introduce a function $\overline{V}(\mu)$, 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 $\overline{V}$ must be concave under this interpretation (as

[22]Formally, for $\lambda \in (\overline{\lambda}, 1)$, ρ concavifies $\lambda \underset{\sim}{V} + (1 - \lambda)\widehat{V}$ 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 9, ρ is thus a robust solution.

otherwise Nature could further increase the Sender's payoff by concavifying $\overline{V}$ with an additional public signal). The result below, however, applies to any concave function $\overline{V}$ such that $\overline{V} \geq \widehat{V} \geq \underset{\smile}{V}$. If Nature is allowed to respond to any common 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 $\underset{\smile}{V}$ and $\overline{V}$. This motivates the following definition of dominance.

Definition 5. A feasible distribution $\rho \in \Delta\Delta\Omega$ dominates a feasible distribution $\rho' \in \Delta\Delta\Omega$ if, for *any* measurable $V : \Delta\Omega \to \mathbb{R}$ such that $V(\mu) \in [\underset{\smile}{V}(\mu), \overline{V}(\mu)]$ for any μ, 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 feasible distribution ρ is undominated if there exists no feasible 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 any robust solution can be dominated only by another robust solution (by the definition of robustness). In turn, this implies that one can always find at least one robust solution that is undominated. The result in part (b) is more convoluted. Heuristically, if follows from the fact that, given any pair of robust solutions ρ and $\rho^\star$, if, for some feasible response V by Nature, ρ performs strictly better than $\rho^\star$, then one can construct another feasible response V' under which $\rho^\star$ 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, that the two policies are both robust solutions does not impose enough structure on the way they may perform under alternative responses by Nature, leaving the door open to the possibility that one dominates the other. For example, 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 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.

7 Conclusions

We introduced and analyzed a novel solution concept for information design in settings in which the Sender faces uncertainty about the Receivers' sources of information and equilibrium selection. Under the proposed approach, 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 Receivers' play in an adversarial fashion. The Sender then picks an information structure that maximizes her expected payoff under her Bayesian conjecture—much like in the standard persuasion model—but among information structures that are worst-case optimal. Our main technical result identified sets of states that can be present together in one of the induced posteriors and states that must be separated. We showed that robust solutions exist and can be characterized using canonical tools; we qualified in what sense they call for more information disclosure then Bayesian solutions; we argued that, under conditions, robustness guarantees that the solution is undominated; and we illustrated the results with applications to both new and existing models of persuasion.

Throughout the analysis, we restricted attention to the case of *public persuasion* in which the Sender discloses the same information to all 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 relied on the assumption that Nature can engineer any signal: One can ask how the properties of robust solutions change as we impose natural constraints on the set of signals that Nature can entertain. In the Online Appendix, we consider one such case by assuming that Nature's signal must be conditionally independent of the Sender's signal. 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

Au, Pak Hung and Keiichi Kawai, "Competitive Disclosure of Correlated Information," *SSRN Working Paper*, 2018.

Aumann, Robert J and Michael Maschler, *Repeated Games with Incomplete Information*, MIT press, 1995.

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 and Ilya Segal, "Persuasion with Rational Inattention," *WP, Stanford University*, 2018.

Board, Simon and Jay Lu, "Competitive Information Disclosure in Search Markets," *Journal of Political Economy*, 2018, *126* (5), 1965–2010.

Börgers, Tilman, "(No) Foundations of Dominant-Strategy Mechanisms: A Comment on Chung and Ely (2007)," *Review of Economic Design*, 2017, *21*, 73–82.

Cui, Zhihan and Dilip Ravindran, "Competing Persuaders in Zero-Sum Games," *WP, Columbia University*, 2020.

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," *WP, ANU*, 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," *mimeo, Peking University*, 2019.

Hurwicz, Leonid, "Optimality Criteria for Decision Making Under Ignorance," *WP, Cowles Foundation*, 1951.

Inostroza, Nicolas and Alessandro Pavan, "Persuasion in Global Games with Application to Stress Testing," *mimeo, Northwestern University*, 2018.

Kamenica, Emir, "Bayesian Persuasion and Information Design," *Annual Review of Economics, forthcoming.*, 2019.

— **and Matthew Gentzkow**, "Bayesian Persuasion," *American Economic Review*, 2011, *101*, 2590–2615.

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," *mimeo, Princeton University*, 2019.

Laclau, Marie and Ludovic Renou, "Public Persuasion," *mimeo, Paris School of Economics*, 2017.

Li, Fei and Peter Norman, "Sequential Persuasion," *SSRN Working Paper*, 2019.

_, Yangbo Song, and Mofei Zhao, "Global Manipulation by Local Obfuscation," *SSRN Working Paper*, 2019.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei, "Attention Management," *American Economic Review: Insights*, 2019.

Mathevet, Laurent, Jacopo Perego, and Ina Taneva, "On Information Design in Games," *Journal of Political Economy*, 2020, *128* (4), 1370–1404.

Matysková, Ludmila, "Bayesian Persuasion With Costly Information Acquisition," *WP, Bonn University*, 2019.

Morris, Stephen, Daisuke Oyama, and Satoru Takahashi, "Information Design in Binary Action Supermodular Games," *mimeo, MIT*, 2019.

Wu, Wenhao, "Sequential Bayesian Persuasion," *Working Paper*, 2018.

Ye, Lintao, "Beneficial Persuasion," *WP, Washington University in St. Louis*, 2019.

Ziegler, Gabriel, "Adversarial Bilateral Information Design," *mimeo, Northwestern University*, 2019.

A Appendix

A.1 Proof of Lemma 1

Note that we have

$$\mu_0^{s,r}(\omega) = \frac{\pi(r|\,\omega,\, s)q(s|\,\omega)\mu_0(\omega)}{\sum_{\omega'}\pi(r|\,\omega',\, s)q(s|\,\omega')\mu_0(\omega')} = \frac{\pi(r|\,\omega,\, s)\mu_0^s(\omega)}{\sum_{\omega'}\pi(r|\,\omega',\, s)\mu_0^s(\omega')} = (\mu_0^s)^r(\omega),$$

where recall that μ^x is the posterior belief induced by observing signal realization x. Thus,

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

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

Therefore, for any $s \in \mathcal{S}$, Nature's problem of minimizing the Sender's payoff is

$$-\sup_{\pi:\,\Omega\to\Delta\mathcal{R}} \sum_{\omega\in\Omega,\, r\in\mathcal{R}} -\underline{V}\left((\mu_0^s)^r\right)\pi(r|\omega)\mu_0^s(\omega), \tag{A.1}$$

where we suppressed the dependence of π on s in the notation because this problem is solved for every $s \in \mathcal{S}$ separately. The optimization problem (A.1) is a standard 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 take $\mathcal{R} = \Omega$, 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. Using Observation 1 and the definition of $\underset{\sim}{V}$, we have that a signal q is worst-case optimal if and only if

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

and, moreover, $\underset{\sim}{V} = -\text{co}(-\underline{V})$. A distribution ρ of posterior beliefs can be induced by some signal function $q : \Omega \to \Delta\mathcal{S}$ 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). This finishes the proof of Lemma 1.

A.2 Proof of Theorem 1

Let $\mathcal{X} = \{\rho \in \Delta\Delta\Omega :\ \rho \text{ satisfies (BP) and } \text{supp}(\rho) \subseteq \Delta_{\mathcal{F}}(\Omega)\}$. It is enough to prove that $\mathcal{W} = \mathcal{X}$ (the rest of the theorem follows directly from definitions).

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 $A \subset \text{supp}(\rho)$, 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 $\underset{\smile}{V} = \text{lco}(\underline{V})$. Because $\text{lco}(\underline{V}) \leq \underline{V}$, we have that $\underset{\smile}{V}(\eta) < \underline{V}_{\text{full}}(\eta)$. 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

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

where the first inequality follows from convexity of $\underset{\smile}{V}$, the second (strict) inequality from $\underset{\smile}{V}(\eta) < \underline{V}_{\text{full}}(\eta)$ and $\underset{\smile}{V} \leq \underline{V}_{\text{full}}$, and the final equality follows 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 worst-case optimal distribution, it must satisfy $\int \underset{\smile}{V}(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0)$ which, by (BP) and the fact that $\underline{V}_{\text{full}}$ is affine, can also be written as

$$\int \left[\underset{\smile}{V}(\mu) - \underline{V}_{\text{full}}(\mu)\right] d\rho(\mu) = 0. \tag{A.4}$$

Because $\underset{\smile}{V} \leq \underline{V}_{\text{full}}$, we must have $\underset{\smile}{V}(\mu) = \underline{V}_{\text{full}}(\mu)$ for all $\mu \in \text{supp}(\rho)$, contradicting (A.3).

Proof of $\mathcal{W} \supseteq \mathcal{X}$: Suppose that $\rho \in \mathcal{X}$. It suffices to show that (WC) holds, or that $\underset{\smile}{V}(\mu) \geq \underline{V}_{\text{full}}(\mu)$ for all $\mu \in \text{supp}(\rho)$. Fix any $\mu \in \text{supp}(\rho)$. Because $\text{supp}(\rho) \subseteq \Delta_{\mathcal{F}}(\Omega)$, we have that $\underline{V}|_{\text{supp}(\mu)} \geq \underline{V}_{\text{full}}|_{\text{supp}(\mu)}$. Because $\underline{V}$ dominates an affine function $\underline{V}_{\text{full}}$ on $\text{supp}(\mu)$, so does its lower convex closure $\underset{\smile}{V}$. This finishes the proof.

A.3 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 $\underset{\smile}{V}(\mu_B) = \text{lco}(\underline{V})(\mu_B) < \underline{V}_{\text{full}}(\mu_B)$. Then let $\delta_B \equiv \underline{V}_{\text{full}}(\mu_B) - \underset{\smile}{V}(\mu_B)$ and $\delta \equiv \min_{B \notin \mathcal{F}} \delta_B > 0$.

Consider any $\mu \in \Delta^c_{\mathcal{F}}\Omega$. Let $B \subseteq \text{supp}(\mu)$ be such that $B \notin \mathcal{F}$. Then, we can write $\mu = \kappa\mu_B + (1-\kappa)\mu'$ for some μ' and κ, as long as κ is such that, for all $\omega \in \text{supp}(\mu)$, $\mu(\omega) \geq \kappa\mu_B(\omega)$. This equality can be written in particular for $\kappa = \min_{\omega \in B}\{\mu(\omega)\}$. Because $\underline{V}_{\text{full}} - \underset{\smile}{V}$ is a concave function as the difference between an affine function and a convex function, we have

$$\left(\underline{V}_{\text{full}} - \underset{\smile}{V}\right)(\mu) = \left(\underline{V}_{\text{full}} - \underset{\smile}{V}\right)(\kappa\mu_B + (1-\kappa)\mu') \geq$$
$$\kappa\left(\underline{V}_{\text{full}} - \underset{\smile}{V}\right)(\mu_B) + (1-\kappa)\left(\underline{V}_{\text{full}} - \underset{\smile}{V}\right)(\mu') \geq \min_{\omega\in B}\{\mu(\omega)\}\delta_B \geq \min_{\omega\in B}\{\mu(\omega)\}\delta.$$

Since B was arbitrary, we also have that $\left(\underline{V}_{\text{full}} - \underset{\smile}{V}\right)(\mu) \geq \delta \cdot \max_{B\subseteq \text{supp}(\mu),\, B\notin\mathcal{F}} \min_{\omega\in B}\{\mu(\omega)\}$.

A.4 Proof of Lemma 7

Before proving Lemma 7, we first prove that regularity implies a seemingly stronger property that will be more convenient to work with.

Property 1. *If the function $\widehat{V}$ is regular, then there exist positive constants K and L such that for every non-degenerate $\mu \in \Delta\Omega$ and every set $A \subsetneq$ supp(μ), there exists $\eta \in \Delta\Omega$ with supp$(\eta) \subseteq A$ such that $d(\mu, \eta) \leq K \max_{\omega\in supp(\mu)\setminus A}\{\mu(\omega)\}$ and $\widehat{V}(\mu) - \widehat{V}(\eta) \leq Ld(\mu, \eta)$.*

Proof of Property 1. The proof is by induction. If the set A is equal to $\text{supp}(\mu)\setminus\{\omega\}$ for some $\omega \in \text{supp}(\mu)$, then the conclusion follows directly from the definition of regularity. This means that we have proven the property for the case $|\text{supp}(\mu)\setminus A| = 1$.

Induction step: Suppose that we have proven the property for all sets A such that $|\text{supp}(\mu)\setminus A| = k$. Next, we prove it for sets A with $|\text{supp}(\mu)\setminus A| = k+1$.

Concretely, suppose that we have a set $A \subsetneq \text{supp}(\mu)$ with $|\text{supp}(\mu)\setminus A| = k+1$. To simplify notation, let $\delta^A := \max_{\omega\in\text{supp}(\mu)\setminus A}\{\mu(\omega)\}$. Define $A' = A\cup\{\omega^\star\}$ for some $\omega^\star \in \text{supp}(\mu)\setminus A$. By the inductive hypothesis, there exists $\eta' \in \Delta\Omega$ with $\text{supp}(\eta') \subseteq A'$ such that $d(\mu, \eta') \leq K\max_{\omega\in\text{supp}(\mu)\setminus A'}\{\mu(\omega)\}$ and $\widehat{V}(\mu) - \widehat{V}(\eta') \leq Ld(\mu, \eta')$.

Next, we apply the definition of regularity to the measure η' and the state $\omega^\star$: There exists η with $\text{supp}(\eta) \subseteq \text{supp}(\eta')\setminus\{\omega^\star\} \subseteq A$ such that $d(\eta', \eta) \leq K\eta'(\omega^\star)$ and $\widehat{V}(\eta') - \widehat{V}(\eta) \leq Ld(\eta', \eta)$.

Since $d(\mu, \eta') \leq K\delta^A$ and $\mu(\omega^\star) \leq \delta^A$ (because $\omega^\star \in \text{supp}(\mu)\setminus A$), it follows that $\eta'(\omega^\star) \leq (K+1)\delta^A$. Thus, we have

$$d(\mu, \eta) \leq d(\mu, \eta') + d(\eta', \eta) \leq K\delta^A + K(K+1)\delta^A \leq K(K+2)\delta^A,$$

and

$$\widehat{V}(\mu) - \widehat{V}(\eta) = \widehat{V}(\mu) - \widehat{V}(\eta') + \widehat{V}(\eta') - \widehat{V}(\eta) \leq L(d(\mu, \eta') + d(\eta', \eta))$$
$$\leq LK(K+2)\delta^A \leq LK(K+2)d(\mu, \eta),$$

where the last inequality follows from the fact that $\mathrm{supp}(\mu) \setminus \mathrm{supp}(\eta)$ contains some ω that has probability δ^A under μ (and 0 under η). Therefore, we obtain the inductive hypothesis with constants $K' = K(K+2)$ and $L' = LK(K+2)$. □

Now we prove Lemma 7: We have to show that there exists a constant $\Delta > 0$ such that for any $\mu \in \Delta^c_{\mathcal{F}}\Omega$,

$$\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) + \Delta \max_{B \subseteq \mathrm{supp}(\mu),\, B \notin \mathcal{F}} \min_{\omega \in B} \{\mu(\omega)\} \geq \widehat{V}(\mu). \tag{A.5}$$

Let $\bar{\delta} \equiv \max_{B \subseteq \mathrm{supp}(\mu),\, B \notin \mathcal{F}} \ \min_{\omega \in B}\{\mu(\omega)\}$. By definition of $\bar{\delta}$, there must exist a set $A \subsetneq \mathrm{supp}(\mu)$, with $A \in \mathcal{F}$, such that for all $\omega \in \mathrm{supp}(\mu) \setminus A$, $\mu(\omega) \leq \bar{\delta}$. To see that, let $C \equiv \{\omega \in \mathrm{supp}(\mu) : \mu(\omega) > \bar{\delta}\}$. Clearly, if $C = \emptyset$, then it suffices to let A coincide with any element of $\mathrm{supp}(\mu)$. If, instead, $C \neq \emptyset$, then let $A = C$. We claim that A defined this way belongs to $\mathcal{F}$. If that was not the case, from the definition of $\bar{\delta}$, we would have that $\bar{\delta} \geq \min_{\omega \in A}\{\mu(\omega)\} > \bar{\delta}$, a contradiction.

By Property 1 applied to μ and the set A (which we can apply since $\widehat{V}$ is regular), there must exist η with $\mathrm{supp}(\eta) \subseteq A$, $d(\mu, \eta) \leq K \max_{\omega \in \mathrm{supp}(\mu) \setminus A}\{\mu(\omega)\} \leq K\bar{\delta}$, such that

$$\widehat{V}(\mu) - \widehat{V}(\eta) \leq Ld(\mu, \eta) \leq LK\bar{\delta}. \tag{A.6}$$

Importantly, $\mathrm{co}(\widehat{V}_{\mathcal{F}})(\eta) \geq \widehat{V}(\eta)$ because $\mathrm{supp}(\eta) \subseteq A \in \mathcal{F}$. Therefore,

$$\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) + \Delta\,\bar{\delta} \geq \mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) - \mathrm{co}(\widehat{V}_{\mathcal{F}})(\eta) + \widehat{V}(\eta) + \Delta\,\bar{\delta}.$$

On the line segment connecting μ and η, $\mathrm{co}(\widehat{V}_{\mathcal{F}})$ is affine: Indeed, we have that $\widehat{V}_{\mathcal{F}}(\kappa\mu + (1-\kappa)\eta) = v_{\mathrm{low}}$ for any $\kappa > 0$, since any such belief $\kappa\mu + (1-\kappa)\eta \notin \mathcal{F}$; but this implies that $\widehat{V}$ lies strictly below its concave closure (except possibly at η), and hence is affine. This means in particular that $\mathrm{co}(\widehat{V}_{\mathcal{F}})$ is Lipschitz continuous on that segment, that is, for some constant N, $\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) - \mathrm{co}(\widehat{V}_{\mathcal{F}})(\eta) \geq -Nd(\mu, \eta)$. Therefore, using (A.6) and $d(\mu, \eta) \leq K\bar{\delta}$, we have that

$$\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) + \Delta\,\bar{\delta} \geq -Nd(\mu, \eta) + \widehat{V}(\eta) + \Delta\,\bar{\delta} \geq \widehat{V}(\mu) + (\Delta - NK - LK)\bar{\delta}.$$

Thus, to prove the desired inequality (A.5), it is enough to set $\Delta = NK + LK$.

A.5 Proof of Lemma 8

It is enough to prove that, for high enough λ, if $\text{supp}(\rho) \nsubseteq \Delta_{\mathcal{F}}\Omega$, then the Sender's objective $\int \left[\lambda \underset{\smile}{V}(\mu) + (1-\lambda)\widehat{V}(\mu)\right] d\rho(\mu)$ increases strictly by splitting any $\mu \in \text{supp}(\rho)$ such that $\mu \in \Delta^c_{\mathcal{F}}\Omega$ 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 $\delta > 0$,

$$\begin{aligned}
&\left[\lambda \underline{V}_{\text{full}}(\mu) + (1-\lambda)\text{co}(\widehat{V}_{\mathcal{F}})(\mu)\right] - \left[\lambda \underset{\smile}{V}(\mu) + (1-\lambda)\widehat{V}(\mu)\right] \\
&\qquad = \lambda \left[\underline{V}_{\text{full}}(\mu) - \underset{\smile}{V}(\mu)\right] + (1-\lambda)\left[\text{co}(\widehat{V}_{\mathcal{F}})(\mu) - \widehat{V}(\mu)\right] \\
&\qquad \geq (\lambda\delta - (1-\lambda)\Delta) \max_{B \subseteq \text{supp}(\mu),\, B \notin \mathcal{F}} \min_{\omega \in B} \{\mu(\omega)\} > 0
\end{aligned}$$

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

A.6 Proof of Theorem 3

Part (a). Let $S^\star$ be the set of robust solutions (understood to be distributions over posterior beliefs). This set is non-empty and closed (by Berge's theorem), hence compact in the weak* topology. Note that if an element $\rho^\star$ of $S^\star$ is dominated, it must be dominated by another element of $S^\star$ (Indeed, a policy that is not a robust solution cannot dominate $\rho^\star$ because, by definition, it is either strictly worse in the worst case, or is strictly worse under the conjecture – by assumption $V = \underset{\smile}{V}$ and $V = \widehat{V}$ are both feasible choices by Nature). Note also that it is enough to show that at least one robust solution is undominated when Nature is restricted to responding with feasible functions V that are upper semi-continuous (this is because any feasible V that is not upper semi-continuous can be approximated point-wise by a sequence of feasible and upper semi-continuous functions). Denote this set of functions by $\mathcal{P}$.

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 $B^\star(V) \subset S^*$ be the set of S^* constructed inductively as follows. Let V_0 be the lowest element of S^* according to the order $\sqsubset$. Then let

$$B^\star(V_0) := \underset{\rho \in S^\star}{\text{argmax}} \left\{ \int V_0(\mu) d\rho(\mu) \right\},$$

that is, the subset of robust solutions that are optimal for the Sender against V_0. The set $B^\star(V_0)$ is non-empty and closed (and hence compact in the weak* topology) because V_0 is upper semi-continuous and $S^\star$ is non-empty and compact. For any $V \in \mathcal{P}$, then let

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

and

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

The sets $B^\star(V)$ are nested, in the sense that $B^\star(V') \subseteq B^\star(V)$ 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 $\bigcap_{V \in \mathcal{P}} B^\star(V) \neq \emptyset$ and any $\rho^\star \in \bigcap_{V \in \mathcal{P}} B^\star(V)$ is an undominated robust solution.

Part (b). We now establish that, when $\text{co}\widehat{V} = \overline{V}$, any robust solution is undominated. Pick any robust solution $\rho^\star$. Again, it suffices to show that $\rho^\star$ is not dominated by any other robust solution ρ. By Corollary 9, any robust solution achieves $\text{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$ under the conjecture, which corresponds to Nature selecting $V = \widehat{V}$. Suppose first that there exists $\mu \in \Delta\Omega$ such that $\text{co}(\widehat{V}_{\mathcal{F}})(\mu) > \underset{\smile}{V}(\mu)$ and $\rho^\star(\mu) \neq \rho(\mu)$. There are two subcases: Either *(a)* $\rho^\star(\mu) > \rho(\mu)$ or *(b)* $\rho^\star(\mu) < \rho(\mu)$.

In case *(a)*, specify the following response for Nature: Conditional on μ, Nature behaves according to the Sender's conjecture; in all other cases, Nature is adversarial. That is, $V(\mu) = \widehat{V}(\mu)$, and $V(\mu') = \underset{\smile}{V}(\mu')$ for all $\mu' \neq \mu$. Since μ is induced in some robust solution (that is, $\mu \in \text{supp}(\rho^\star) \cup \text{supp}(\rho)$), by Corollary 9, it must be that $\widehat{V}(\mu) = \text{co}(\widehat{V}_{\mathcal{F}})(\mu)$. Thus, under this response of Nature, the Sender's expected payoff under a robust solution $\rho' \in \{\rho^\star, \rho\}$ is given by

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

In a robust solution, by Lemma 1, we have $\int \underset{\smile}{V}(\mu') d\rho'(\mu') = \underline{V}_{\text{full}}(\mu_0)$, and thus the difference in expected payoffs between $\rho^\star$ and ρ is

$$[\rho^\star(\mu) - \rho(\mu)] \left[\text{co}(\widehat{V}_{\mathcal{F}})(\mu) - \underset{\smile}{V}(\mu) \right] > 0,$$

by assumption. Thus, ρ does not dominate $\rho^\star$.

In case *(b)*, let Nature be adversarial conditional on μ, and behave according to the Sender's conjecture in all other cases. Under this response of Nature, the expected payoff under a robust solution $\rho' \in \{\rho^\star, \rho\}$ is

$$\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu_0) - \rho'(\mu)[\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) - \underaccent{\smile}{V}(\mu)].$$

To see this, recall that, when Nature responds to any induced posterior with $\widehat{V}$, then ρ' generates an expected payoff equal to $\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$ – this follows directly from the fact that ρ' is a robust solution.[23] Conditional on inducing μ (which has probability $\rho'(\mu)$), instead of $\widehat{V}(\mu) = \mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu)$, the Sender gets $\underaccent{\smile}{V}(\mu)$. Thus, the difference in expected payoffs between $\rho^\star$ and ρ is given by

$$[\rho(\mu) - \rho^\star(\mu)]\left[\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) - \underaccent{\smile}{V}(\mu)\right] > 0,$$

by assumption. Thus, ρ does not dominate $\rho^\star$ also in this case.

The final case to consider is when there exists no $\mu \in \Delta\Omega$ such that $\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) > \underaccent{\smile}{V}(\mu)$ and $\rho^\star(\mu) \neq \rho(\mu)$. Put differently, for any μ such that $\rho^\star(\mu) \neq \rho(\mu)$ (such a μ must exist because otherwise the two signals would coincide), we must have $\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) = \underaccent{\smile}{V}(\mu)$ (since we always have $\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu) \geq \underaccent{\smile}{V}(\mu)$). Note, however, that $\mathrm{co}(\widehat{V}_{\mathcal{F}})$ is a concave function while $\underaccent{\smile}{V}$ is a convex function, and thus they can be equal at μ if and only if they are both affine functions on $\Delta(\mathrm{supp}(\mu))$: In fact, we must have $\widehat{V} = \underaccent{\smile}{V} = \underline{V}_{\text{full}}$ on $\Delta(\mathrm{supp}(\mu))$. Moreover, because $\widehat{V}$ is affine on $\Delta(\mathrm{supp}(\mu))$, we have that $\mathrm{co}\widehat{V}(\mu) = \widehat{V}(\mu)$ for any such μ. Finally, using the assumption of Theorem 3 that $\mathrm{co}\widehat{V} = \overline{V}$, we conclude that $\overline{V} = \underaccent{\smile}{V}$ on $\Delta(\mathrm{supp}(\mu))$. But this means that *any* V that Nature can select is affine on $\Delta(\mathrm{supp}(\mu))$. This implies that Nature's response conditional on any such μ is payoff-equivalent for the Sender: The Sender's payoff is the same irrespective of the signal and the strategy profile (compatible with the assumed solution concept) selected by Nature in response to any such μ. Because this is true for any μ at which $\rho^\star$ 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 $\rho^\star$.

[23] In fact, from Corollary 9, $\int \widehat{V}_{\mathcal{F}}(\mu)d\rho'(\mu) = \mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$. The property then follows from the fact that, for any $\mu' \in \mathrm{supp}(\rho')$, $\widehat{V}(\mu') = \widehat{V}_{\mathcal{F}}(\mu')$.

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 $v(\lambda) \equiv \underline{V}|_{\Delta B}(\lambda\delta_\omega + (1-\lambda)\delta_{\omega'})$. 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, $\mathbf{1}_{\{\mathbb{E}_\mu[\tilde\omega|\tilde\omega\le p]+D>p\}} \le \mathbf{1}_{\{\mathbb{E}_\mu[\tilde\omega]+D>p\}}$. Suppose that the inequality is strict for some $p \ge \underline{\omega}_\mu : \mathbb{E}_\mu[\tilde\omega] + D > p$ but $\mathbb{E}_\mu[\tilde\omega|\tilde\omega \le p] + D \le p$. This is only possible when $p < \overline{\omega}_\mu$, where $\overline{\omega}_\mu$ is the maximum of $\text{supp}(\mu)$. But then

$$p \ge \mathbb{E}_\mu[\tilde\omega|\tilde\omega \le p] + D \ge \underline{\omega}_\mu + D \ge (\overline{\omega}_\mu - D) + D = \overline{\omega}_\mu > p,$$

a contradiction.

OA.1.3 Proof of Lemma 4

By Lemma 3, we can write

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

Let $B = \{\omega_1, ..., \omega_n\}$ with $\omega_1 < \omega_2 < ... < \omega_n$, 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 a function $\widetilde{\underline{V}}$ defined by $\widetilde{\underline{V}}(\mu_2, ..., \mu_n) = \underline{V}(1 - \mu_2 - ... - \mu_n, \mu_2, ..., \mu_n)$ has a negative semi-definite hessian. By a direct calculation, denoting $\omega_{-1} = [\omega_2, ..., \omega_n]$, we obtain Hessian$(\widetilde{\underline{V}}) = -(\omega_{-1} - \omega_1)^T \cdot (\omega_{-1} - \omega_1)$, which is a negative semi-definite matrix (of rank 1).

OA.1.4 Proof of Proposition 3

Given any $\mu \in \Delta\Omega$, let $\mu^+ \equiv \mu(\omega > 0)$ denote the probability that μ assigns to the event that $\omega > 0$. In this application, the upper selection features all agents attacking if $\mu^+ < \alpha$, and all agents refraining from attacking if $\mu^+ \geq \alpha$, where $\alpha \equiv g/(g + |b|)$, implying that $\widehat{V}(\mu) = 0$ if $\mu^+ < \alpha$ and $\widehat{V}(\mu) = 1$ 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 $Pr(\omega > 0|s = 0) = \mu_0^+/[\mu_0^+ + (1-\mu_0^+)(1-\phi_{BP})] = \alpha$. 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 signal $s = 0$ is disclosed, $Pr(\omega > 0|s = 0) \geq \alpha$ and all agents refrain from attacking, whereas when signal $s = 1$ is disclosed, $Pr(\omega > 0|s = 1) < \alpha$ and all agents attack.[24] Next note that, starting from any binary policy sending signal $s = 1$ with positive probability over a positive measure subset of $\mathbb{R}_+$, one can construct another binary policy that sends signal $s = 0$ (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 send signal $s = 0$ with certainty for all $\omega > 0$. Furthermore, under any Bayesian solution, the ex-ante

[24]The 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.

probability $\int_{-\infty}^{0^-} \pi(0|\omega)d\mu_0(\omega)$ that signal $s = 0$ is sent when $\omega < 0$ is uniquely pinned down by the condition $Pr(\omega > 0|s = 0) = \mu_0^+/[\mu_0^+ + \int_{-\infty}^{0^-} \pi(0|\omega)d\mu_0(\omega)] = \alpha$. Because the Sender's preferences depend only on $1-A$, the specific way the policy sends signal $s = 0$ when $\omega < 0$ is irrelevant, thus implying that the binary policy described above is indeed a Bayesian solution. By the same token, it is also easy to see that the above binary policy is payoff-equivalent to one that sends signal $s = 0$ with certainty when $\omega > 0$, whereas, when $\omega < 0$, with probability ϕ_{BP} fully reveals the state and with the complementary probability sends signal $s = 0$. Signal $s = 0$ can then be interpreted as the "null" signal $s = \emptyset$ as claimed in the proposition.

To see that the above Bayesian policy is not robust, let $\mu^{(0,1]} \equiv \mu(\omega \in (0, 1])$ denote the probability that μ assigns to the interval $(0, 1]$. Recall that, given any posterior μ, if $\mu^+ \equiv \mu(\omega > 0) < \alpha$, 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 $\bar{\Omega} \equiv \{\omega \in \Omega : \omega \leq 1\}$, 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 sends signal $s = 0$ with certainty. When, instead, $\omega \leq 0$, with probability $\phi_{RS} > \phi_{BP}$, the Sender fully reveals the state, and with the complementary probability $1 - \phi_{RS}$, the Sender sends signal $s = 0$. 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 sends the signal $s=\emptyset$. When $\omega \in (0,1]$, the Sender discloses the signal $s=\emptyset$ 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 $n \to \infty$, 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,

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

When combined with Lemma 6, the above inequality implies that

$$\int_{\Delta\Omega}\widehat{V}(\mu)d\rho_n(\mu)-\mathrm{co}(\widehat{V}_{\mathcal{F}})(\mu_0) \geq \frac{\lambda_n}{1-\lambda_n}\cdot\delta\cdot\int_{\Delta^c_{\mathcal{F}}\Omega}\left[\max_{B\subseteq \mathrm{supp}(\mu),\, B\notin\mathcal{F}}\ \min_{\omega\in B}\{\mu(\omega)\}\right]d\rho_n(\mu), \tag{OA.1}$$

where $\Delta^c_{\mathcal{F}}\Omega$ denotes the complement of $\Delta_{\mathcal{F}}\Omega$. Because the left hand side of the above inequality is bounded, and $\lambda_n/(1-\lambda_n)$ diverges to infinity, we must have that

$$\int_{\Delta^c_{\mathcal{F}}\Omega}\left[\max_{B\subseteq \mathrm{supp}(\mu),\, B\notin\mathcal{F}}\ \min_{\omega\in B}\{\mu(\omega)\}\right]d\rho_n(\mu) \to 0.$$

Because the set of possible supports is finite (since Ω is finite), this implies that for any $A \subset \Omega$ such that $A \notin \mathcal{F}$,

$$\int_{\{\mu\in\Delta\Omega:\, \mathrm{supp}(\mu)=A\}}\left[\max_{B\subseteq A,\, B\notin\mathcal{F}}\ \min_{\omega\in B}\{\mu(\omega)\}\right]d\rho_n(\mu) \to 0.$$

On the set $\{\mu \in \Delta\Omega : \operatorname{supp}(\mu) = A\}$ the function $\max_{B \subseteq A,\, B \notin \mathcal{F}} \ \min_{\omega \in B}\{\mu(\omega)\}$ is continuous, bounded, and strictly positive. By definition of convergence in the weak* topology, we have,

$$\int_{\{\mu \in \Delta\Omega:\, \operatorname{supp}(\mu)=A\}} \left[\max_{B \subseteq A,\, B \notin \mathcal{F}} \ \min_{\omega \in B}\{\mu(\omega)\}\right] d\rho(\mu) = 0.$$

Because the integrand is strictly positive, we must have $\rho\left(\{\mu \in \Delta\Omega : \operatorname{supp}(\mu) = A\}\right) = 0$. Because this is true for any $A \notin \mathcal{F}$, and there are finitely many such A, this implies that $\operatorname{supp}(\rho) \subseteq \Delta_{\mathcal{F}}\Omega$, and thus ρ is worst-case optimal.

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

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

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 $\operatorname{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$ is the upper bound on the best-case payoff that a worst-case optimal distribution can yield. This, however, means that $\int_{\Delta\Omega} \widehat{V}(\mu) d\rho(\mu) = \operatorname{co}(\widehat{V}_{\mathcal{F}})(\mu_0)$, and thus ρ is a robust solution, by Corollary 9. $\square$

Next, we show that, without the regularity condition, 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 $x \to 0$, $\widehat{V}$ 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 $\underline{V}$, 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 at which $\widehat{V} = 0$. However, we will show that in the limit as $\lambda \to 1$, all λ-solutions must put positive (bounded away from zero) mass on the belief $(1, 0, 0)$.

Therefore, the distribution ρ_{RS} that puts mass $1/3$ on $(1/2, 1/2, 0)$ and on $(1/2, 0, 1/2)$ and mass $1/6$ on $(0, 1, 0)$ and on $(0, 0, 1)$ is a robust solution but is not a limit of λ-solutions.

Note first that $\underset{\sim}{V}(\mu) = \text{lco}(\underline{V})(\mu) = -3 \min_{\omega} \mu(\omega)$. 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 $\widehat{V}_{\lambda}(\mu) \equiv \lambda \underset{\sim}{V}(\mu) + (1-\lambda)\widehat{V}(\mu)$ is strictly concave on that line segment, a λ-solution attaches the entire weight m to a single $x^{\star}$. For a fixed λ, the optimal choice of $x^{\star}$ is

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

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

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

subject to the Bayes-plausbility constraint. Because the above function is increasing in m, any λ-solution, $\lambda < 1$, puts probability $m^{\star}$ to the the belief $(1-2x^{\star},\, x^{\star},\, x^{\star})$, where $m^{\star} \geq 1/3$ is the largest value of m consistent with Bayes plausibility. Next observe that $(1-2x^{\star}, x^{\star}, x^{\star})$ converges to $(1, 0, 0)$ as $\lambda \to 1$. 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 Convex lower semi-continuous functions on $\Delta\Omega$ are continuous

Lemma OA.1. *A convex lower semi-continuous function $V : \Delta\Omega \to \mathbb{R}$ is continuous.*

Proof. Suppose not. Then, since the function is lower semi-continuous, there must exist a sequence $\mu_n \to \mu$ such that $\liminf V(\mu_n) > V(\mu)$. We can write $\mu_n = \kappa_n \mu + (1-\kappa_n)\zeta_n$, where

$$\kappa_n = \min_{\omega \in \text{supp}(\mu)} \min\left\{\frac{\mu_n(\omega)}{\mu(\omega)},\, \frac{1-\mu_n(\omega)}{1-\mu(\omega)}\right\},$$

and

$$\zeta_n = \frac{\mu_n - \kappa_n \mu}{1 - \kappa_n}.$$

Because $\mu_n \to \mu$ and Ω is finite, we have that $\kappa_n < 1$ and $\kappa_n \to 1$. Thus, ζ_n is a well-defined probability measure for high enough n because, for any $\omega \in \Omega$,

$$0 \leq \frac{\mu_n(\omega) - \kappa_n \mu(\omega)}{1 - \kappa_n} \leq 1.$$

By convexity of V, we have

$$V(\mu_n) = V(\kappa_n \mu + (1-\kappa_n)\zeta_n) \leq \kappa_n V(\mu) + (1-\kappa_n) V(\zeta_n) \leq \kappa_n V(\mu) + (1-\kappa_n) M \overset{n\to\infty}{\to} V(\mu),$$

where M is an upper bound on the value of the function V that can be defined as $M = \max_{\omega \in \Omega} V(\delta_\omega)$ given that the function is convex. This is a contradiction with $\liminf V(\mu_n) > V(\mu)$. $\square$

OA.2.3 Example showing that Bayesian solutions can be dominated

Consider the following conjecture $\widehat{V}$ (equal to $\underline{V}$) defined over the set $[0,1]$ of posteriors over a binary state, with prior $\mu_0 = 1/2$: $\widehat{V}(\mu) = \left(|\mu - \frac{1}{2}| - \frac{1}{4}\right)^2$. 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 $\underset{\smile}{V} = \text{lco}\widehat{V}$ in the definition 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 easy to see which step of the proof of Theorem 3 fails for Bayesian solutions: In case (a) of that proof, we relied on Lemma 1 to argue that for a robust solution ρ, $\int \underset{\smile}{V}(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0)$, 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 Conditionally independent 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's signal must be conditionally independent (conditional on the state) of the Sender's signal. This assumption might be appropriate for settings in which Nature's move reflects the Sender's ambiguity over the information the Receivers might 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 will work with the baseline model of Section 2, except that we will allow for general conjectures (as explained in Section 4; it will be clear that the results also extend to the multi-Receiver case). Unless specified otherwise, we maintain all the assumptions imposed in the main text.

The Sender continues to choose an information structure $q : \Omega \to \Delta\mathcal{S}$ which maps states ω into probability distributions over signal realizations $s \in \mathcal{S}$, but we no longer assume that $\mathcal{S}$ is finite (this would be with loss of generality). We also modify Nature's strategy space: Nature selects a signal $\pi : \Omega \to \Delta\mathcal{R}$ 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 new set of signals available to Nature.[25]

The best-case payoff $\widehat{v}(q)$ from selecting a signal q is computed under the conjecture that Nature selects some fixed (conditionally independent) signal $\pi_0 : \Omega \to \Delta\mathcal{R}$:

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

where ξ_0 is the conjectured tie-breaking rule, with $\xi_0(A^\star(\mu)|\mu) = 1$ for all μ. 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 $\pi_{CI} \in \Pi_{CI}$ (π_{CI} is not a function of the posterior belief generated by the Sender). Throughout, we assume that $\widehat{V}$ is upper semi-continuous. Similarly, the Sender's worst-case payoff is defined as

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

[25] As before, we assume that $\mathcal{R}$ and $\mathcal{S}$ are subsets of some sufficiently rich but fixed space $\mathcal{X}$.

with $\underline{V}$ defined as before. With that modification, the definition of a worst-case optimal signal (Definition 1) remains the same. To distinguish between the two solution concepts, we call signals that are optimal in the worst case over all Nature's signals that are conditionally independent *CI-worst-case optimal*. We use W_{CI} to denote the set of CI-worst-case optimal signals. Observation 1 remains valid: Full disclosure is always CI-worst-case optimal, and a signal is in W_{CI} if and only if it achieves the full-disclosure payoff in the worst case. Then, we define a *CI-robust solution* analogously to Definition 2: A signal q is a CI-robust solution if it maximizes $\widehat{v}(q)$ over W_{CI}.

OA.3.1 Summary of results

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

Characterizing CI-robust solutions turns out to be significantly more complicated than characterizing robust solutions. In particular, the restrictions imposed by CI-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 CI-robust solutions, i.e., a CI-robust solution may fail to exist. We show in Subsection OA.3.3 (Theorem OA.2) that a CI-robust solution exists under a stronger assumption of continuity of $\underline{V}$. Moreover, we introduce a notion of weak CI-robust solutions (that relaxes the condition of CI-worst-case optimality), and show that a weak CI-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 CI-robust solution. This condition is weaker than the one in Corollary 2; that is, whenever two states must be separated under a CI-robust solution, they also must be separated under a robust solution.

- Corollaries 4, and 5 do not extend to CI-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 CI-robust solution, or for all distributions to be CI-worst-case optimal.

- In Subsection OA.3.4, we analyze the binary-state case. Unlike robust solutions, as described by Corollary 3, CI-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 CI-robust solutions.

- In Subsection OA.3.6, we show that Corollary 6 and Corollary 7 fail for CI-robust solutions. That is, it is possible that a Bayesian solution is strictly more informative than all CI-robust solutions.

- Corollaries 8 and 9 also fail: In fact, a CI-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 CI-robust solution.

Lemma OA.2. *The set of CI-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 $\pi : \Omega \to \Delta\mathcal{R}$,

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

where

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

Therefore,

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

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 $\int_{\text{supp}(\nu)} \mu d\nu(\mu) = \mu_0$. 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

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

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 the Bayesian persuasion literature). Thus, it is without loss of generality to set $\mathcal{R} = \Omega$ to find all CI-worst-case optimal signals. □

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

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

and hence

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

where $Q_{\mu_0,q} \in \Delta\mathcal{S}$ is the unconditional distribution over $\mathcal{S}$ induced by μ_0 and q. Recall that a distribution $\rho \in \Delta\Delta\Omega$ is feasible if it satisfies the Bayes plausibility constraint (BP). Therefore, the problem of finding a CI-robust solution is equivalent to the problem of finding a feasible $\rho \in \Delta\Delta\Omega$ that maximizes $\int \widehat{V}(\mu)d\rho(\mu)$ among all CI-worst-case optimal distributions, that is, among all distributions that satisfy

$$\min_{\{\pi:\,\Omega \to \Delta\Omega\}} \int_{\mathrm{supp}(\rho)} \underline{V}_\pi(\mu)d\rho(\mu) = \underline{V}_{\mathrm{full}}(\mu_0). \tag{OA.1}$$

As before, we will abuse terminology slightly by calling ρ the CI-robust solution.

Condition (OA.1), contrasted with Lemma 1, highlights the difference between worst-case optimality and CI-worst-case optimality. In Lemma 1, the minimum operator is inside the integral, i.e., it is computed posterior by posterior. For CI-worst-case optimality, instead, the minimum operator is outside the integral, and Nature's problem involves a trade-off because it 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 CI-robust solution cannot be guaranteed. Example OA.2 illustrates the difficulty.

Example OA.2. Suppose the state is binary, $\Delta\Omega = [0, 1]$, $\mu \in [0, 1]$ is the probability that the state is 1, and $\mu_0 = 1/2$. Define

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

and let $\widehat{V}$ and $\underline{V}$ be, respectively, the point-wise highest and lowest selections from the correspondence $\mathcal{V}$. Then, $\widehat{V}$ is continuous, whereas $\underline{V}$ has a discontinuity at $\mu = 1/2$. A distribution ρ is CI-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 feasible 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 must be conditionally independent of the Sender's signal. The set W_{CI} is not closed: Any sequence of continuous distributions converging to a Dirac delta at 1/2 lies in W_{CI} but its limit does not. At the same time, any such sequence yields values that converge to the upper bound of 1 – the best achievable payoff to the Sender in the best case. It is also clear that the supremum of 1 cannot be achieved by any CI-worst-case optimal signal (because the only candidate – a Dirac delta at 1/2 – is not CI-worst-case optimal). This shows that a CI-robust solution may fail to exist. Note, however, that a Dirac delta at 1/2 (which corresponds to no disclosure by the Sender) can be approximated by a sequence of distributions that are themselves CI-worst-case optimal.

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

Definition OA.1. A feasible distribution over posterior beliefs $\rho \in \Delta\Delta\Omega$ is a *weak* CI-robust solution if it maximizes $\int_{\text{supp}(\rho)} \widehat{V}(\mu) d\rho(\mu)$ over $cl(W_{CI})$, where $cl(W_{CI})$ denotes the closure (in the weak* topology) of the set of CI-worst-case optimal distributions of posterior beliefs.

A weak solution thus relaxes the requirement that the distribution ρ is CI-worst-case optimal. Instead, it requires that it can be approximated by distributions that are CI-worst-case optimal. With this in mind, we establish our main existence result.

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

Proof. Define

$$v(\rho) \equiv \inf_{\pi} \int_{\text{supp}(\rho)} \underline{V}_{\pi}(\mu) d\rho(\mu)$$

as the CI-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.

First, by a result in Kamenica and Gentzkow (2011), for any feasible distribution of posterior beliefs $\rho \in \Delta\Delta\Omega$ there exists a signal function $q_\rho : \Omega \to \Delta\mathcal{S}$ that induces this distribution (the subsequent results do not depend on which particular q_ρ we pick). 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: $\inf_{\nu \in \Delta\Delta\Omega} \int_{\text{supp}(\nu)} \underline{V}_{q_\rho}(\mu) d\nu(\mu)$ subject to $\int_{\text{supp}(\nu)} \mu d\nu(\mu) = \mu_0$, where, for any signal function 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, $\int_{\text{supp}(\nu)} \underline{V}_{q_\rho}(\mu) d\nu(\mu)$ 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 $\int_{\text{supp}(\nu)} \mu d\nu(\mu) = \mu_0$ 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, which is what we wanted to prove. Moreover, the problem of finding a distribution $\rho \in \Delta\Delta\Omega$ that maximizes $v(\rho)$ subject to the Bayes plausibility condition $\int_{\text{supp}(\rho)} \mu d\rho(\mu) = \mu_0$ has a solution, and the set of solutions, W_{CI}, is non-empty and compact.

When, instead, $\underline{V}$ is not continuous, what remains true is that the set $cl(W_{CI})$ is non-empty (by Observation 1 in the main text) and compact because it is a closed subset of a compact space.

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 $W_{CI} = cl(W_{CI})$. Thus, the problem of finding a (weak) CI-robust solution is equivalent to the problem of finding a distribution $\rho \in \Delta\Delta\Omega$ that maximizes $\int_{\text{supp}(\rho)} \widehat{V}(\mu) d\rho(\mu)$ over $cl(W_{CI})$. 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 $cl(W_{CI})$ is compact, a solution to the above problem always exists, thus establishing existence of (weak) CI-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). This is because, in that case, given any induced posterior μ, Nature can always induce a conditional expected payoff to the Sender equal to $\text{lco}(\underline{V})(\mu)$ – the lower convex closure of $\underline{V}$ evaluated at μ. The convex closure is a convex function, and convex functions are continuous on the interior of the domain. This guarantees that the set W of worst-case optimal distributions is closed, while, in general the set of CI-worst-case optimal distributions W_{CI} need not be closed. Note that we have not defined an analog of the function $\underset{\sim}{V}$ in the present case: This is because the worst-case response by Nature does not have the posterior-separability property when the signal must be conditionally independent.

OA.3.4 CI-robustness for binary state

In this subsection, we consider the case where Ω is binary. Unlike in the case where Nature can condition on the realization of the Sender's signal, 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 $s \equiv \underline{V}(1) - \underline{V}(0)$ 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 CI-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 CI-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 feasible $\rho \in \Delta\Delta\Omega$ 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 $\mu_0\underline{V}(1)$.

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

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

In particular, we have that $U_\rho(1) = \mu_0\underline{V}(1)$ because $\pi = 1$ corresponds to a signal by Nature that fully discloses the state. Let $U_\rho'(1)$ denote the left derivative of $U_\rho(\pi)$ with respect to π, evaluated at $\pi = 1$ (let $\Delta\rho(1)$ be the probability mass that ρ puts on the belief $\mu = 1$). We then have that

$$U_\rho'(1) = \lim_{\epsilon\to 0}\frac{U_\rho(1) - U_\rho(1-\epsilon)}{\epsilon} = \mu_0\underline{V}(1) - \lim_{\epsilon\to 0}\frac{\int_0^1 (1-\mu(1-\epsilon))\underline{V}\left(\frac{\epsilon\mu}{1-\mu(1-\epsilon)}\right) d\rho(\mu)}{\epsilon}$$
$$\overset{(1)}{=} \mu_0\underline{V}(1) - \int_{[0,\,1)}\left(\lim_{\epsilon\to 0}\frac{\underline{V}\left(\frac{\epsilon\mu}{1-\mu(1-\epsilon)}\right)}{\frac{\epsilon\mu}{1-\mu(1-\epsilon)}}\frac{\mu-\mu^2+\mu^2\epsilon}{1-\mu+\mu\epsilon}\right) d\rho(\mu) - \underline{V}(1)\Delta\rho(1)$$
$$= \mu_0\underline{V}(1) - \underline{V}'(0)\left[\mu_0 - \Delta\rho(1)\right] - \underline{V}(1)\Delta\rho(1) = \left[\mu_0 - \Delta\rho(1)\right]\left[s - \underline{V}'(0)\right] > 0, \tag{OA.2}$$

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 $\lim_{\epsilon\to 0}\frac{\epsilon\mu}{1-\mu(1-\epsilon)} = 0$, 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 information payoff $\underline{V}_{\text{full}}(\mu_0)$ by selecting a binary signal π that is almost fully revealing. Therefore, full disclosure is the unique CI-worst-case optimal distribution, and hence the unique CI-robust solution. □

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 CI-robust solution is full disclosure of the state.

The proof of Proposition OA.1 shows that, through an appropriate binary signal, Nature can make sure that any non-degenerate posterior belief μ induced by the Sender can be decomposed 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 close to (but different from) a Dirac at $\omega = 0$ yields the Sender a payoff strictly less that 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 information payoff. Therefore, in such cases, full disclosure is the unique CI-robust solution.

Loosely speaking, under the conditions in Proposition OA.1, the Sender fully reveals the state not because she is worried that, else, Nature will do it, but because she realizes that if she does not fully reveal the state herself, Nature will *almost* fully reveal the state, and being exposed to almost full revelation is strictly worse than being exposed to full revelation.

The above intuition can also be used to compare CI-worst-case optimality to worst-case optimality (and hence CI-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 CI-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 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 make sure 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 (thus, 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 distributions special in the analysis of CI-worst-case optimality.

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

Proposition OA.2. *If $\underline{V}(\mu) \geq \underline{V}_{full}(\mu)$ for all μ, then all feasible distributions $\rho \in \Delta\Delta\Omega$ are CI-worst-case optimal. In this case, a distribution $\rho \in \Delta\Delta\Omega$ is a CI-robust solution if and only if it is a Bayesian solution.*

Proof. By Theorem 1 in the main text, under the assumptions of the proposition, all feasible distributions are worst-case optimal, and hence they are also CI-worst-case optimal. Hence, for $\rho \in \Delta\Delta\Omega$ to be a CI-robust solution, ρ must maximize $\widehat{V}$ over the entire set of feasible 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 CI-worst-case optimal solutions and hence it is CI-robust. □

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 CI-worst-case optimality have any bite. In this case, the set of CI-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 CI-robust solution. However, full disclosure is the unique CI-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 CI-robust solutions can be difficult to characterize.

OA.3.5 State separation under CI-robustness

In this subsection, we characterize properties of CI-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 $V : \Delta\Omega \to \mathbb{R}$, let $dV(\mu;\, \mu')$ denote the Gateaux derivative of V at μ in the direction of μ'. The latter is defined by

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

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 distribution $\mu = \delta_\omega$, the Gateaux derivative of $\underline{V}_{\text{full}}(\mu)$ in the direction of the Dirac distribution $\delta_{\omega'}$ is equal to

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

Theorem OA.3. *Suppose that for some pair of $\omega,\, \omega' \in \Omega$, $d\underline{V}(\delta_\omega;\, \delta_{\omega'}) < \underline{V}(\delta_{\omega'}) - \underline{V}(\delta_\omega)$. Then, any CI-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 CI-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-information payoff, and hence ρ is not a CI-worst-case optimal distribution. $\square$

Theorem OA.3 can be used to provide a weaker version of Corollary 4.

Corollary OA.1. *If for all $\omega,\, \omega' \in \Omega$, the condition of Theorem OA.3 holds, then full disclosure is the unique CI-robust solution.*

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

Corollary OA.2. *If $\underline{V} \geq \underline{V}_{full}$, then all feasible distributions are CI-worst-case optimal.*

This is the same condition as the one identified by Corollary 5 in the main text. Moreover, Corollary 5 actually implies Corollary OA.2 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 gains a more tractable form in the case where $\Omega \subset \mathbb{R}$, and the Sender's payoff depends only on the expected state.

Corollary OA.3. *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 CI-worst-case optimal distribution must separate the states ω and ω' with probability one.*

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

Corollary 7 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 CI-robust solutions. We do this by means of a counterexample. Our counterexample is rather contrived and has no immediate economic interpretation. We suspect that the conclusion of Corollary 7 can only fail, when one replaces robustness with CI robustness, in very special cases.

The example exploits the fact that Corollary 6 in the main text does extend to CI-robust solutions: a mean-preserving spread of a CI-worst-case optimal distribution need not be CI-worst-case optimal. For intuition, think of a mean preserving spread as an additional signal, on top of the original signal. 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 herself such spreads 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 signal. 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 CI-worst-case optimality. In turn, this implies that the Sender can strictly benefit from withholding information, whereas this is never the case when Nature can condition its signal on the realization of the Sender's signal.

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 payoff under the favorable selection satisfies $\widehat{V}(\mu) = 2$ if $\mu \notin G$ and $\widehat{V}(\mu) = 3$ if $\mu \in G$, where $G \equiv \{1/3,\ 7/12,\ 2/3,\ 3/4\}$. Clearly, given $\widehat{V}$, there are many Bayesian solutions—any feasible 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 value 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 red solid arrows). The counterexample is constructed by selecting the Sender's payoff under the adversarial selection $\underline{V}$ so that ρ_R is the unique CI-robust solution.

The idea is to construct a function $\underline{V}$ under which the Sender gets a low payoff from inducing beliefs 7/12 and 3/4 (that is, by splitting 2/3 into 7/12 and 3/4) so that ρ_{BP} is not CI-worst-case optimal. Let $\underline{V}(\mu) = 0$ except over a finite set of points specified below.[26] Suppose that $\underline{V}(7/12) = \underline{V}(3/4) = -1$, whereas $\underline{V}(\mu) = 0$ for all $\mu \neq \{7/12, 3/4\}$. Then ρ_{BP} is clearly not CI-worst-case optimal, for, by not disclosing any information, Nature guarantees that the Sender's expected payoff under ρ_{BP} is strictly below her full information payoff, which is equal to zero. Note, however, that this is not enough, because under such $\underline{V}$, ρ_R is also not CI-worst-case optimal.

[26]Note that, contrary to what 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.

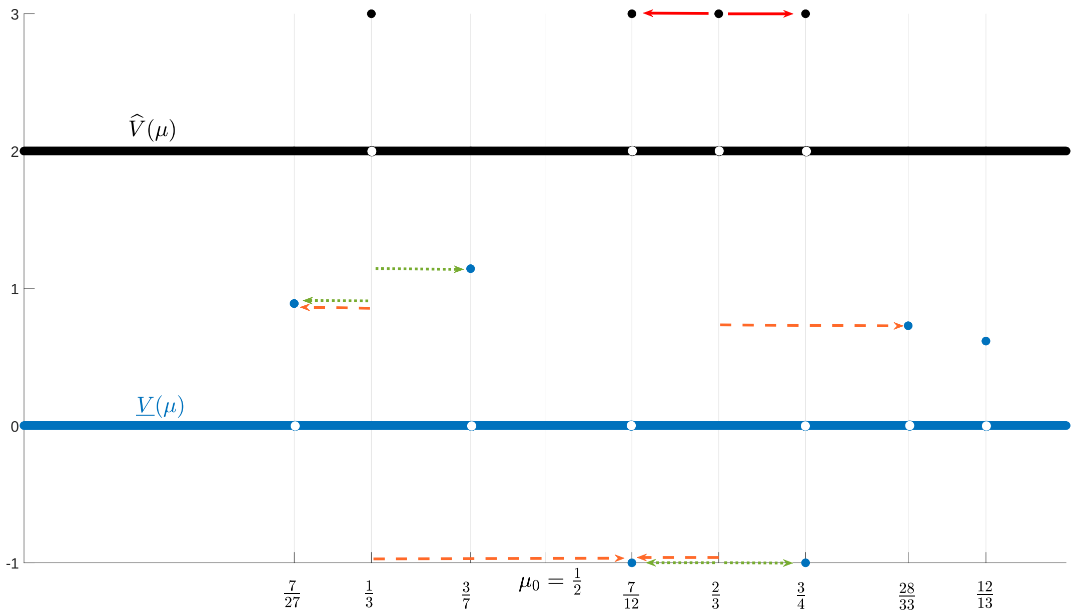

Figure OA.3.1: The functions $\underline{V}$ and $\widehat{V}$

Indeed, by choosing π appropriately, Nature can induce a posterior of 7/12 and/or of $\mu = 3/4$ with positive probability, thus bringing the Sender's payoff strictly below the full-information payoff. In particular, Nature could use the same signal that the Sender uses to split 2/3 into 7/12 and 3/4. Therefore, we need to construct $\underline{V}$ so that, if Nature chooses such a signal, when the Sender's induced posterior is 1/3 instead of 2/3, the Sender's expected payoff is sufficiently above zero to compensate for the loss that Nature imposes to the Sender when the latter induces the posterior 2/3.

Observe that there is a unique binary signal that splits 2/3 into 7/12 and 3/4 (the effects of such decomposition on $\underline{V}$ are illustrated by the green dotted arrows in Figure OA.3.1). Conditional on the Sender inducing a posterior of 1/3, the same signal then decomposes 1/3 into 7/27 and 3/7 with conditional probabilities that are pinned down uniquely. We can then choose the values $\underline{V}(7/27)$ and $\underline{V}(3/7)$ in such a way that, when Nature selects the binary signal π and the Sender induces a distribution ρ_R, the Sender's expected payoff is exactly equal to 0 – her full-disclosure payoff.

However, Nature does not need to pick a signal π that decomposes 2/3 into 7/12 and 3/4 (which also decomposes 1/3 into 7/27 and 3/7). To minimize the Sender's payoff, Nature can pick a signal that induces only one of the two posteriors $7/12$ and $3/4$ with positive probability when the Sender induces a posterior of $2/3$. An example of such a signal is the one corresponding to the orange dashed arrows in Figure OA.3.1 (Such a signal decomposes $1/3$ into $7/27$ and $7/12$ and $2/3$ into $7/12$ and $28/33$). For ρ_R to be CI-worse case optimal, the value of $\underline{V}(28/33)$ must then be selected in a way that the Sender's ex-ante expected payoff is at least zero.

To complete the characterization of $\underline{V}$, we use Lemma OA.2 which says that, to minimize the Sender's expected payoff, Nature can restrict itself to binary signals. If $\underline{V}(7/12) = \underline{V}(3/4) - 1$, and $\underline{V}(\mu) \geq 0$ for all $\mu \neq 7/12, 3/4$, 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. To construct a function $\underline{V}$ that makes ρ_R CI-worst-case optimal, we can use the proof of Lemma OA.2 which states 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 feasible distributions, where q is any signal by the Sender that induces ρ_R. One such signal is the binary signal given by $\mathcal{S} = \{l, h\}$, $q(l|0) = 2/3$, and $q(l|1) = 1/3$. This q induces ρ_R when the prior is $\mu_0 = 1/2$. Given such a signal, we then have that the Sender's expected payoff when Nature induces the posterior μ is equal to

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

To guarantee that ρ_R is a CI-worst-case optimal distribution it then suffices to choose a $\underline{V}$ that takes value 0 almost everywhere (including at $\mu = 0$ and at $\mu = 1$), is such that $\underline{V}(\mu) < 0$ only for $\mu = 7/12, 3/4$, at which it takes value $\underline{V}(7/12) = \underline{V}(3/4) = -1$, and is such that $\underline{V}_q(\mu) \geq 0$ for all μ. Under such a $\underline{V}$, when the Sender picks the above signal q, no matter the signal selected by Nature, the Sender's expected payoff is at least equal to her full-information payoff (which is equal to 0). Hence q is CI-worst-case optimal. There are only four values of μ at which $\underline{V}_q(\mu)$ can be negative: $\mu = 7/17$, $3/5$, $14/19$, $6/7$. Indeed, only for these four posteriors, given the Sender's signal q, the final posterior takes value equal to $7/12$ 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 aforementioned four posterior beliefs. Solving this system, we obtain that

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

| as illustrated in Figure OA.3.1. This completes the construction of the function $\underline{V}$, as summarized in the following claim.

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.3) 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 CI-robust solution ρ_R.*

By the construction of $\underline{V}$, ρ_R is CI-worst-case optimal, and because it yields the maximal payoff of 3 under $\widehat{V}$, it is a CI-robust solution. It only remains to show that ρ_R is the *unique* CI-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 $\{1/3, 7/12, 2/3, 3/4\}$ (since this is the only way to guarantee an expected payoff of 3 which is required for being a CI-robust solution). Furthermore, for ρ' to be CI-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 in which there exists a Bayesian solution ρ_{BP} that strictly dominates the unique CI-robust solution ρ_R in the Blackwell order.