DISCUSSION PAPER SERIES

DP15017

PREPARING FOR THE WORST BUT HOPING FOR THE BEST: ROBUST (BAYESIAN) PERSUASION

Piotr Dworczak and Alessandro Pavan

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

PREPARING FOR THE WORST BUT HOPING FOR THE BEST: ROBUST (BAYESIAN) PERSUASION

Piotr Dworczak and Alessandro Pavan

Discussion Paper DP15017
Published 08 July 2020
Submitted 05 July 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

PREPARING FOR THE WORST BUT HOPING FOR THE BEST: ROBUST (BAYESIAN) PERSUASION

Abstract

We propose a robust solution concept for Bayesian persuasion that accounts for the Sender's ambiguity over (i) the exogenous sources of information the Receivers may learn from, and (ii) the way the Receivers play (when multiple strategy profiles are consistent with the assumed solution concept and the available information). The Sender proceeds in two steps. First, she identifies all information structures 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. Second, she picks an information structure that, in case Nature and the Receivers play favorably to her, maximizes her expected payoff over all information structures that are "worst-case optimal." 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.

JEL Classification: D82

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

Piotr Dworczak - piotr.dworczak@northwestern.edu
Northwestern University

Alessandro Pavan - alepavan@northwestern.edu
Northwestern University and CEPR

Acknowledgements
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.

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

Piotr Dworczak and Alessandro Pavan†

May 27, 2020

Abstract

We propose a robust solution concept for Bayesian persuasion that accounts for the Sender's ambiguity over *(i)* the exogenous sources of information the Receivers may learn from, and *(ii)* the way the Receivers play (when multiple strategy profiles are consistent with the assumed solution concept and the available information). The Sender proceeds in two steps. First, she identifies all information structures 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. Second, she picks an information structure that, in case Nature and the Receivers play favorably to her, maximizes her expected payoff over all information structures that are "worst-case optimal." 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 standard Bayesian persuasion model, a Sender designs an information structure to influence the behavior of one or multiple Receivers. The Sender is Bayesian, and in particular knows the distribution of the Receivers' prior beliefs. Furthermore, in case there are multiple strategy profiles consistent with the disclosed information and the assumed solution concept (e.g., Bayes-Nash equilibrium), the Sender trusts that the Receivers will select the strategy profile that is most favorable to her. As a result of such assumptions, the Sender's optimal signal typically depends on the fine details of the information the Receivers are endowed with.

In many economic environments of interest, the Sender may not know which information sources the Receivers have access to, prior to receiving the Sender's information, and/or which additional information they may be able to collect after receiving the Sender's information. Furthermore, the Sender may not trust her ability to coordinate the Receivers on the strategy profile most favorable to her.

We propose a novel formulation of the persuasion problem that accounts for the above type of uncertainty. We assume the Sender proceeds in two steps. First, she identifies all information structures that are "worst-case optimal," i.e., that yield the largest payoff guarantee when both Nature and the Receivers play adversarially. Adversarial play means that Nature provides the Receivers with exogenous information that, once paired with the Sender's signal, minimizes the Sender's payoff. Moreover, whenever multiple strategy profiles are consistent with the assumed solution concept, the Sender also expects Nature to coordinate the Receivers on the profile that is worst for her. Second, among all information structures that are worst-case optimal, the Sender selects one that maximizes her payoff in case Nature does not provide the Receivers with additional information and the Receivers play according to the strategy profile that is most favorable to the Sender, among those that are consistent with the assumed solution concept and the available information. This second step – in which the Sender uses the "best-case scenario" to evaluate her remaining policies – is analogous to the optimization in the standard Bayesian persuasion model but on the restricted set of (worst-case optimal) signals. We refer to solutions to the above problem as *robust solutions*. Robust solutions are thus policies that are best-case optimal among all policies that are worst-case optimal.

Our main technical result is a separation theorem identifying states that cannot appear together in the support of any of the posterior beliefs induced by the Sender. Separation of such states is both necessary and sufficient for worst-case optimality. As a result, robust solutions maximize the Sender's best-case-scenario payoff over all information policies that induce posterior beliefs with admissible supports.

We show how the separation theorem permits us to identify various properties of robust solutions and 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. 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.

We also show that robust solutions are characterized by an optimization program analogous to the one describing Bayesian solutions but with the state-separation condition added as an additional constraint. Moreover, because such constraints can be incorporated into the Sender's objective function, the problem of finding a robust solution is equivalent to the problem of finding a Bayesian solution with a modified objective function. This implies that robust solutions can be found using techniques developed for the standard persuasion model, and that they inherit many of the structural properties of Bayesian solutions, for example, sufficiency of using as many signals as there are states. Even when a closed-form solution is not available, our state separation condition is easy to check and often provides economically meaningful restrictions on the minimal informativeness of the Sender's signal required for robustness.

Finally, we revisit some of the applications considered in the literature. We identify instances in which a Bayesian solution is worst-case optimal (and hence is also a robust solution), cases where, contrary to what is predicted in the literature, the unique robust solution entails full disclosure, and cases where a robust solution does not coincide with either full disclosure or a standard Bayesian solution.

On the interpretation of the model. A robust solution is best-case optimal among policies that are worst-case optimal. Importantly, the Sender's ambiguity over possible scenarios arises from Nature's ability to select additional signals and coordinate the Receivers on any strategy profile consistent with the assumed solution concept. Thus, our best-case scenario coincides with the standard Bayesian model that assumes that the Sender knows the exact probability distribution of the Receivers' information and can select her preferred strategy profile. As a consequence, a robust solution can be seen as a constrained-optimal Bayesian solution, where the constraint is that feasible policies must pass a "robustness test." Passing the robustness test means being optimal in some adversarial scenario that can be flexibly chosen depending on the application of interest through the choice of the solution concept. For example, suppose that the Sender hopes to induce the Receivers to play a certain (Bayesian Nash) equilibrium profile but is concerned about the possibility of strategic mis-coordination. The robustness test can be obtained by assuming that the Receivers will play the rationalizable profile that is worst for the Sender. In this case, a robust solution maximizes the Sender's expected (Bayesian-Nash-equilibrium) payoff among all policies that perform optimally under the worst-case scenario defined above.

In Section 5, we show that defining the worst-case scenario through Nature's choice of additional signals implies certain properties of the function that maps posterior beliefs induced by the Sender to her final payoff. This allows us to generalize the class of robustness tests that can be accommodated. For example, our results continue to hold in certain cases where the set of signals that Nature can use is limited, as well as in environments where the robustness constraint reflects the obligation to provide a minimal payoff guarantee to a third party.

We also discuss the role of the assumption that the Sender has lexicographic preferences over her payoffs in the worst- and best-case scenario. We consider an alternative version of the problem in which the Sender maximizes a weighted sum of these payoffs, and show that, for objective functions that satisfy a permissive regularity condition, solutions to this problem coincide with robust solutions as long as the weight on the worst-case scenario is sufficiently large. This result builds on the special structure of the problem under consideration and in particular on our state-separation theorem that implies that worst-case optimality is equivalent to ruling out posterior beliefs with certain supports.

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). More closely related are papers that adopt an adversarial approach to the design of the optimal information structure. Inostroza and Pavan (2018), Mathevet et al. (2019), Morris et al. (2019), and Ziegler (2019) focus on the adversarial selection of the continuation strategy profile (rather than of the Receivers' exogenous information structure). 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). Cui and Ravindran (2020) consider persuasion by competing designers in zero-sum games and identify conditions under which full disclosure is the unique outcome.[1] Related are also Kolotilin et al. (2017) and Laclau and Renou (2017). These papers 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.[2]

Our results are different from those in 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 in the best-case scenario (as in the canonical Bayesian persuasion model). In particular, our solution concept reflects the idea that there is no reason for the Sender to fully disclose the state if she can 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. As anticipated above, our robust solutions are also solutions to the problem of a designer maximizing a weighted average of her payoff under the worst- and the best-case scenarios (as in the literature on alpha-max-min preferences[3]), with a large weight on the worst-case payoff.

Organization. The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 contains all main results. Section 4 presents a few applications. Section 5 discusses alternative robustness constraints and the connection between our lexicographic robust solutions and the solutions to weighted max-min objectives. The Appendix at the end of the document contains all missing proofs. Finally, the Online Appendix contains a couple of additional results for the case of weighted objective functions and a complete analysis of the case of conditionally independent signals, whereby Nature cannot condition its disclosures on the realization of the Sender's information.

[1] Kosterina (2019) assumes that Nature cannot condition its information 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. The same assumption is made in the Stackelberg version of the zero-sum game between competing designers examined in Cui and Ravindran (2020).

[2] 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.

[3] See, among others, Hurwicz (1951), Gul and Pesendorfer (2015), and Grant et al. (2020).

2 Model

A payoff-relevant state ω is drawn from a finite set Ω according to the distribution $\mu_0 \in \Delta\Omega$. A Sender (she) chooses an information structure $q : \Omega \to \Delta\mathcal{S}$ that maps states into probability distributions over signal realizations $s \in \mathcal{S}$: We denote by $dq(\cdot\,|\omega)$ the probability distribution over the set of signal realizations $\mathcal{S}$ 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 Receivers may have access to, when learning about the state. In particular, the Sender need not know the distribution of Receivers' beliefs prior to receiving information from her, and/or the information they may be able to collect about ω, after observing the realization of her signal. We capture both types of uncertainty by allowing Nature to disclose additional information to the Receivers that can be arbitrarily correlated with both the state and the realization of the Sender's signal. That is, in the eyes of the Sender, Nature chooses an information structure $\pi : \Omega \times \mathcal{S} \to \Delta\mathcal{R}$ that maps $(\omega, s) \in \Omega \times \mathcal{S}$ into a distribution over a set of signal realizations $\mathcal{R}$. We denote by $d\pi(\cdot\,|\omega,\, s)$ the probability distribution (under the signal π) over $\mathcal{R}$ when the state is ω and the realization of the Sender's signal is s. Hereafter, we treat the (measurable) signal spaces $\mathcal{S}$ and $\mathcal{R}$ as exogenous and assume that they are subsets of some sufficiently rich space $\mathcal{X}$. As it will become clear from the analysis below, under our solution concept, there will be no loss of optimality for either the Sender or Nature in restricting attention to signals that take finitely many values. We then denote by Q and Π the set of all feasible signals for the Sender and Nature, respectively.

The possibility for Nature to condition her signal on the realization of the Sender's signal reflects the Sender's concern that the Receivers may be able to acquire additional information after seeing the realization of her signal. Clearly, this is the most adversarial scenario for the Sender. In certain applications, though, the Sender may find it appropriate to exclude certain types of signals that Nature can use. As an example of such restrictions, in the Online Appendix, we consider the case where the Sender expects Nature to select only signals that are independent of her signal, conditional on the state. This restriction may capture the idea that the Sender faces ambiguity over the Receivers' information prior to observing the realization of her own signal but is certain that they will not be able to collect additional information after observing the realization of her signal.

Fixing the 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 the signals (where x could be a vector). For example, given q and π, we denote by $\mu_0^{s,r}$ the common 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, using

Bayes rule.

To facilitate the exposition, we abstain from modeling the game played by the Receivers in response to the information provided by the Sender. We do so by assuming the Sender's payoff is described by a pair of functions $\overline{V}$, $\underline{V}: \Delta\Omega \to \mathbb{R}$ that map the induced (common) posteriors into the Sender's expected payoff. We assume that $\overline{V}$ is upper semi-continuous, and that $\underline{V}$ is lower semi-continuous. We interpret $\overline{V}(\mu)$ as the conditional expected payoff to the Sender when, given the induced (common) posterior $\mu \in \Delta\Omega$, the Receivers play (in the un-modeled game) according to the strategy profile most favorable to the Sender, within those consistent with the assumed (and un-modeled) solution concept. Similarly, we interpret $\underline{V}(\mu)$ as the Sender's conditional expected payoff when, given the induced posterior μ, the Receivers play according to the strategy profile most adversarial to the Sender.[4] Under the above interpretation, $\overline{V} \geq \underline{V}$. However, this property plays no role for any of our results. In Section 5, we leverage on the fact that $\underline{V}$ does not need to be related to $\overline{V}$ to provide alternative interpretations of our model.

The only restriction the above specification imposes is that the Sender discloses the same information to all Receivers, which seems plausible for many of the applications of interest. That Nature discloses the same information to all Receivers is not important for our results. This is because one can always select $\underline{V}(\mu)$ to be the infimum over the set of the Sender's payoffs that, given μ, Nature can induce by providing the agents with (possibly private) signals consistent with the common posterior μ, and coordinating them to play according to the strategy profile most adversarial to the Sender among those consistent with the assumed solution concept (e.g., Bayes-Nash equilibrium). For example, the payoffs $\overline{V}(\mu)$ and $\underline{V}(\mu)$ could be equal to the Sender-best and Sender-worst *Bayes Correlated Equilibrium*, respectively, when μ is the common prior of the basic game.

While the distinction between $\underline{V}$ and $\overline{V}$ allows our solution concept to capture the Sender's ambiguity over equilibrium selection, our results do not hinge on that interpretation. In particular, the above "black-box" formulation includes the simple case in which $\underline{V} = \overline{V} = V$ for some continuous payoff function V, so that the designer faces no ambiguity over the outcome that prevails for any given (final) posterior. See also Section 5 for alternative interpretations of the $\overline{V}$ and $\underline{V}$ functions.

[4] In the case of a single Receiver, the distinction between $\overline{V}$ and $\underline{V}$ may reflect different tie-breaking rules the Sender may expect the Receiver to follow when indifferent over multiple actions.

3 Robust solutions

Let

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

denote the expected payoff to the Sender under the adversarial selection rule $\underline{V}$ when she chooses signal q and Nature chooses signal π. Similarly, we denote by $\overline{v}(q,\,\pi)$ her expected payoff under the favorable selection (with $\underline{V}$ replaced by $\overline{V}$ in the formula above).

Definition 1. A signal $q \in Q$ is *worst-case optimal* (for the Sender) if for all signals $q' \in Q$,

$$\inf_{\pi\in\Pi} \underline{v}(q,\,\pi) \geq \inf_{\pi\in\Pi} \underline{v}(q',\,\pi).$$

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 $\underline{V}_{\text{full}}(\mu_0)$. For any distribution $\mu \in \Delta\Omega$, we let

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

denote the Sender's expected payoff when, starting from the belief μ, the Sender expects the state to be fully disclosed (here δ_ω denotes the Dirac distribution assigning measure one to the state ω). Clearly, this upper bound can be achieved if the Sender discloses the state herself. Hence, we can make the following simple observation.

Observation 1. A signal q is worst-case optimal (i.e., $q \in W$) if and only if $\inf_{\pi\in\Pi} \underline{v}(q,\,\pi) = \underline{V}_{\text{full}}(\mu_0)$. 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 in mind, we can now formally define the notion of a robust solution. We let $\emptyset$ denote a degenerate signal that discloses no information about the state.

Definition 2. A signal $q \in Q$ is a *robust solution* if it maximizes $\overline{v}(q,\,\emptyset)$ over 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 and the Receivers play adversarially. Second, if there are multiple signals that pass this test, the Sender seeks one among them that maximizes her payoff in the most favorable case, under the assumption that Nature does not interfere by providing the Receivers with exogenous information and that the Receivers play according to the selection most advantageous to the Sender (this second step is thus analogous to the one in the standard Bayesian persuasion model). In short, a robust solution is best-case optimal among those that are worst-case optimal.

To see that this interpretation is valid, note that a robust solution q also maximizes $\sup_{\pi\in\Pi} \overline{v}(q, \pi)$ over W (which explicitly captures the assumption that Nature favors the Sender). The converse, however, is not true: If q maximizes $\sup_{\pi\in\Pi} \overline{v}(q, \pi)$ over W, it need not maximize $\overline{v}(q, \emptyset)$ over the same set. In this sense, our definition of a robust solution is conservative: Even though the Sender expects that, once the "robustness test" is passed, Nature and the Receivers play favorably to her, she discloses information herself as opposed to relying on Nature to do it.

Because the Sender's payoff depends only on the induced (common) 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. Given any measurable set X and any probability measure $\gamma \in \Delta X$ over X, let $\text{supp}(\gamma)$ denote the support of γ, i.e., the smallest closed subset of X whose complement has zero measure according to γ.

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 $\int \overline{V}(\mu)d\rho(\mu)$ over $\mathcal{W}$, where $\mathcal{W} \subset \Delta\Delta\Omega$ is the set of distributions over posterior beliefs satisfying*

$$\int \left(\min_{\pi:\,\Omega\to\Delta\Omega} \left\{ \sum_{\omega,\,r\in\Omega} \underline{V}(\mu^r)\pi(r|\omega)\mu(\omega) \right\} \right) d\rho(\mu) = \underline{V}_{full}(\mu_0), \tag{WC}$$

and Bayes plausibility

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

Lemma 1 is intuitive. Any feasible distribution over posterior beliefs ρ must satisfy Bayes plausibility (BP). 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. (Note that the choice of π can depend on which μ has realized, reflecting the idea that the choice of Nature's signal may depend on the realization of the Sender's signal.) Moreover, as we formally show in the Appendix, without loss of optimality, Nature can be assumed to select a signal whose signal space is given by $\mathcal{R} = \Omega$. 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 $\mathcal{W}$.

Now let $\text{lco}(\underline{V})$ denote the lower convex closure of $\underline{V}$. That is, if $\text{co}(\underline{V})$ denotes the concave closure of $\underline{V}$, as defined by Kamenica and Gentzkow (2011), then $\text{lco}(\underline{V}) = -\text{co}(-\underline{V})$. Clearly, for any $\mu \in \Delta\Omega$,

$$\min_{\pi:\,\Omega\to\Delta\Omega} \left\{ \sum_{\omega,\, r\in\Omega} \underline{V}(\mu^r)\pi(r|\omega)\mu(\omega) \right\} = \text{lco}(\underline{V})(\mu).$$

Note that, for any $\mu \in \Delta\Omega$, $\text{lco}(\underline{V})(\mu) \le \underline{V}_{\text{full}}(\mu)$, with the inequality holding as an equality when $\mu = \delta_\omega$, that is, when μ is a Dirac delta assigning measure one to some state $\omega \in \Omega$. Hence, $\text{co}(\text{lco}(\underline{V})) = \underline{V}_{\text{full}}$. With this notation in hands, the worst-case optimality condition (WC) can be conveniently rewritten as

$$\int \text{lco}(\underline{V})(\mu) d\rho(\mu) = \underline{V}_{\text{full}}(\mu_0). \tag{3.1}$$

Hereafter, we will abuse terminology and call ρ_{RS} a robust solution if it maximizes the Sender's expected payoff under the favorable selection $\int \overline{V}(\mu)d\rho(\mu)$ over all distributions $\rho \in \Delta\Delta\Omega$ satisfying (BP) and (3.1), with no further reference to the underlying signal.

It is useful at this point to contrast the definition of a robust solution to the definition of a Bayesian-persuasion solution (henceforth, *Bayesian solution*) as defined in Kamenica and Gentzkow (2011).

Definition 3. A signal q_{BP} is a *Bayesian solution* if it maximizes $\overline{v}(q, \emptyset)$ over 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 $\int \overline{V}(\mu)d\rho(\mu)$ over all ρ satisfying (BP).

Throughout, we will refer to any $\rho \in \Delta\Delta\Omega$ satisfying (BP) as a *feasible* distribution.

We are now 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.

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

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

Then,

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

Therefore, $\rho_{RS} \in \Delta\Delta\Omega$ is a robust solution if and only if it maximizes

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

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

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

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) $\underline{V}(\mu) \geq \underline{V}_{\text{full}}(\mu)$ for all $\mu \in \Delta B$. That is, if and only if *any* posterior 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.

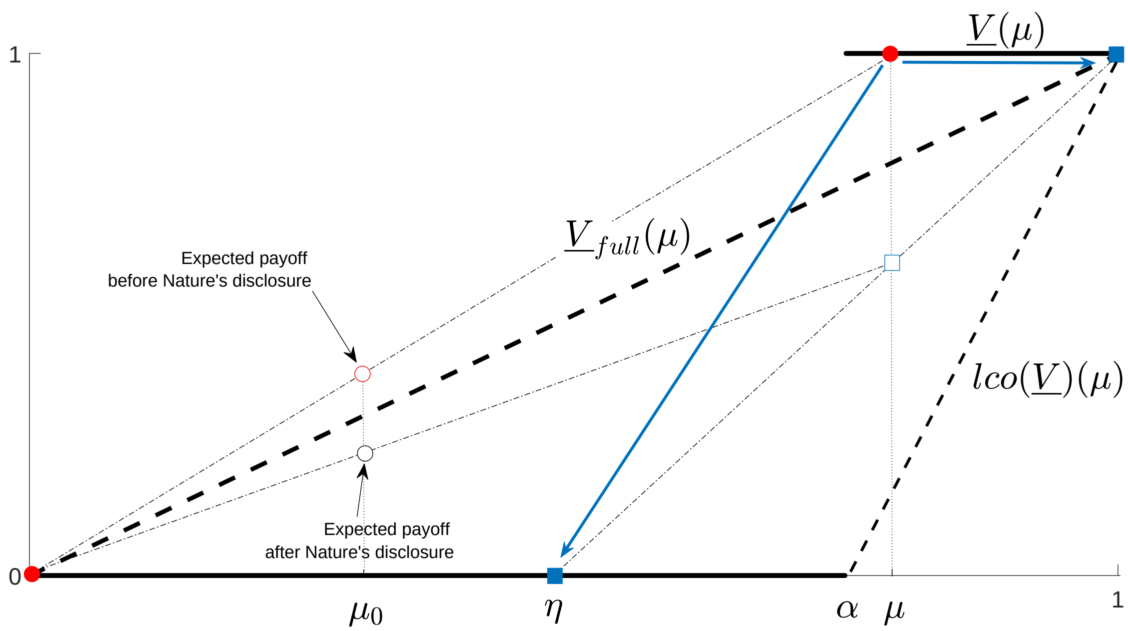

Figure 3.1: Illustration of Theorem 1 in the judge example of Kamenica and Gentzkow (2011)

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 otherwise). 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.

Figure 3.1 illustrates the logic of Theorem 1 in the context of the "judge" example of Kamenica and Gentzkow (2011). The state is binary, i.e., $\omega \in \{\omega_L, \omega_H\}$ and the Sender gets a payoff of 1 if and only if the posterior belief she induces assigns probability at least $\alpha \in (\mu_0, 1)$ to the state being ω_H, where μ_0 denotes the prior belief that the state is ω_H. The payoff from full disclosure is given by the dashed line connecting the extreme points $(0, \underline{V}(0))$ and $(1, \underline{V}(1))$. As Kamenica and Gentzkow show, the Sender can achieve a better payoff by splitting the prior into a posterior belief of 0 and $\mu \in (\alpha, 1)$. However, no such signal is worst-case optimal. Indeed, because μ has full support, Nature can induce any belief $\eta \in (0, \alpha)$ with positive probability by disclosing the state ω_L with appropriate frequency. By decomposing μ into η and 1, Nature pushes the Sender's expected payoff below her full-disclosure payoff. In fact, Nature's problem, given μ, can be seen as a Bayesian persuasion problem with prior μ and objective function $-\underline{V}$. Thus, Nature can induce a conditional expected payoff given by $\text{lco}(\underline{V})(\mu)$ – the lower convex closure of $\underline{V}$ evaluated at μ. In the judge example, this implies that full-disclosure is the unique worst-case optimal signal and hence that the only robust solution consists in fulling revealing the state.

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 $\overline{V}$ 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; and 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'. If all states must be separated, full disclosure is the unique worst-case optimal distribution.

Corollary 3 (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 4 (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 3-4, suppose that $\underline{V}$ is strictly convex. Then, for any two states $\omega, \omega' \in \Omega$, $\underline{V}(\delta_\omega) = \underline{V}_{\text{full}}(\delta_\omega)$, $\underline{V}(\delta_{\omega'}) = \underline{V}_{\text{full}}(\delta_{\omega'})$, but

$$\underline{V}(\lambda\delta_\omega + (1-\lambda)\delta_{\omega'}) < \lambda\underline{V}(\delta_\omega) + (1-\lambda)\underline{V}(\delta_{\omega'}) = \underline{V}_{\text{full}}(\lambda\delta_\omega + (1-\lambda)\delta_{\omega'})$$

implying that ω and ω' must be separated. Since ω and ω' are arbitrary, $\mathcal{F}$ can only contain singletons, and thus Corollary 3 implies that full disclosure is the unique robust solution. Conversely, suppose that $\underline{V}$ is concave. Then, by definition of concavity, for any $\mu \in \Delta\Omega$, $\underline{V}(\mu) \geq \sum_{\omega\in\Omega} \underline{V}(\delta_\omega)\mu(\omega) = \underline{V}_{\text{full}}(\mu)$. Thus, $\Omega \in \mathcal{F}$, and Corollary 4 implies that all distributions are worst-case optimal. In this case, because worst-case optimality imposes no restrictions, any robust solution must coincide with a Bayesian solution. While $\Omega \in \mathcal{F}$ is sufficient for robustness of a Bayesian solution, it is not necessary.

Corollary 5 (Robustness of Bayesian Solutions)**.** *A Bayesian solution is robust if and only if the support of any posterior belief it induces is in $\mathcal{F}$.*

A special case of interest is when the state space is binary, as in the judge example above. The result below extends our analysis of the example from Figure 3.1 to all binary problems.

Corollary 6 (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.*

By Corollary 6, in the binary-state case, a robust solution discloses weakly more information than a Bayesian solution. To see whether this property holds more generally, we say that $\rho \in \Delta\Delta\Omega$ Blackwell dominates $\rho' \in \Delta\Delta\Omega$ if there exist signals $q' : \Omega \to \Delta\mathcal{S}'$ inducing ρ', and $q : \Omega \to \Delta(\mathcal{S}' \times \mathcal{S})$ inducing ρ such that the marginal distribution of q on $\mathcal{S}'$ conditional on any $\omega \in \Omega$ coincides with that of q'.

Corollary 7 (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 7, that Bayesian solution must be worst-case optimal, and hence it is also a robust solution. Therefore, we obtain the following conclusion.

Corollary 8 (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 8 provides a formal sense in which (maximally informative) robust solutions provide more information than Bayesian solutions.[5] 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 8 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 8 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 8 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

[5]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 statements is obviously false. For example, when both $\underline{V}$ and $\overline{V}$ are affine, all distributions are both robust and Bayesian solutions and hence trivially there exist Bayesian solutions that strictly Blackwell dominate some robust solutions.

the state). The counterexample we provide is based on failure of the property identified in Corollary 7. What makes Corollary 7 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.[6]

Corollary 9 (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. We illustrate this and previous points with some examples in the next section.

Finally, we show that robust solutions can be found using the concavification technique (see Aumann and Maschler, 1995, and Kamenica and Gentzkow, 2011). Indeed, because the state-separation condition applies posterior by posterior, we can incorporate the constraints into the objective function $\overline{V}$ 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 $v_{\text{low}} := \min_{\omega\in\Omega} \overline{V}(\delta_\omega) - 1$, and define

$$\overline{V}_{\mathcal{F}}(\mu) = \begin{cases} \overline{V}(\mu) & \text{if } \mu \in \Delta_{\mathcal{F}}\Omega \text{ and } \overline{V}(\mu) \geq v_{\text{low}}, \\ v_{\text{low}} & \text{otherwise .} \end{cases} \tag{3.2}$$

Observe that posteriors μ with $\overline{V}(\mu) \leq v_{\text{low}}$ 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} (which is a feasible operation for the Sender, under both solution concepts). Therefore, Bayesian solutions under the objective function $\overline{V}_{\mathcal{F}}$ correspond exactly to robust solution with the original objective, by Theorem 1. Moreover, we have defined the modification $\overline{V}_{\mathcal{F}}$ of $\overline{V}$ so that it remains upper-semi-continuous because the set $\{\mu \in \Delta\Omega : \text{supp}(\mu) \in \mathcal{F} \text{ and } \overline{V}(\mu) \geq v_{\text{low}}\}$ is closed. By the above reasoning, we obtain the following corollary.

Corollary 10. *A feasible distribution $\rho \in \Delta\Delta\Omega$ is a robust solution if and only if*

$$\int \overline{V}_{\mathcal{F}}(\mu) d\rho(\mu) = co(\overline{V}_{\mathcal{F}})(\mu_0).$$

[6] When 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.

Corollary 10 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.

Corollary 11. *There always exists a robust solution ρ with $|supp(\rho)| \leq |\Omega|$.*

4 Applications

4.1 Buyer-seller interactions

To illustrate the implications of our solution concept, we start with a few applications of information design to buyer-seller interactions. We first revisit two well-known results and show how the conclusions change once robustness is accounted for. Next, we consider a novel application to a lemons problem.

Privately informed Receiver: Guo and Shmaya (2019)

We first analyze a 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 buyer has private information about the product's quality ω summarized by the realization t of a signal f mapping each $\omega \in \Omega$ into a distribution on a finite set $\mathcal{T}$, and this fact is common knowledge. The seller can decide to disclose more information to the buyer about the product's quality, taking f as given, but does not observe the signal realization t. Guo and Shmaya (2019) show that, when the signal f satisfies MLRP (formally, when the probability distribution $f(t|\omega)$ is log-supermodular), a Bayesian solution has an *interval structure*: each buyer's type t 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 in the above environment by a seller who is uncertain about what information the buyer may have access to, over and above the one contained in the signal f. To avoid uninteresting cases, assume that f is not fully revealing.[7]

In such an environment, given any posterior $\mu \in \Delta\Omega$ induced by the seller (where the posterior is obtained by conditioning only on the realization of the seller's signal, *before*

[7]That is, conditional on any state ω, there is positive conditional probability that the signal realization t from f does not reveal that the state is ω.

conditioning on the buyer's type), the seller's payoff can be written as

$$\underline{V}(\mu) = \sum_{\omega \in \Omega} \sum_{t \in \mathcal{T}} \mathbf{1}\left(\frac{\sum_{\omega' \in \Omega} \omega' f(t|\omega')\mu(\omega')}{\sum_{\omega' \in \Omega} f(t|\omega')\mu(\omega')} > p \right) f(t|\omega)\mu(\omega)$$

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

The following is a simple implication of Corollary 2.

Proposition 1. *Any robust solution must separate 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. Disclosing a binary signal informing the buyer of whether the state is (weakly) below or (strictly) above p is an example of a robust solution. Another one is to fully disclose the state.

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, regardless of the realization of her type t, even when the good is of high quality. By fully disclosing the state, the seller guards herself against this possibility and ensures that all high-quality goods ($\omega > p$) are bought with certainty.

The limits of price discrimination: Bergemann, Brooks and Morris (2015)

In an influential paper, Bergemann et al. (2015) consider the problem of a monopolistic seller quoting a price to a buyer who is privately informed about her value ω for the seller's good. They characterize all feasible distributions over consumer surplus and seller's profit, as the seller's information about the buyer's value varies (which we can model as a consequence of information about ω revealed by a Sender). Here, we argue that the model is robust to the uncertainty the Sender may face about the monopolist's exogenous information when the Sender maximizes *(i)* seller's profit or *(ii)* buyer's surplus.

In case *(i)*, the Bayesian solution is full disclosure. Because full disclosure is always worst-case optimal, this is also a robust solution. In the more interesting case *(ii)*, the Bayesian solution is quite complicated and consists of multiple signals that carefully segment the market. Nevertheless, this solution is also robust. To see that, we can apply Theorem 1: Because the buyer's surplus is 0 at degenerate beliefs, we have that $\underline{V}_{\text{full}}(\mu) = 0$ for all μ, while $\underline{V}(\mu) \geq 0$ for all μ. Thus, $\Omega \in \mathcal{F}$, and any Bayesian solution is robust.

That being said, there are other objective functions for which robust solutions differ from Bayesian solutions. For example, suppose that the Sender wants to maximize the expected price quoted by the seller. Bayesian solutions always involve some randomization (except for degenerate cases) and might in general be complicated. However, the unique robust solution is to fully disclose the state because the Sender's objective function projected to any $\Delta\{\omega, \omega'\}$ takes the shape analogous to the objective function in the judge example (see Figure 3.1), implying that any two states must be separated. Similarly, if the Sender maximizes a weighted combination of consumer's and producer's surplus, with strictly positive weight to both, then full disclosure is the unique robust solution whereas the Bayesian solution involves a more complex market segmentation.

Lemons problem

In the previous two examples, either full disclosure or the Bayesian solution emerged as robust solutions. We now consider a simple example with adverse selection in which this is not the case.

A 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 $\mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega} \leq p] + D > p$.[8] 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.

Given the induced posterior μ, the payoff to the seller is given by

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

Note that, in this example, $V = \underline{V} = \overline{V}$, as 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.

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

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

Proof. 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 $v(\lambda) \equiv V|_{\Delta B}(\lambda\delta_\omega + (1-\lambda)\delta_{\omega'})$. 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 $V|_{\Delta B}(\mu)$ is below the full-disclosure payoff $V_{\text{full}}|_{\Delta B}(\mu)$ 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 $\max\Omega < 1$. $\square$

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 $p \geq \underline{\omega}_\mu$.*

Proof. Clearly, $\mathbf{1}_{\{\mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega}\leq p]+D>p\}} \leq \mathbf{1}_{\{\mathbb{E}_\mu[\tilde{\omega}]+D>p\}}$. Suppose that the inequality is strict for some $p \geq \underline{\omega}_\mu : \mathbb{E}_\mu[\tilde{\omega}] + D > p$ but $\mathbb{E}_\mu[\tilde{\omega}|\tilde{\omega} \leq p] + D \leq p$. This is only possible when $p < \overline{\omega}_\mu$, where

$\overline{\omega}_\mu$ is the maximum of supp(μ). But then

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

a contradiction. □

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, $V|_{\Delta B}(\mu)$ is concave on ΔB (and non-affine if $|B| \geq 2$).*

The proof is technical and hence relegated to the Appendix. Intuitively, Lemma 4 states that the seller does not benefit from splitting posterior beliefs with sufficiently small supports. The next result is then a simple implication of the above lemmas:

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 $V|_{\Delta B}(\mu)$ 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, Lemma 4 implies, any robust solution discloses just enough information to eliminate the adverse selection problem.

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

Indeed, since $V|_{\Delta B}(\mu)$ 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 a robust solution must disclose just enough information to neutralize the adverse selection problem. We believe that the same approach – deriving properties of robust solutions without completing solving for optimal policies – is likely to be fruitful in other applications as well.

4.2 Supermodular games

In this subsection, we consider an application featuring multiple Receivers in which the lowest and the highest selections $\underline{V}$ and $\overline{V}$ correspond to different rationalizable profiles in the continuation game among the Receivers.

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.

Suppose first that the Sender knew that the only information the agents possess is the common prior μ_0 and were confident in her ability to coordinate the Receivers on the best rationalizable profile – corresponding to a Bayesian solution ρ_{BP}. A Bayesian solution would then be similar to the one in the judge example in Section 3 (see the proof of Proposition 3 in the Appendix). 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 $\omega > 0$.[9] Now let $\mu_0^+ \equiv \mu_0(\omega > 0)$ denote the probability assigned by the prior μ_0 to

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

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. The Sender can then maximize her payoff through a policy that, when $\omega > 0$, sends the "null" signal $s = \emptyset$ with certainty, whereas, when $\omega \leq 0$, fully discloses the state with probability $\phi_{BP} \in (0, 1)$ and sends the signal $s = \emptyset$ with the complementary probability, where ϕ_{BP} is defined by $\mu_0^+/[\mu_0^+ + (1 - \mu_0^+)(1 - \phi_{BP})] = \alpha$.

This policy, 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 her ability to coordinate the agents 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 Sender faces ambiguity over the information the agents are exogenously endowed with, worst-case optimality requires that all states $\omega > 1$ be separated from all states $\omega \leq 1$. (The arguments are similar to those in the judge's example in Section 3: For any induced posterior whose support contains both states $\omega > 1$ and states $\omega \leq 1$, Nature can construct another posterior that induces the agents to attack also when $\omega > 1$, thus bringing the Sender's payoff below her full information payoff.) One may then conjecture that, as in the judge example, full disclosure of the state is a robust solution. 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. Formally, we have the following result:

Proposition 3. *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 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 separates 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 8 and 9.

4.3 Mean-measurable case

Finally, we consider an example in which $\Omega \subseteq [0, 1]$, $\underline{V} = \overline{V}$, and $\underline{V}(\mu) = u(\mathbb{E}_\mu[\tilde{\omega}])$, for all $\mu \in \Delta\Omega$, for some continuous function $u : [0, 1] \to \mathbb{R}$. The Sender's payoff depends only on the mean of each posterior distribution over states that she induces. Corollary 2 implies that under the above assumptions, two states ω and ω' must be separated in a robust solution if there exists $\lambda \in (0, 1)$ such that $u(\lambda\omega + (1-\lambda)\omega') < \lambda u(\omega) + (1-\lambda)u(\omega')$.

We will show how the above condition can be leveraged to find robust solutions in an example that received significant attention in previous work: We assume that $u(x)$ is strictly convex, strictly increasing, and differentiable for all $x < \alpha$, and equal to $u(\alpha)$ for all $x \geq \alpha$, for some $\alpha \in (0, 1)$, as depicted in Figure 4.1. Kolotilin (2018) obtains a similar objective function in a model with a privately informed Receiver who takes a binary action (under some assumptions on the distribution of the Receiver's information); Dworczak and Martini (2019) obtain it in a model in which a principal motivates an agent to take costly effort by rewarding her with a fixed share of the prize and disclosing information about the prize value (the kink corresponds to the minimum level of the agent's expected value of the prize necessary for the agent to choose full effort); Duffie et al. (2017) obtain it in a model where a market regulator designs a benchmark that discloses the size of gains from trade to market participants to induce an efficient level of entry in the presence of externalities (the kink corresponds to the minimum level of the expected value of the gains from trade at which all market participants enter). These papers show that the optimal Bayesian solution is upper censorship: It reveals the state up to a threshold, and then pools the remaining states into a single message that induces the posterior mean α.

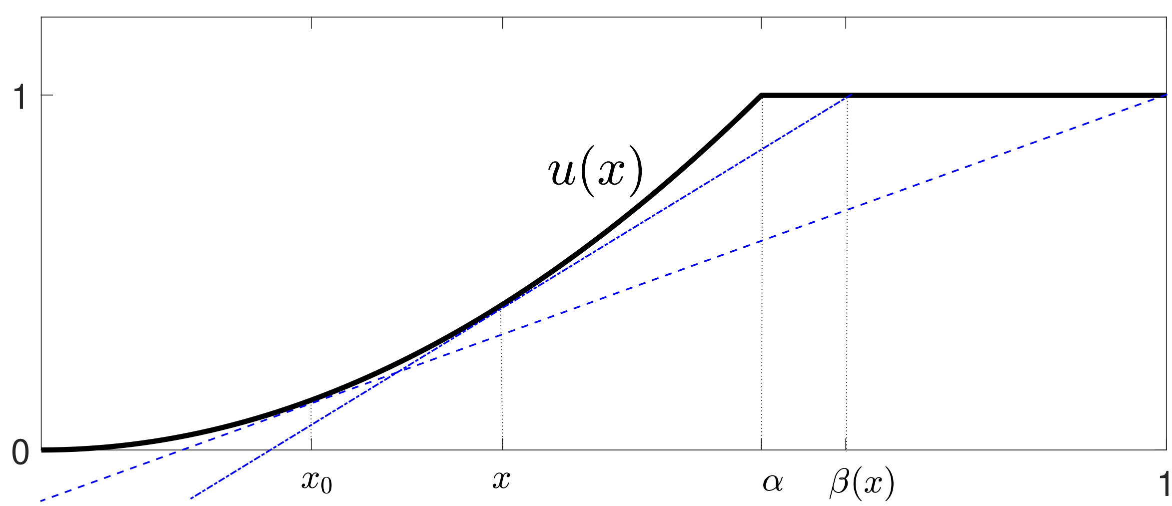

Figure 4.1: The function $u(x)$

Our first result shows which states *must* be separated by a robust solution. For any $x \in [0,\, 1]$, let

$$\beta(x) \equiv \begin{cases} x + \frac{u(\alpha)-u(x)}{u'(x)} & x \in [0,\, \alpha), \\ \alpha & x \geq \alpha. \end{cases}$$

The function $\beta(x)$ is a decreasing function that outputs the x-coordinate of the point at which a line tangent to u at x crosses the horizontal line $y = u(\alpha)$.

Lemma 6. *The set $\mathcal{F}$ of admissible supports for the posteriors induced by worst-case optimal policies is given by*[10]

$$\mathcal{F} = \{B \subset \Omega :\ B \subseteq \{\omega\} \cup [\beta(\omega),\, 1],\ \textit{for some } \omega \in \Omega\}.$$

The set $\mathcal{F}$ consists of all sets B contained in subsets of Ω of the form $\{\omega\} \cup [\beta(\omega), 1]$ for some $\omega \in \Omega$. Worst-case optimal distributions need not separate states that lie above α. On the other extreme, let x_0 be the supremum of $\omega \in \Omega$, $\omega < \alpha$, such that $u(\omega) + u'(\omega)(1-\omega) < u(\alpha)$. All states $\omega < x_0$ must be fully revealed because for any such ω, $\{\omega\} \cup [\beta(\omega),\, 1] = \{\omega\}$. For any intermediate state $\omega \in [x_0,\, \alpha)$, instead, worst-case optimality requires that ω must be separated from all states except the ones in $[\beta(\omega),\, 1]$. Note that because $\beta(\omega) \geq \alpha$ for any ω, this implies that all states below α must be separated from one another.

Given Lemma 6, to arrive at a robust solution, one needs to solve an optimal persuasion problem subject to the additional constraint that the supports of induced posterior beliefs satisfy the condition in Lemma 6.

Proposition 4 (Mean-measurable case)**.** *There exists a robust solution with the following structure: For some $x^\star \in [x_0,\, \alpha)$:*

- *States $\omega < x^\star$ are fully revealed;*
- *States $\omega \in (x^\star,\, \alpha)$ are pooled (under separate signals) with states above $\beta(\omega)$ to induce a posterior mean of at least α;*
- *There may exist a state $\omega = x^\star$ which is revealed with some probability and, with the remaining probability, pooled with states above $\beta(x^\star)$ to induce a posterior mean of exactly α.*

In certain cases, deriving a robust solution is particularly simple. We know that all states $x < x_0$ must be revealed to guarantee worst-case optimality (by Lemma 6). Conditional on

[10] If $\beta(\omega) > 1$, $[\beta(\omega),\, 1] = \emptyset$.

the state being above x_0, the Sender's payoff is upper-bounded by $u(\alpha)$. If this upper bound can be achieved by a worst-case optimal distribution, then this distribution must be a robust solution.

Corollary 12. *Suppose that a feasible distribution $\rho \in \Delta\Delta\Omega$, in addition to satisfying the restrictions of Lemma 6, is such that conditional on any $\omega \geq x_0$, any of the induced posteriors has a mean above α. Then, ρ is a robust solution.*

A robust solution resembles a Bayesian solution in that low states are fully revealed and sufficiently high states are pooled to induce a posterior mean above α. However, there is a crucial difference. In the Bayesian solution, all sufficiently high states are pooled into a single signal which allows to maximize the pooling region. In contrast, in a robust solution, a posterior mean above α is induced by many different signal realizations because no two states below α can be pooled with one another.

5 Discussion

5.1 Alternative robustness tests

In this subsection, we discuss what properties of our setting are key for the main results, including the conclusion that robustness only constrains the supports of posterior beliefs; subsequently, we observe that our methods extend to alternative robustness tests.

Instead of modeling Nature playing against the Sender, consider a function $\mathcal{V} : \Delta\Omega \to \mathbb{R}$ mapping posterior beliefs induced by the Sender to payoffs. Suppose that $\mathcal{V} \leq \mathcal{V}_{\text{full}}$ and that

$$\mathcal{V}(\mu) < \mathcal{V}_{\text{full}}(\mu) \implies \mathcal{V}(\eta) < \mathcal{V}_{\text{full}}(\eta), \forall \eta \in \Delta\Omega \text{ s.t. } \operatorname{supp}(\eta) = \operatorname{supp}(\mu). \tag{5.1}$$

Define the robustness test by requiring that $\rho \in \Delta\Delta\Omega$ selected by the Sender satisfies

$$\int \mathcal{V}(\mu) d\rho(\mu) \geq \mathcal{V}_{\text{full}}(\mu_0). \tag{5.2}$$

That is, when posteriors are evaluated according to $\mathcal{V}$, the Sender's signals must be at least as good as full disclosure. Then, by $\mathcal{V} \leq \mathcal{V}_{\text{full}}$ and (5.1), the robustness property is equivalent to requiring that for all $\mu \in \operatorname{supp}(\rho)$, $\operatorname{supp}(\mu) \in \{B \subseteq \Omega : \mathcal{V}|_{\Delta(B)} \geq \mathcal{V}_{\text{full}}|_{\Delta(B)}\}$, analogously to Theorem 1.

Our baseline model can be understood as a special case of the above framework where $\mathcal{V} = \text{lco}(\underline{V})$ with $\underline{V}$ interpreted as the Sender's payoff under adversarial equilibrium selection. In our framework, the property that $\mathcal{V} \leq \mathcal{V}_{\text{full}}$ is a consequence of the fact that Nature can always disclose the state, while (5.1) follows from convexity of $\text{lco}(\underline{V})$.

The more general formulation shows that Nature does not need to have access to all possible signals for our results to hold. As long as $\mathcal{V} \leq \mathcal{V}_{\text{full}}$, for property (5.1) to hold, it suffices that starting from any belief η, Nature can induce any other belief μ with the same support as μ, with positive probability. This is consistent with various constraints that can be imposed on Nature's choice, for example, with a capacity constraint on the informativeness of Nature's signal (e.g., based on entropy reduction).

We also notice that the above constraint may admit alternative interpretations. For example, consider a setting where two different information designers move sequentially and where the upstream designer faces ambiguity over the downstream designer's response.[11] The downstream designer's objective, or instruments, may not be known to the upstream designer. In this setting, $\mathcal{V}(\mu)$ represents the upstream designer's lowest payoff over all possible reactions by the downstream designer. Our results carry over to this setting provided that the upstream designer expects the downstream designer to be able to fully reveal the state and generate with positive probability any belief η with the same support as the belief μ induced by the upstream designer.

Finally, the function $\mathcal{V}$ need not be related to the Sender's payoff. For example, the constraint (5.2) may reflect the obligation by the Sender to provide a minimal payoff guarantee to a third party (e.g., another principal in an organization). In this case, the constraint in (5.2) may reflect ambiguity aversion by the third party instead of the Sender herself. The Sender may be a standard expected-utility maximizer with preferences represented by the function $\overline{V}$. The constraint in (5.2) may arise from the third party's concern that the information provided by the Sender may be manipulated (by other, un-modeled players) prior to reaching the Receivers (in this example, $\underline{V}$ is the third party's payoff and $\mathcal{V} = \text{lco}(\underline{V})$).

5.2 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 their best-case performance. 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 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

[11] Cui and Ravindran (2020) consider a zero-sum game between two information designers in the spirit of the exercise described here, but where the upstream designer faces no ambiguity over the downstream designer's best response.

favorable, respectively. 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.[12]

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

$$\sup_{q \in Q} \left\{ \lambda \inf_{\pi \in \Pi} \underline{v}(q, \pi) + (1 - \lambda)\overline{v}(q, \emptyset) \right\}.$$

Our previous results imply that this is equivalent to

$$\sup_{\rho \in \Delta\Delta\Omega} \left\{ \lambda \int_{\Delta\Omega} \text{lco}(\underline{V})(\mu) d\rho(\mu) + (1 - \lambda) \int_{\Delta\Omega} \overline{V}(\mu) d\rho(\mu) \right\} \tag{5.3}$$

subject to (BP). The function $\overline{V}(\mu)$ is upper semi-continuous by assumption. Moreover, any convex lower-semi-continuous function on $\Delta\Omega$ is continuous,[13] and thus $\text{lco}(\underline{V})$ is continuous. Therefore, the problem for a fixed λ is equivalent to a standard Bayesian persuasion problem with an upper semi-continuous objective function $V_\lambda(\mu) \equiv \lambda \text{lco}(\underline{V})(\mu) + (1 - \lambda)\overline{V}(\mu)$, and a feasible ρ is a solution if and only if it concavifies V_λ at the prior μ_0.

Our goal is to relate the solutions to the problem defined by (5.3) (which we will denote by $S(\lambda)$ and call 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 $\overline{V}$ 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 $\text{supp}(\eta) \subseteq \text{supp}(\mu) \setminus \{\omega\}$ such that $d(\mu, \eta) \leq K\mu(\omega)$ and $\overline{V}(\mu) - \overline{V}(\eta) \leq Ld(\mu, \eta)$.

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 favorable selection $\overline{V}$. 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 $\overline{V}$. 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

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

[13]For completeness, we provide a proof in the Online Appendix OA.1.1.

(the condition rules out functions $\overline{V}(\mu)$ that *decrease* at an infinite rate as $\mu(\omega)$ approaches 0). And, indeed, regularity allows for highly discontinuous objective functions (we maintain that $\overline{V}$ is upper semi-continuous). For example, $\overline{V}(\mu) = v(\mathbb{E}_{\mu}[\omega])$, for some real-valued function v is regular, no matter the 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 $\overline{V}(\mu) = \sum_{i=1}^{k} a_i \mathbf{1}_{\{\mu \in A_i\}}$ for some partition $(A_1, ..., A_k)$ 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 $\overline{V}$ is regular. Then, 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 without the regularity condition we can prove a slightly weaker version of one direction of the equivalence: 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 that shed some light on the proof (the full proof 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 7. *There exists a constant $\delta > 0$ such that for any $\mu \in \Delta^c_{\mathcal{F}}\Omega$,*

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

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 $\text{co}(\overline{V}_{\mathcal{F}})(\mu)$ without sacrificing worst-case optimality, by Corollary 10. For $\mu \in \Delta^c_{\mathcal{F}}\Omega$, it is possible that $\overline{V}(\mu) > \text{co}(\overline{V}_{\mathcal{F}})(\mu)$ but the difference can be upper bounded.

Lemma 8. *For a regular function $\overline{V}$, there exists a constant $\Delta > 0$ such that for any $\mu \in \Delta^c_{\mathcal{F}}\Omega$,*

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

Together, the above two lemmas imply the following result:

Lemma 9. *Suppose that $\overline{V}$ is regular. There exists $\overline{\lambda} < 1$ such that, for all $\lambda \in (\overline{\lambda}, 1]$, if ρ solves the problem defined by (5.3), then ρ cannot assign positive probability to $\Delta^c_{\mathcal{F}}\Omega$.*

Theorem 2 follows from Lemma 9. 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. 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.

6 Conclusions

We introduce and analyze a novel solution concept for information design in settings in which the Sender faces uncertainty about the Receivers' sources of information. 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 and the Receivers play adversarially to her. The Sender then picks an information structure that, in case Nature and the Receivers play favorably to her, maximizes her expected payoff over all information structures that are worst-case optimal. We call such solutions robust. Our main result is a separation theorem that identifies states that can be present together in one of the induced posteriors and states that must be separated. We then use the characterization to identify various properties of robust solutions, and to illustrate the implications of robustness in a few applications considered in the literature. In the Online Appendix, we discuss how the analysis and the key results must be adapted to accommodate the possibility that the Sender expects Nature to be unable to condition her disclosure on the realization of the Sender's signal.

Throughout the analysis, we restrict attention to the case of *public persuasion* in which the Sender discloses the same information to all Receivers (but expects Nature to possibly engage in discriminatory disclosures). In future work, it would be interesting to extend the analysis to private persuasion, whereby the Sender discloses different signals to different Receivers. It would also be interesting to examine how the properties of robust solutions change as one considers different solution concepts for the game among the Receivers.

References

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

Bergemann, Dirk and Stephen Morris, "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.

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.

Duffie, Darrell, Piotr Dworczak, and Haoxiang Zhu, "Benchmarks in search markets," *Journal of Finance*, 2017, *72*, 1983 – 2044.

Dworczak, Piotr and Giorgio Martini, "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 2019, *127* (5), 1993–2048.

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, "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, 2018, *13*, 607–636.

_ **, 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.

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, forthcoming.*, 2019.

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.

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

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

A Missing Proofs

A.1 Proof of Lemma 1

Because the order in which signals are observed does not influence the posterior belief under Bayesian updating, we have $\mu_0^{s,r} = (\mu_0^r)^s = (\mu_0^s)^r$, where recall that μ^x is the posterior belief induced by observing signal realization x. Thus, using Fubini's theorem, and the law of iterated expectations,

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

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

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

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

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 $\mu = \mu_0^s$. Since $\mathcal{R}$ is now finite, we can let $\pi(r|\omega)$ denote the probability of sending signal r in state ω. Define

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

and using Observation 1, we then have that a signal q is worst-case optimal if and only if

$$\sum_{\Omega} \int_{\mathcal{S}} \left(\min_{\pi:\, \Omega \to \Delta\Omega} \underline{V}_\pi(\mu^s) \right) dq(s|\omega)\mu_0(\omega) = \underline{V}_{\text{full}}(\mu_0). \tag{A.2}$$

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 $\underline{V}(\eta) < \underline{V}_{\text{full}}(\eta)$. Because $\text{lco}(\underline{V}) \leq \underline{V}$ (where recall that $\text{lco}(\underline{V})$ denotes the lower convex closure of $\underline{V}$), we have that $\text{lco}(\underline{V})(\eta) < \underline{V}_{\text{full}}(\eta)$. Because $\text{supp}(\eta) \subseteq \text{supp}(\mu)$, there exists a small enough $\lambda > 0$ such that $\mu = \lambda\eta + (1-\lambda)\eta'$, for some $\eta' \in \Delta\Omega$. We have

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

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

We are ready to obtain a contradiction. Recall from equation (3.1) that since ρ is a worst-case optimal distribution, it must satisfy

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

which, by (BP) and the fact that $\underline{V}_{\text{full}}$ is affine, can also be written as

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

Because $\text{lco}(\underline{V}) \leq \underline{V}_{\text{full}}$, we must have $\text{lco}(\underline{V})(\mu) = \underline{V}_{\text{full}}(\mu)$ 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 (3.1) holds, or that $\text{lco}(\underline{V})(\mu) \geq \underline{V}_{\text{full}}(\mu)$ 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 $\text{lco}(\underline{V})$. In particular, $\text{lco}(\underline{V})(\mu) \geq \underline{V}_{\text{full}}(\mu)$ which finishes the proof.

A.3 Proof of Lemma 4

By Lemma 3, we can write

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

Let $B = \{\omega_1, ..., \omega_n\}$ with $\omega_1 < \omega_2 < ... < \omega_n$, and let $\mu_i = \mu(\omega_i)$. Then, V can be treated as a function defined on a unit simplex in $\mathbb{R}^n$: $V(\mu) = \frac{1}{2} \sum_{i=1}^n \mu_i \left(\sum_{j=1}^n \mu_j \omega_j + D - \omega_i \right)^2$. To prove the lemma, it is enough to prove that a function $\tilde{V}$ defined by $\tilde{V}(\mu_2, ..., \mu_n) = V(1 - \mu_2 - ... - \mu_n, \mu_2, ..., \mu_n)$ has a negative semi-definite hessian. By a direct calculation, denoting $\omega_{-1} = [\omega_2, ..., \omega_n]$, we obtain $\text{Hessian}(\tilde{V}) = -(\omega_{-1} - \omega_1)^T \cdot (\omega_{-1} - \omega_1)$, which is a negative semi-definite matrix (of rank 1).

A.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 $\overline{V}(\mu) = 0$ if $\mu^+ < \alpha$ and $\overline{V}(\mu) = 1$ 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.[14] 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 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

[14]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.

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.

A.5 Proof of Lemma 6

Suppose that $\omega < \alpha$. By Corollary 2, ω must be separated from each ω' such that for some $\lambda \in (0,\, 1)$, $u(\lambda\omega + (1-\lambda)\omega') < \lambda u(\omega) + (1-\lambda)u(\omega')$. This condition holds for any $\omega' < \alpha$ because the function u is strictly convex on $[0,\, \alpha)$. When $\omega' \geq \alpha$, the condition is equivalent to $u'(\omega) < (u(\alpha) - u(\omega)) \,/\, (\omega' - \omega)$, or $\omega' < \beta(\omega)$.

Thus, a necessary condition for ρ to be worst-case optimal is that, for any $\mu \in \text{supp}(\rho)$, if $\omega \in \text{supp}(\mu)$, and $\omega < \alpha$, then only $[\beta(\omega),\, 1]$ can have a non-empty intersection with $\text{supp}(\mu) \setminus \{\omega\}$.

We now show that this condition is also sufficient. For any $\omega \in \Omega$, take any subset B of $\{\omega\} \cup [\beta(\omega),\, 1]$ and any $\mu \in \Delta B$. If $\underline{V}(\mu) = u(\alpha)$ (that is, if $\mathbb{E}_{\tilde{\omega}\sim\mu}[\tilde{\omega}] \geq \alpha$) then $\underline{V}(\mu) \geq \underline{V}_{\text{full}}(\mu)$ because $u(\alpha)$ upper-bounds $\underline{V}$ (and hence also $\underline{V}_{\text{full}}$). If, on the other hand, $\underline{V}(\mu) < u(\alpha)$ (equivalently, $\mathbb{E}_{\tilde{\omega}\sim\mu}[\tilde{\omega}] < \alpha$), then necessarily $\omega < \alpha$, and

$$\begin{aligned}\underline{V}(\mu) = u\left(\mathbb{E}_{\tilde{\omega}\sim\mu}[\tilde{\omega}]\right) \geq u(\omega) + u'(\omega)(\mathbb{E}_{\tilde{\omega}\sim\mu}[\tilde{\omega}] - \omega) = \mathbb{E}_{\tilde{\omega}\sim\mu}\left[u(\omega) + u'(\omega)(\tilde{\omega} - \omega)\right] \\ \geq \mathbb{E}_{\tilde{\omega}\sim\mu}\left[u(\tilde{\omega})\right] = \underline{V}_{\text{full}}(\mu),\end{aligned}$$

where the first inequality follows from the fact that u is convex on $[0,\, \alpha]$ (and hence lies above the line tangent to it at $\omega < \alpha$), and the second inequality holds because, by definition of $\beta(\cdot)$, $u(\omega) + u'(\omega)(\tilde{\omega} - \omega) \geq u(\tilde{\omega})$ for $\tilde{\omega} \in \text{supp}(\mu)$. Hence, we have shown that any $\rho \in \Delta\Delta\Omega$ with the property that, for any $\mu \in \text{supp}(\rho)$, if $\omega \in \text{supp}(\mu)$ and $\omega < \alpha$, then $[0, \beta(\omega)] \cap \text{supp}(\mu) \setminus \{\omega\}\} = \emptyset$, is worst-case optimal.

A.6 Proof of Proposition 4

We order the state space so that $\Omega = \{x_1, ..., x_k, \omega_1, ..., \omega_n\}$ and $x_1 < ... < x_k < \alpha \leq \omega_1 < ... < \omega_n$. We also normalize $u(\alpha) = 1$ to simplify notation. The proof of Proposition 4 follows from Lemma 6 combined with the following lemma (and the fact that a robust solution exists, by Corollary 1).

Lemma 10. *Suppose that ρ is a robust solution. Conditional on state x_i, a posterior mean induced by ρ is either equal to x_i (full revelation) or lies above α. Moreover, if conditional on some x_i there is positive conditional probability of inducing a posterior mean above α, then conditional on all states $\omega > x_i$ the posterior mean lies above α with probability one.*

To prove the first part of the lemma, towards a contradiction, suppose that there is $\mu \in \text{supp}(\rho)$ such that $x_i \in \text{supp}(\mu)$ and $y \equiv \mathbb{E}_\mu[\tilde{\omega}] \in (x_i, \alpha)$ (the posterior mean cannot be lower than x_i because, by Lemma 6, x_i cannot be pooled with any lower state). By Lemma 6, $\text{supp}(\mu) \subseteq \{x_i\} \cup [\beta(x_i), 1]$. We will show that we can improve upon ρ by revealing the state x_i with (small) probability $\epsilon > 0$ conditional on inducing belief μ under ρ. The expected payoff given μ is

$$V_{old} \equiv u\left(x_i\mu(x_i) + \sum_{\omega \geq \beta(x_i)} \omega\mu(\omega)\right),$$

and under the alternative policy it becomes

$$V_{new} \equiv \mu(x_i)\epsilon\; u(x_i) + \left(\mu(x_i)(1-\epsilon) + \sum_{\omega \geq \beta(x_i)} \mu(\omega)\right) u\left(\frac{x_i\mu(x_i)(1-\epsilon) + \sum_{\omega \geq \beta(x_i)} \omega\mu(\omega)}{\mu(x_i)(1-\epsilon) + \sum_{\omega \geq \beta(x_i)} \mu(\omega)}\right).$$

We have

$$\lim_{\epsilon \to 0} \frac{V_{new} - V_{old}}{\epsilon} = \mu(x_i)(y - x_i)\left[u'(y) - \frac{u(y) - u(x_i)}{y - x_i}\right] > 0$$

by strict convexity of u on $[0, \alpha)$ and the fact that $y > x_i$. Therefore, ρ can be improved upon by a distribution that is worst-case optimal, and thus it could not have been a robust solution.

To prove the second part of Lemma 10, towards a contradiction, suppose that there are μ, $\mu' \in \text{supp}(\rho)$ such that $x_i \in \text{supp}(\mu)$, $\omega > x_i$, $\omega \in \text{supp}(\mu')$, $y \equiv \mathbb{E}_\mu[\tilde{\omega}] \geq \alpha$ but $y' \equiv \mathbb{E}_{\mu'}[\tilde{\omega}] < \alpha$. Because $\omega > x_i$ and $\beta(x)$ is a monotone function, we know from Lemma 6 that ω can be pooled with any state with which x_i can be pooled (we do not lose worst-case optimality). Moreover, by the first part, we know that ω is fully revealed in this case, and in particular we must have $\omega = x_j$ for some $j > i$, and $y' = x_j$. We can assume without loss of generality that $y = \alpha$ because we can reveal some states above α conditional on μ with sufficiently high probability so that the conditional mean falls to exactly α, and the expected payoff stays the same (because $u(x)$ is constant for $x \geq \alpha$). To simplify notation, let $\overline{\mu} = \sum_{\omega \geq \beta(x_i)} \mu(\omega)$ and

$$\overline{\omega} = \frac{\sum_{\omega \geq \beta(x_i)} \omega \mu(\omega)}{\sum_{\omega \geq \beta(x_i)} \mu(\omega)}.$$

Then, all subsequent calculations are unaffected if we assume that μ puts mass $\overline{\mu}$ on the "synthetic" state $\overline{\omega}$, instead of the individual ω_i's in its support. Thus, it is without loss of generality to assume that μ puts mass $1 - \overline{\mu}$ on x_i and mass $\overline{\mu}$ on $\overline{\omega}$ while μ' is a Dirac delta at $x_j > x_i$.

Let p be the conditional probability under ρ of inducing belief μ conditional on inducing either μ or μ'. We can assume that $p \in (0, 1)$ as otherwise one of these beliefs is induced with unconditional probability 0.

Consider a modification of ρ conditional on the realization of beliefs $\{\mu, \mu'\}$: Instead of inducing μ with probability p and μ' with probability $1 - p$, consider inducing *(1)* μ' (equal to δ_{x_j}) with probability $(1 - p)(1 - \epsilon)$; *(2)* μ_γ with probability $(1 - p)\epsilon + p\overline{\mu}\gamma$ supported on $\{x_j, \overline{\omega}\}$, where γ is such that

$$\mathbb{E}_{\mu_\gamma}[\tilde{\omega}] = \frac{x_j(1 - p)\epsilon + \overline{\omega} p \overline{\mu} \gamma}{(1 - p)\epsilon + p\overline{\mu}\gamma} = \alpha;$$

(3) μ with probability $p(1 - \gamma)$; *(4)* δ_{x_i} with probability $p(1 - \overline{\mu})\gamma$. Intuitively, we are modifying the signal so that the higher state x_j is pooled towards posterior mean α instead of the lower state x_i. By Lemma 6, the modified signal is still worst-case optimal. We have

$$\gamma = \frac{\epsilon}{\overline{\mu}} \left(\frac{\alpha - x_j}{\overline{\omega} - \alpha} \frac{1 - p}{p} \right)$$

which lies in $(0, 1)$ for small enough ϵ. Denoting the conditional expected payoff (conditional on $\{\mu, \mu'\}$) under the original signal by V_{old} and the conditional expected payoff under the modified signal by V_{new}, we get that

$$V_{new} - V_{old} = -\epsilon(1 - p)u(x_j) - \gamma p + (1 - p)\epsilon + p\overline{\mu}\gamma + p(1 - \overline{\mu})\gamma u(x_i).$$

Therefore, we have

$$\frac{V_{new}-V_{old}}{\epsilon} = \frac{1-p}{\overline{\mu}}\left[-\overline{\mu}u(x_j) - \frac{\alpha - x_j}{\overline{\omega}-\alpha} + \overline{\mu} + \overline{\mu}\frac{\alpha - x_j}{\overline{\omega}-\alpha} + (1-\overline{\mu})\frac{\alpha - x_j}{\overline{\omega}-\alpha}u(x_i)\right]$$
$$= (1-p)(1-u(x_i))\left[\frac{1-u(x_j)}{1-u(x_i)} - \frac{1-\overline{\mu}}{\overline{\mu}}\frac{\alpha - x_j}{\overline{\omega}-\alpha}\right]. \quad \text{(A.5)}$$

Note that $\overline{\mu}\,\overline{\omega} + (1-\overline{\mu})x_i = \alpha$, and thus $(1-\overline{\mu})/\overline{\mu} = (\overline{\omega}-\alpha)\,/\,(\alpha - x_i)$ so that

$$\frac{V_{new}-V_{old}}{\epsilon} = (1-p)(1-u(x_i))\left[\frac{1-u(x_j)}{1-u(x_i)} - \frac{\alpha - x_j}{\alpha - x_i}\right] > 0 \iff \frac{1-u(x_j)}{\alpha - x_j} > \frac{1-u(x_i)}{\alpha - x_i}$$

which is true because u is strictly convex in the relevant range, $u(\alpha) = 1$, and $x_j > x_i$. Thus, we have shown that ρ can be improved upon by a worst-case optimal signal, and thus could not have been a robust solution.

A.7 Proof of Lemma 7

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 $\text{lco}(\underline{V})(\mu_B) < \underline{V}_{\text{full}}(\mu_B)$. Then let $\delta_B \equiv \underline{V}_{\text{full}}(\mu_B) - \text{lco}(\underline{V})(\mu_B)$ 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 $\underline{V}_{\text{full}} - \text{lco}(\underline{V})$ is a concave function as the difference between an affine function and a convex function, we have

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

Since B was arbitrary, we also have that

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

A.8 Proof of Lemma 8

Before proving Lemma 8, we first prove that regularity implies a seemingly stronger property that will be more convenient to work with.

Lemma 11. *If the function $\overline{V}$ 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* $\overline{V}(\mu) - \overline{V}(\eta) \leq Ld(\mu, \eta)$.

Proof of Lemma 11. 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 lemma for the case $|\text{supp}(\mu) \setminus A| = 1$.

Induction step: Suppose that we have proven the lemma 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 $\overline{V}(\mu) - \overline{V}(\eta') \leq Ld(\mu, \eta')$.

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 $\overline{V}(\eta')-\overline{V}(\eta) \leq Ld(\eta', \eta)$.

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

$$\begin{aligned}\overline{V}(\mu) - \overline{V}(\eta) = \overline{V}(\mu) - \overline{V}(\eta') + \overline{V}(\eta') - \overline{V}(\eta) &\leq L(d(\mu, \eta') + d(\eta', \eta)) \\ &\leq LK(K + 2)\delta^A \leq LK(K + 2)d(\mu, \eta),\end{aligned}$$

where the last inequality follows from the fact that $\text{supp}(\mu) \setminus \text{supp}(\eta)$ 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 $L' = LK(K + 2)$.

This finishes the proof of Lemma 11. □

Now we prove Lemma 8: We have to show that there exists a constant $\Delta > 0$ such that for any $\mu \in \Delta^c_{\mathcal{F}}\Omega$,

$$\text{co}(\overline{V}_{\mathcal{F}})(\mu) + \Delta \max_{B\subseteq\text{supp}(\mu),\, B\notin\mathcal{F}} \min_{\omega\in B}\{\mu(\omega)\} \geq \overline{V}(\mu). \tag{A.6}$$

Let $\bar{\delta} \equiv \max_{B\subseteq\text{supp}(\mu),\, B\notin\mathcal{F}} \min_{\omega\in B}\{\mu(\omega)\}$. By definition of $\bar{\delta}$, there must exist a set $A \subsetneq \text{supp}(\mu)$, with $A \in \mathcal{F}$, such that for all $\omega \in \text{supp}(\mu) \setminus A$, $\mu(\omega) \leq \bar{\delta}$. To see that, let $C \equiv \{\omega \in \text{supp}(\mu) : \mu(\omega) > \bar{\delta}\}$. Clearly, if $C = \emptyset$, then it suffices to let A coincide with any element of $\text{supp}(\mu)$. 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 Lemma 11 applied to μ and the set A (which we can apply since $\overline{V}$ is regular), there must exist η with $\text{supp}(\eta) \subseteq A$, $d(\mu, \eta) \leq K \max_{\omega \in \text{supp}(\mu) \setminus A}\{\mu(\omega)\} \leq K\bar{\delta}$, such that

$$\overline{V}(\mu) - \overline{V}(\eta) \leq Ld(\mu, \eta) \leq LK\bar{\delta}. \tag{A.7}$$

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

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

On the line segment connecting μ and η, $\text{co}(\overline{V}_{\mathcal{F}})$ is affine: Indeed, we have that $\overline{V}_{\mathcal{F}}(\kappa\mu + (1 - \kappa)\eta) = v_{\text{low}}$ for any $\kappa > 0$, since any such belief $\kappa\mu + (1-\kappa)\eta \notin \mathcal{F}$; but this implies that $\overline{V}$ lies strictly below its concave closure (except possibly at η), and hence is affine. This means in particular that $\text{co}(\overline{V}_{\mathcal{F}})$ is Lipschitz continuous on that segment, that is, for some constant N, $\text{co}(\overline{V}_{\mathcal{F}})(\mu) - \text{co}(\overline{V}_{\mathcal{F}})(\eta) \geq -Nd(\mu, \eta)$. Therefore, using (A.7) and $d(\mu, \eta) \leq K\bar{\delta}$, we have that

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

Thus, to prove the desired inequality (A.6), it is enough to set $\Delta = NK + LK$.

A.9 Proof of Lemma 9

It is enough to prove that, for high enough λ, if $\text{supp}(\rho) \not\subseteq \Delta_{\mathcal{F}}\Omega$, then the Sender's objective $\int \left[\lambda \text{lco}(\underline{V})(\mu) + (1-\lambda)\overline{V}(\mu)\right] d\rho(\mu)$ increases strictly by splitting any $\mu \in \text{supp}(\rho)$ such that $\mu \in \Delta^c_{\mathcal{F}}\Omega$ into beliefs that yield $\text{co}(\overline{V}_{\mathcal{F}})(\mu)$ – such a split is always available to the Sender and, by definition of $\text{co}(\overline{V}_{\mathcal{F}})$, yields the payoff $\underline{V}_{\text{full}}(\mu)$ in the worst-case scenario (if μ has zero probability under ρ, then the same argument can be applied to a set $S \subseteq \text{supp}(\rho)$ of positive measure under ρ such that $S \cap \Delta_{\mathcal{F}}\Omega = \emptyset$). By Lemma 7 and 8, we have that, for some $\Delta > 0$ and $\delta > 0$,

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

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

A.10 Proof of Theorem 2

It is enough to prove that for $\lambda \in (\overline{\lambda}, 1)$, where $\overline{\lambda}$ as defined in Lemma 9, ρ concavifies $\overline{V}_{\mathcal{F}}$ at μ_0 if and only if ρ concavifies $\lambda \mathrm{lco}(\underline{V}) + (1-\lambda)\overline{V}$ at μ_0. This, however, follows directly from Lemma 9. For $\lambda > \overline{\lambda}$, a feasible $\rho \in \Delta\Delta\Omega$ that concavifies $\lambda \mathrm{lco}(\underline{V}) + (1-\lambda)\overline{V}$ at μ_0 (i.e., is a λ-solution) must not put positive probability on beliefs $\mu \in \Delta^c_{\mathcal{F}}\Omega$. However, $\mathrm{lco}(\underline{V})$ is affine on such beliefs, by definition of $\mathcal{F}$. Therefore, as long as $\lambda \in (\overline{\lambda}, 1)$, ρ concavifies $\lambda \mathrm{lco}(\underline{V}) + (1-\lambda)\overline{V}(\mu)$ at μ_0 if and only if it concavifies $\overline{V}$ at μ_0 on $\Delta_{\mathcal{F}}\Omega$. This, however, is equivalent to concavifying $\overline{V}_{\mathcal{F}}$ at μ_0.

Online Appendix

OA.1 Supplementary results for Section 5.2

In this Online Appendix, we revisit Section 5.2 and examine the consequences of relaxing the regularity condition.

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,

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

When combined with Lemma 7, the above inequality implies that

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

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 $\{\mu \in \Delta\Omega : \mathrm{supp}(\mu) = A\}$ 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,

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

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

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

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

where the second inequality comes from upper-semi continuity of $\overline{V}$, and the last inequality follows from the fact that ρ is worst-case optimal, while $\text{co}\overline{V}_{\mathcal{F}}(\mu_0)$ is the upper bound on the best-case payoff that a worst-case optimal distribution can yield. This, however, means that $\int_{\Delta\Omega} \overline{V}(\mu) d\rho(\mu) = \text{co}\overline{V}_{\mathcal{F}}(\mu_0)$, and thus ρ is a robust solution, by Corollary 10. □

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 $\overline{V}$ be such that

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

and

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

and $\overline{V}(\mu) = -1$ anywhere else. Notice that $\overline{V}$ violates regularity because along the line segment $(1 - 2x, x, x)$, as $x \to 0$, $\overline{V}$ decreases at an infinite rate to 0, while $\overline{V}(\mu) \leq 0$ 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 $\overline{V} = 0$. 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 $\text{lco}(\underline{V})(\mu) = -3 \min_\omega \mu(\omega)$. 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 $V_\lambda(\mu) \equiv \lambda \text{lco}(\underline{V})(\mu) + (1 - \lambda)\overline{V}(\mu)$ 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.1.1 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

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

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 $\liminf V(\mu_n) > V(\mu)$. □

OA.2 The case of conditionally independent signals

In the main version of the 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.

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. However, we modify Nature's strategy space: Nature selects a conditionally independent signal distribution $\pi : \Omega \to \Delta\mathcal{R}$. When the spaces $\mathcal{S}$ and $\mathcal{R}$ of signal realizations are finite, we can write the Bayesian posterior belief given the prior μ_0 and the signal realizations (s, r) as

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

We assume that $\overline{V}$ is upper semi-continuous and lower-bounded, and that $\underline{V}$ is lower semi-continuous and upper-bounded.

The Sender's expected payoffs $\overline{v}(q, \pi)$ and $\underline{v}(q, \pi)$ under the favorable and adversarial selections, respectively, continue to be defined as in the main text, but with $d\pi(r|\omega, s)$ replaced by $d\pi(r|\omega)$.

With that modification in mind, the definition of a worst-case optimal signal (Definition 1 in the main text) remains the same, except for the fact that the infimum over Nature's signals is now taken over the (smaller) space of conditionally independent signals. 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 $\overline{v}(q, \emptyset)$ over W_{CI}.

OA.2.1 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.2.3 (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.2.5, 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. We show that in the applications considered in Section 4.3, the sufficient condition for state separation from Theorem OA.3 applies, and CI-robust solutions coincide with robust solutions.

- Corollary 3, Corollary 4, and Corollary 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.2.2 and Subsection OA.2.5, 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.2.4, we analyze the binary-state case. Unlike robust solutions, as described by Corollary 6, 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.2.6, we show that Corollary 7 and Corollary 8 fail for CI-robust solutions. That is, it is possible that a Bayesian solution is strictly more informative than all CI-robust solutions.

OA.2.2 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

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

where

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

Therefore,

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

where $\Pi_{\mu_0,\pi} \in \Delta\mathcal{R}$ denotes the unconditional distribution over $\mathcal{R}$ induced by μ_0 and π. From this observation, it is easy to see that, without loss of generality, we can assume that Nature chooses a distribution $\nu \in \Delta\Delta\Omega$ over posterior beliefs over Ω, subject to Bayes plausibility. In particular, to minimize the Sender's payoff, Nature solves the following problem:

$$\inf_{\nu \in \Delta\Delta\Omega} \int_{\text{supp}(\nu)} \underline{V}_q(\mu) d\nu(\mu)$$

subject to

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

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 $\int \overline{V}(\mu)d\rho(\mu)$ among all CI-worst-case optimal distributions, that is, among all distributions that satisfy

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

As before, we will abuse notation 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 infimum is inside the integral, i.e., it is computed posterior by posterior. For CI-worst-case optimality, instead, the infimum 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.

Before proceeding, we make two simple observations that show that CI-robust solutions coincide with Bayesian solutions in certain simple cases.

Observation OA.1. When $\underline{V}$ and $\overline{V}$ are convex, full disclosure is a CI-robust solution.

Proof. Let $V \in \{\underline{V}, \overline{V}\}$. Note that, by Jensen's inequality,

$$\begin{aligned} V_\pi(\mu) &= \sum_{\omega,\,r} V\left(\sum_{\omega'} \mu^r(\omega')\delta_{\omega'}\right)\pi(r|\omega)\mu(\omega) \\ &\leq \sum_{\omega,\,r,\,\omega'} V(\delta_{\omega'})\frac{\pi(r|\omega')\mu(\omega')}{\sum_{\tilde{\omega}}\pi(r|\tilde{\omega})\mu(\tilde{\omega})}\pi(r|\omega)\mu(\omega) = \sum_{\omega} V(\delta_\omega)\mu(\omega), \end{aligned} \tag{OA.2}$$

and hence, for any feasible $\rho \in \Delta\Delta\Omega$,

$$\int_{\text{supp}(\rho)} V_\pi(\mu)d\rho(\mu) \leq \sum_{\omega} V(\delta_\omega)\mu_0(\omega) = \underline{V}_{\text{full}}(\mu_0).$$

for any π. □

Note that we have established an even stronger property in the proof: Full disclosure is a dominant strategy for the Sender (it is optimal irrespective of the signal chosen by Nature).

Observation OA.2. When $\underline{V}$ and $\overline{V}$ are concave, no disclosure is a CI-robust solution.

Proof. First, note that, irrespective of the Sender's signal, Nature always fully discloses the state – this follows from the proof of Observation OA.1 applied to the negative of $\underline{V}$ (which is a convex function under the assumptions of the observation). This implies that all feasible distributions for the Sender are CI-worst-case optimal.

On the other hand, we have, for any π and ρ (using the concavity of $\overline{V}$, along with Jensen's inequality)

$$\int_{\text{supp}(\rho)} \overline{V}_\pi(\mu) d\rho(\mu) = \int_{\text{supp}(\rho)} \sum_{\omega, r} \overline{V}(\mu^r)\pi(r|\omega)\mu(\omega) d\rho(\mu)$$

$$= \sum_r \left(\int_{\text{supp}(\rho)} \overline{V}(\mu^r) \frac{\sum_\omega \pi(r|\omega)\mu(\omega)}{\sum_\omega \pi(r|\omega)\mu_0(\omega)} d\rho(\mu) \right) \sum_\omega \pi(r|\omega)\mu_0(\omega)$$

$$\leq \sum_r \overline{V} \left(\int_{\text{supp}(\rho)} \mu^r \frac{\sum_\omega \pi(r|\omega)\mu(\omega)}{\sum_\omega \pi(r|\omega)\mu_0(\omega)} d\rho(\mu) \right) \sum_\omega \pi(r|\omega)\mu_0(\omega)$$

$$= \sum_{r, \omega} \overline{V}\left(\mu_0^r\right) \pi(r|\omega)\mu_0(\omega) = \overline{V}_\pi(\mu_0),$$

where the last term is the expected payoff for the Sender when the Sender chooses no disclosure and Nature chooses π. □

Again, in the proof of Observation OA.2, we have established a stronger property – no disclosure is dominant for the Sender when the objective function is concave.

OA.2.3 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 $\mu_0 = 1/2$. Let

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

and let $\overline{V}$ and $\underline{V}$ be, respectively, the point-wise highest and lowest selections from the correspondence $\mathcal{V}$. Then, $\overline{V}$ 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 $\int_{\text{supp}(\rho)} \overline{V}(\mu)d\rho(\mu)$ 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 $\int_{\text{supp}(\rho)} \overline{V}(\mu)d\rho(\mu)$ over $cl(W_{CI})$. Because the objective function is upper semi-continuous in ρ (this follows from the fact that, by assumption, $\overline{V}$ 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. □

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.

OA.2.4 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

$$U_\rho'(1) = \lim_{\epsilon\to 0} \frac{U_\rho(1) - U_\rho(1-\epsilon)}{\epsilon} = \mu_0 \underline{V}(1) - \lim_{\epsilon\to 0} \frac{\int_0^1 (1-\mu(1-\epsilon))\underline{V}\left(\frac{\epsilon\mu}{1-\mu(1-\epsilon)}\right) d\rho(\mu)}{\epsilon}$$

$$\stackrel{(1)}{=} \mu_0 \underline{V}(1) - \int_{[0,\,1)} \left(\lim_{\epsilon\to 0} \frac{\underline{V}\left(\frac{\epsilon\mu}{1-\mu(1-\epsilon)}\right)}{\frac{\epsilon\mu}{1-\mu(1-\epsilon)}} \frac{\mu - \mu^2 + \mu^2\epsilon}{1-\mu+\mu\epsilon} \right) d\rho(\mu) - \underline{V}(1)\Delta\rho(1)$$

$$= \mu_0 \underline{V}(1) - \underline{V}'(0)\left[\mu_0 - \Delta\rho(1)\right] - \underline{V}(1)\Delta\rho(1) = \left[\mu_0 - \Delta\rho(1)\right]\left[s - \underline{V}'(0)\right] > 0, \quad \text{(OA.3)}$$

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) (see Figure 3.1, and the discussion around it in the main text) 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 $\overline{V}$ 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 $\overline{V}$ 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.2.5 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. □

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

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 4 in the main text. Moreover, Corollary 4 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, as in Section 4.3.

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.*

While Theorem OA.3 and Corollary OA.3 have more limited power than Theorem 1, they imply that the CI-robust solutions in the application considered in Section 4.3 satisfy the same properties as their robust-solution counterparts.

Claim OA.1. *In the application in Section 4.3, the set of CI-worst-case optimal distributions is the same as the set of worst-case optimal solutions.*

Proof. It is enough to observe that all the arguments in the proof of Lemma 6 in the main text hold true when Nature is restricted to selecting conditionally independent signals. The result then follows directly from Corollary OA.3. □

Intuitively, the proof of Lemma 6 in the main text treats separately each pair of states ω and ω' and is established by observing that the payoff $\underline{V}(\mu)$ lies strictly below the full-disclosure payoff $\underline{V}_{\text{full}}(\mu)$ when the support of μ includes states ω and ω' that satisfy the condition in Corollary OA.3. Hence, Lemma 6 applies also to the case of conditionally independent signals.

OA.2.6 A Bayesian solution can Blackwell dominate a CI-robust solution

Corollary 8 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 8 can only fail, when one replaces robustness with CI robustness, in very special cases.

The example exploits the fact that Corollary 7 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 $\overline{V}(\mu) = 2$ if $\mu \notin G$ and $\overline{V}(\mu) = 3$ if $\mu \in G$, where $G \equiv \{1/3,\ 7/12,\ 2/3,\ 3/4\}$. Clearly, given $\overline{V}$, there are many Bayesian solutions – any feasible distribution of posteri-

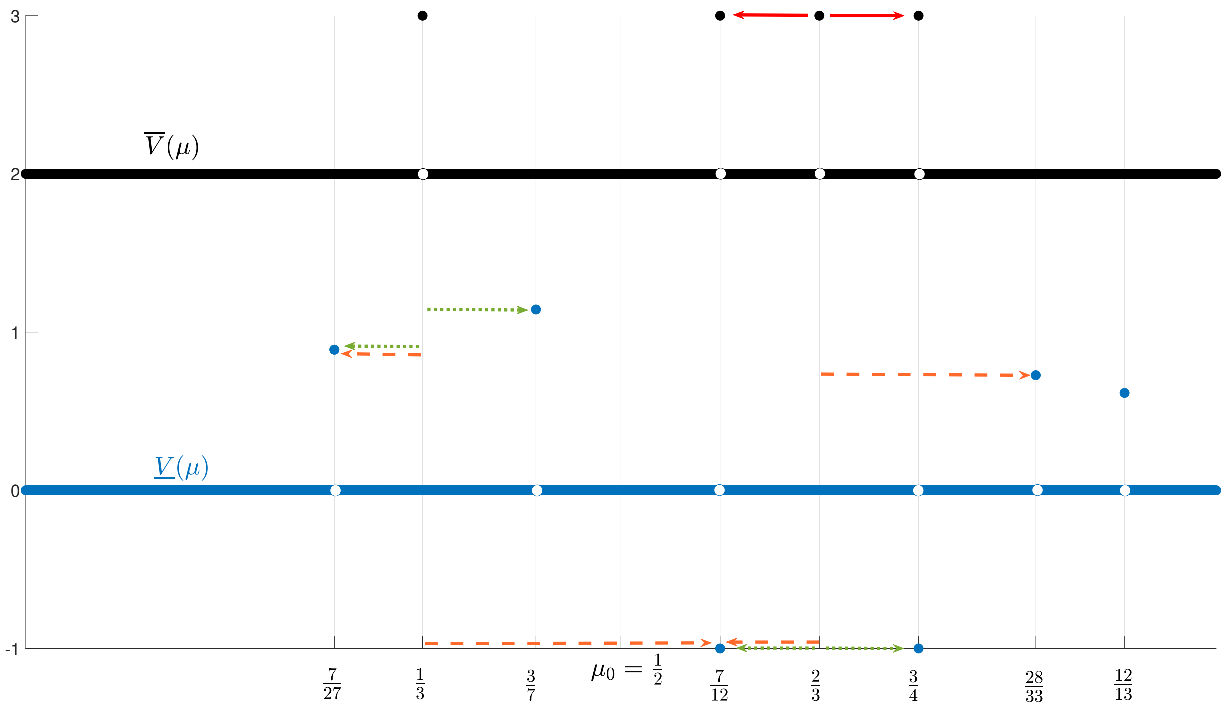

Figure OA.2.1: The functions $\underline{V}$ and $\overline{V}$

ors 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.2.1 illustrates the value function $\overline{V}$ (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.[15] 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

[15] 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.

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. 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.2.1). 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.2.1 (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

$$\underline{V}\left(\frac{7}{27}\right) = \frac{8}{9}, \ \underline{V}\left(\frac{3}{7}\right) = \frac{8}{7}, \ \underline{V}\left(\frac{28}{33}\right) = \frac{8}{11}, \ \underline{V}\left(\frac{12}{23}\right) = \frac{8}{13}, \tag{OA.4}$$

| as illustrated in Figure OA.2.1. This completes the construction of the function $\underline{V}$, as summarized in the following claim.

Claim OA.2. *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.4) holds, and $\overline{V}(\mu) = 2$ except that $\overline{V}(1/3) = \overline{V}(7/12) = \overline{V}(2/3) = \overline{V}(3/4) = 3$. 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 $\overline{V}$, 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 $\overline{V}$ 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.