Econometrica Supplementary Material

SUPPLEMENT TO “PREPARING FOR THE WORST BUT HOPING FOR THE BEST: ROBUST (BAYESIAN) PERSUASION”

(*Econometrica*, Vol. 90, No. 5, September 2022, 2017–2051)

PIOTR DWORCZAK
Department of Economics, Northwestern University

ALESSANDRO PAVAN
Department of Economics, Northwestern University

OA.1. PROOFS FOR SECTION 5

OA.1.1. *Proof of Lemma 2*

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

OA.1.2. *Proof of Lemma 3*

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

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

a contradiction.

OA.1.3. *Proof of Lemma 4*

By Lemma 3, we can write

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

Piotr Dworczak: piotr.dworczak@northwestern.edu
Alessandro Pavan: alepavan@northwestern.edu

https://doi.org/10.3982/ECTA19107

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

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

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

OA.1.4. *Proof of Proposition 4*

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

Let $\mu_0^+ < \alpha$, as assumed in the main text. The following policy is then a Bayesian solution. The Sender randomizes over two announcements, $s = 0$ and $s = 1$. She announces $s = 0$ with certainty when $\omega > 0$, and with probability $(1 - \phi_{\mathrm{BP}}) \in (0, 1)$ when $\omega \leq 0$, with ϕ_{BP} satisfying $\mathbb{P}(\omega > 0|s = 0) = \mu_0^+/[\mu_0^+ + (1 - \mu_0^+)(1 - \phi_{\mathrm{BP}})] = \alpha$. To see that this is a Bayesian solution, first note that, without loss of optimality, the Sender can confine attention to policies with two signal realizations, $s = 0$ and $s = 1$, such that, when $s = 0$ is announced, $\mathbb{P}(\omega > 0|s = 0) \geq \alpha$ and all agents refrain from attacking, whereas when $s = 1$ is announced, $\mathbb{P}(\omega > 0|s = 1) < \alpha$ and all agents attack.[1] Next, note that, starting from any binary policy announcing $s = 1$ with positive probability over a positive measure subset of $\mathbb{R}_+$, one can construct another binary policy that announces $s = 0$ (thus inducing all agents to refrain from attacking) with a higher ex ante probability, contradicting the optimality of the original policy. Hence, any binary Bayesian solution must announce $s = 0$ with certainty for all $\omega > 0$. Furthermore, under any Bayesian solution, the ex ante probability $\sum_{\omega\in\Omega:\omega<0}\pi(0|\omega)\mu_0(\omega)$ is uniquely pinned down by the condition $\mathbb{P}(\omega > 0|s = 0) = \mu_0^+/[\mu_0^+ + \sum_{\omega\in\Omega:\omega<0}\pi(0|\omega)\mu_0(\omega)] = \alpha$. Because the Sender's preferences depend only on $1 - A$, the specific way the policy announces $s = 0$ when $\omega < 0$ is irrelevant, thus implying that the binary policy described above is indeed a Bayesian solution. By the same token, the above binary policy is payoff-equivalent to one that announces $s = 0$ with certainty when $\omega > 0$, whereas when $\omega < 0$, it fully reveals the state with probability ϕ_{BP}, and announces $s = 0$ with the complementary probability. The signal realization $s = 0$ can then be interpreted as the decision not to disclose any information (equivalently, as the "null signal"$s = \emptyset$), as claimed in the proposition.

To see that the above Bayesian policy is not robust, let $\mu^{(0,1]} := \mu(\omega \in (0, 1])$ denote the probability that μ assigns to the interval $(0, 1]$. Recall that, given any posterior μ, if $\mu^+ := \mu(\omega > 0) < \alpha$, the unique rationalizable action is to attack. If $\mu^+ \in [\alpha, \alpha + \mu^{(0,1]}]$,

[1] 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.

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

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

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

OA.2. AUXILIARY RESULTS FOR SECTION 6

OA.2.1. *Relaxing the Regularity Assumption in Theorem 2*

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

THEOREM OA.1: *If $\lambda_n \nearrow 1$, and $\rho_n \in S(\lambda_n)$ converges to ρ in the weak* topology as $n \to \infty$, 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,

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

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

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

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

$$\int_{\Delta^c_{\mathcal{F}}\Omega} d(\mu, \Delta_{\mathcal{F}}\Omega)\, d\rho_n(\mu) \to 0.$$

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

subject to the Bayes-plausbility constraint. Because the above function is increasing in m, any λ-solution, for $\lambda < 1$, attaches probability $m^\star$ to 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 $\lambda \nearrow 1$. Hence, all limits of λ-solutions put at least $1/3$ mass on $(1,0,0)$, which is what we wanted to prove.

OA.2.2. *Example Showing That Bayesian Solutions Can Be Dominated*

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

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

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

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

OA.3. SIMULTANEOUS-MOVE-ROBUST SOLUTIONS

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

To simplify exposition, we work with the baseline model of Section 2, except that we allow for general conjectures. Unless specified otherwise, we maintain all the assumptions imposed in the main text.

The Sender continues to choose an information structure $q : \Omega \to \Delta\mathcal{S}$, which maps states ω into probability distributions of signal realizations $s \in \mathcal{S}$, but we do not assume that $\mathcal{S}$ is finite (this would be with loss of generality). We also modify Nature's strategy space: Nature selects a signal $\pi : \Omega \to \Delta\mathcal{R}$ that is independent of the Sender's signal conditional on the state, with a signal space $\mathcal{R}$ that is potentially infinite. Let Π_{CI} be the set of signals available to Nature, where "CI" stands for "conditionally independent."[2]

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

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

where ξ_0 is the conjectured tie-breaking rule, with $\xi_0(A^\star(\mu)|\mu) = 1$ for all μ.[3] We can similarly define $\widehat{V}$ as in formula (4.1) in Section 4, except that the conjecture about Nature is that it uses a signal $\pi_0 \in \Pi_{\text{CI}}$ (π_0 is not a function of the posterior belief generated by the Sender). Throughout, we assume that $\widehat{V}$ is upper semicontinuous.

Let

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

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

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

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

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

$$q \in \operatorname*{argmax}_{q'\in Q} \underline{v}(q', \pi),$$

$$\pi \in \operatorname*{argmin}_{\pi'\in\Pi_{\text{CI}}} \underline{v}(q, \pi').$$

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

[2] We assume that $\mathcal{R}$ and $\mathcal{S}$ are subsets of some sufficiently rich but fixed space.

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

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

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

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

PROOF: (1) $\implies$ (2). Suppose that $q \in \operatorname{argmax}_{q'\in Q}\{\inf_{\pi\in\Pi_{\mathrm{CI}}} \underline{v}(q',\pi)\}$. We argue that (q, π_{full}) is a Bayes–Nash equilibrium of the simultaneous-move game between Nature and the Sender, where π_{full} is the full-disclosure signal. Optimality of q for the Sender is trivial since any policy q' leads to the full-disclosure payoff $\underline{V}_{\mathrm{full}}(\mu_0)$ against π_{full}. Optimality of π_{full} for Nature follows from the fact that q maximizes $\inf_{\pi\in\Pi_{\mathrm{CI}}} \underline{v}(q',\pi)$ over all $q'\in Q$, which implies that, given q, Nature cannot bring the Sender's payoff below $\underline{V}_{\mathrm{full}}$.[4]

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

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

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

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

OA.3.1. *Summary of Results*

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

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

- Corollary 1 fails for SM-robust solutions, that is, a SM-robust solution may fail to exist. We show in Section OA.3.3 (Theorem OA.2) that a SM-robust solution exists under a stronger assumption of continuity of $\underline{V}$. Moreover, we introduce a notion of weak SM-robust solutions (that relaxes the condition of SM-worst-case optimality), and show that a weak SM-robust solution exists under no further assumptions on $\underline{V}$.

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

- In Section OA.3.5, we provide a sufficient condition (Theorem OA.3) for state separation under a SM-robust solution. This condition is weaker than the one in Corollary 2; that is, whenever two states must be separated under a SM-robust solution, they also must be separated under a robust solution.
- Corollary 4 does not extend to SM-robust solutions because we do not have a characterization similar to the one in Theorem 1. In Section OA.3.2 and Section OA.3.5, we obtain various (weaker) sufficient conditions for either full-disclosure to be the unique SM-robust solution, or for all distributions to be SM-worst-case optimal.
- In Section OA.3.4, we analyze the binary-state case. Unlike robust solutions, as described by Corollary 3, SM-robust solutions for binary-state problems may coincide with neither Bayesian solutions nor full disclosure. However, we give sufficient conditions for Bayesian solutions and full disclosure, respectively, to constitute SM-robust solutions.
- In Section OA.3.6, we show that Corollary 5 and Corollary 6 fail for SM-robust solutions. That is, it is possible that a Bayesian solution is strictly more informative than all SM-robust solutions.
- Corollaries 7 and 8 also fail: In fact, a SM-robust solution may require infinitely many signal realizations even when the state space is finite.

OA.3.2. *Preliminary Observations*

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

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

PROOF: Observe that, for any $\pi : \Omega \to \Delta\mathcal{R}$,

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

where

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

Therefore,

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

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

$\underline{V}(\mu)$ is lower semicontinuous, so is $\underline{V}_q(\mu)$, for any q. Formally, for any sequence $\{\mu_n\}$ of posterior beliefs over Ω converging to $\mu \in \Delta\Omega$, we have that

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

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

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

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,

$$\begin{aligned}
&\sum_{\omega, r \in \Omega} \int_S \underline{V}(\mu_0^{s,r}) \pi(r|\omega)\, dq(s|\omega) \mu_0(\omega) \\
&\quad = \int_S \underbrace{\left(\sum_{\omega, r \in \Omega} \underline{V}(\mu_0^{s,r}) \pi(r|\omega) \mu_0^s(\omega) \right)}_{\underline{V}_\pi(\mu_0^s)} \left(\sum_{\omega \in \Omega} dq(s|\omega) \mu_0(\omega) \right),
\end{aligned}$$

where

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

and hence

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

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

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

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

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

OA.3.3. *Existence*

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

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

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

and let $\widehat{V}$ and $\underline{V}$ be, respectively, the pointwise highest and lowest selection from the correspondence $\mathcal{V}$. Then $\widehat{V}$ is continuous, whereas $\underline{V}$ has a discontinuity at $\mu=1/2$. A distribution ρ is SM-worst-case optimal if and only if it guarantees the Sender a payoff of 0 (this is the payoff from full disclosure of the binary state). Any Bayes-plausible continuous distribution of posterior beliefs (e.g., $\rho\in\Delta\Delta\Omega$ that is uniform on $[0,1]$) yields a payoff guarantee of 0 because Nature cannot induce a posterior belief of 1/2 with positive probability. This conclusion relies crucially on the assumption that Nature's signal is conditionally independent of the Sender's signal.

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

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

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

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

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

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

PROOF: Define

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

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

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

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

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

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

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

When Nature can send arbitrary signals, including signals that are correlated with the Sender's signal, existence of robust solutions does not require the additional assumption that $\underline{V}$ is continuous (see Corollary 1 in the main text). This is because, in that case, given any induced posterior μ, adversarial Nature always brings the conditional expected payoff of the Sender down to $\mathrm{lco}(\underline{V})(\mu)$—the lower convex closure of $\underline{V}$ evaluated at μ. The lower convex closure is a convex function, and convex functions are continuous on the interior of the domain. This guarantees that the set $\mathcal{W}$ of worst-case optimal distributions is closed, while, in general, the set of SM-worst-case optimal distributions $\mathcal{W}_{\mathrm{SM}}$ need not be closed.

OA.3.4. *SM-Robustness for the Binary State*

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

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

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

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

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

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

$$\begin{aligned}
U'_\rho(1) &= \lim_{\epsilon \to 0} \frac{U_\rho(1) - U_\rho(1-\epsilon)}{\epsilon} \\
&= \mu_0 \underline{V}(1) - \lim_{\epsilon \to 0} \frac{\displaystyle\int_0^1 (1-\mu(1-\epsilon))\underline{V}\left(\frac{\epsilon\mu}{1-\mu(1-\epsilon)}\right) d\rho(\mu)}{\epsilon} \\
&\overset{(1)}{=} \mu_0 \underline{V}(1) - \int_{[0,1)} \left(\lim_{\epsilon \to 0} \frac{\underline{V}\left(\dfrac{\epsilon\mu}{1-\mu(1-\epsilon)}\right)}{\dfrac{\epsilon\mu}{1-\mu(1-\epsilon)}} \frac{\mu - \mu^2 + \mu^2\epsilon}{1-\mu+\mu\epsilon} \right) d\rho(\mu) - \underline{V}(1)\rho(1) \\
&= \mu_0 \underline{V}(1) - \underline{V}'(0)\big[\mu_0 - \rho(1)\big] - \underline{V}(1)\rho(1) \\
&= \big[\mu_0 - \rho(1)\big]\big[s - \underline{V}'(0)\big] > 0, \qquad \text{(OA.2)}
\end{aligned}$$

as long as $\mu_0 > \rho(1)$—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-disclosure payoff $\underline{V}_{\text{full}}(\mu_0)$ by selecting a binary signal π that is almost fully revealing. Therefore, full disclosure is the unique SM-worst-case optimal distribution, and hence the unique SM-robust solution. *Q.E.D.*

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

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

Loosely speaking, the Sender fully reveals the state not because she is worried that, else, Nature will do it, but because she is worried that Nature will *almost* fully reveal the state. Under the conditions in Proposition OA.1, almost full revelation is strictly worse than full revelation.

The above intuition can also be used to compare SM-worst-case optimality to worst-case optimality (and hence SM-robustness to robustness). As explained in the main text, a sufficient condition for full disclosure to be the unique robust solution is that the payoff $\underline{V}(\mu)$ lies below the full-disclosure payoff $(1-\mu)\underline{V}(0) + \mu\underline{V}(1)$ at *some* interior $\hat{\mu}$. A sufficient condition for full disclosure to be the unique SM-robust solution is that $\underline{V}(\mu)$

is below the full-disclosure payoff $(1-\mu)\underline{V}(0)+\mu\underline{V}(1)$ for μ sufficiently close to one of the two bounds, $\mu=0$ or $\mu=1$. When Nature can condition her disclosure on the *realization* of the Sender's signal (equivalently, on the posterior μ induced by the Sender), for any interior μ, Nature can induce the "final" posterior belief $\hat{\mu}$ with positive probability, without restricting its own ability to influence the Receivers' beliefs conditional on other realizations of the Sender's signal. In contrast, when Nature's signal is conditionally independent, and Nature chooses to induce the posterior belief $\hat{\mu}$ with positive probability conditional on the Sender inducing μ, it can no longer independently choose what posterior beliefs the Receivers will have conditional on other realizations of the Sender's signal. In particular, even if Nature's signal realization shifts μ to a $\hat{\mu}$ that yields a low payoff to the Sender, the same signal realization could shift some other η induced by the Sender to a $\hat{\eta}$ that has a high payoff to the Sender. In short, Nature cannot target the same posterior belief $\hat{\mu}$ regardless of the realization of the Sender's signal.

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

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

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

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

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

OA.3.5. *State Separation Under SM-Robustness*

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

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

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

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

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

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

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

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

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

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

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

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

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

OA.3.6. *A Bayesian Solution Can Blackwell Dominate a SM-Robust Solution*

Corollary 6 in the main text states that, for any Bayesian solution ρ_{BP}, one can find a robust solution ρ_{RS} that is either incomparable to, or more informative than, ρ_{BP} in the

Blackwell sense. In this subsection, we show that this conclusion does not extend to SM-robust solutions. We do this by means of a counterexample. Our counterexample is rather contrived and has no immediate economic interpretation.

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

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

Counterexample

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

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

The idea is to construct a function $\underline{V}$ under which the Sender gets a low payoff from beliefs $7/12$ and $3/4$, so that ρ_{BP} is not SM-worst-case optimal. Suppose that $\underline{V}(\mu) = 0$ except over a finite set of points, and that $\underline{V}(7/12) = \underline{V}(3/4) = -1$. Then, ρ_{BP} is clearly not SM-worst-case optimal, because by not disclosing any information, Nature guarantees that the Sender's expected payoff under ρ_{BP} is strictly below her full-disclosure payoff, which is equal to zero. Note, however, that this is not enough, because under such $\underline{V}$, ρ_R is also not SM-worst-case optimal. Indeed, starting from any interior posterior belief, by

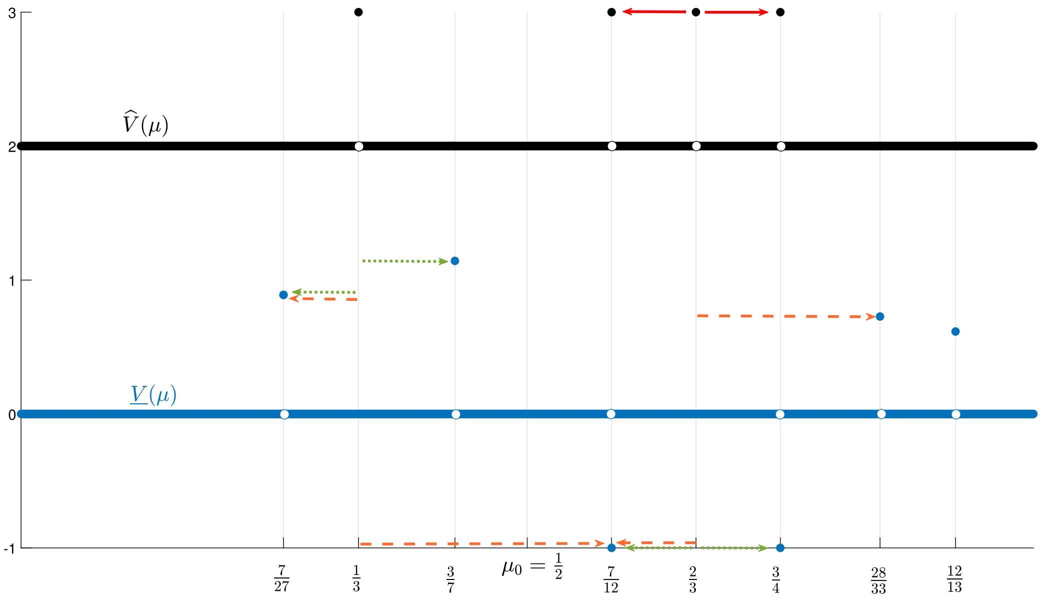

FIGURE OA.1.—The functions $\underline{V}$ and $\widehat{V}$.

choosing π appropriately, Nature can induce the posterior $\mu = 7/12$ and/or the posterior $\mu = 3/4$ with positive probability, thus bringing the Sender's payoff strictly below the full-disclosure payoff. Therefore, we construct $\underline{V}$ so that, whenever Nature's response results in a low payoff for the Sender conditional on one of her posterior beliefs under ρ_R, it must result in a sufficiently high payoff for the Sender conditional on the other posterior belief. For example, Nature can split 2/3 into 7/12 and 3/4; however, the unique signal that achieves this split—precisely because the signal must be conditionally independent—must also split the other posterior 1/3 into 7/27 and 3/7 (as illustrated by the green dotted arrows in Figure OA.1). Thus, we choose the values of $\underline{V}$ to be sufficiently high at 7/27 and 3/7.

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

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

$$= \left(\frac{2}{3} - \frac{1}{3}\mu\right)\underline{V}\left(\frac{\mu}{2-\mu}\right) + \left(\frac{1}{3} + \frac{1}{3}\mu\right)\underline{V}\left(\frac{2\mu}{1+\mu}\right).$$

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

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

as illustrated in Figure OA.1. This completes the construction of the function $\underline{V}$. The following claim is then true.[5]

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

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

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

REFERENCES

KAMENICA, EMIR, AND MATTHEW GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615. [9,11,13]

Co-editor Barton L. Lipman handled this manuscript.

Manuscript received 29 October, 2020; final version accepted 22 March, 2022; available online 25 March, 2022.

[5] Note that, contrary to what we assumed throughout the analysis, the function $\underline{V}$ considered in this example is not lower semicontinuous. 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 semicontinuous functions.