Robust Narrative Persuasion

Song Li*

Universitat Autònoma de Barcelona and Barcelona School of Economics

July 28, 2026

Abstract

This paper studies a sender who designs a public menu of narratives for a heterogeneous population whose cross-sectional distribution of priors is known only through moment restrictions. After observing a public event, each receiver selects the narrative that best fits his prior and updates within it. I show that under binary states, one threshold-setting narrative per realization suffices. The implementable set is bounded by a persuasion frontier, along which there is a trade-off across signal realizations. In a political-narrative application, the dictator's optimal value can fall as the average prior rises: a higher average prior can let Nature meet the variance lower bound with less mass strictly above the persuasion threshold.

JEL Classification: D82; D83; D81; C72.
Keywords: Persuasion; Narratives; Robustness; Ambiguity.

1 Introduction

The same public event can support opposing interpretations. Voters may read an election outcome as a legitimate mandate or as evidence of manipulation; consumers may read a product review as quality evidence or as a paid endorsement; jurors may read forensic evidence as proof of guilt or as procedurally compromised.

Each such interpretation is a *narrative*: a likelihood model that assigns probabilities to possible observations in each state of the world. Each receiver adopts the narrative that best fits what he observes under his prior.[1] Which narrative he chooses depends on what he already believes. After an unfavorable protest report, skeptical citizens may see evidence of regime weakness, while loyal citizens may see evidence of foreign interference.

A sender offering a public menu of narratives must anticipate this self-selection. By designing the menu, she shapes narrative selection. For a given realization, persuasion

*I am especially grateful to Fernando Payró for his invaluable guidance and generous support throughout this project. I also thank Antoine Zerbini for helpful feedback and discussions.

[1]Throughout, I refer to the sender as she and each receiver as he.

succeeds for a receiver if his selected narrative induces a sender-favorable action. If few receivers lie near the margin of being persuaded, a population with a favorable average prior may still be hard to persuade. Conversely, a population with a lower average prior can be more responsive if its mass is concentrated among the marginal receivers. This makes the cross-sectional distribution of priors central.

In practice, the sender rarely observes the population distribution. I call this uncertainty *ambiguity*. A dictator infers public dissent through rally attendance and social media monitoring; a seller derives consumer attitudes from aggregated market surveys; a prosecutor prepares for jury selection using records of collective past behavior. Such information typically takes the form of sample statistics, such as averages and measures of dispersion. These sample statistics define an *ambiguity set*. The sender evaluates each menu against all distributions in this set; her worst-case expected payoff over this set is the menu's *payoff guarantee*. I ask: How should she design a public menu of narratives to maximize her payoff guarantee in a heterogeneous population?

To answer this question, I formulate a problem of robust narrative persuasion. The sender searches for an optimal public menu that is *robust* in two senses. First, it performs well against the most unfavorable distribution in the ambiguity set, as if a hypothetical Nature were choosing the distribution adversarially. Second, its payoff guarantee is stable to small perturbations of the ambiguity set: when the ambiguity set is slightly misspecified, the payoff guarantee does not drop discontinuously.

In the paper's framework, the sender commits ex ante to a public menu of narratives, each explaining the signal realizations produced by the true information structure, which I call the *data-generating model* following Jain (2023).[2] This departs from Bayesian persuasion (Kamenica and Gentzkow, 2011), where the sender chooses the true information structure.[3]

The public nature of the menu requires the sender to anticipate varied reactions across receivers; its choice before the realization makes her account for how a receiver's narrative selection changes across realizations; and her knowledge of only sample statistics about priors requires the menu to perform well for every distribution consistent with those statistics. Therefore, the sender faces a public-menu design problem under ambiguity. The same design problem also fits a single receiver whose prior is privately known: the sender is then uncertain both about that prior and about the distribution it is drawn from, knowing only sample statistics of the latter.

The key tradeoff arises around the margin of persuasion. After each realization, the menu separates receivers who take the sender-preferred action from those who do not. Given any menu, Nature selects from the ambiguity set the least favorable distribution, placing as much mass as possible among receivers who are not persuaded by that menu. The sender chooses the menu with the highest payoff guarantee, but because the same menu is used across realizations, making persuasion easier after one realization can make it harder after others.

Under binary states, this logic reduces to a threshold profile. One threshold-setting narrative per realization is enough, and implementability is characterized by a persuasion frontier in threshold-odds coordinates.

[2]By "explain," I mean specifying a probabilistic relationship between states and signal realizations.

[3]See Bergemann and Morris (2019) and Kamenica (2019) for surveys of Bayesian persuasion and its subsequent extensions.

I illustrate this tradeoff in an application. There, I assume that whenever a receiver believes it is weakly more likely to be in the state where the sender-preferred action is also the correct choice, he takes the sender-preferred action. Under this assumption, a boundary menu is optimal: after the realization with the larger sender payoff weight, every receiver takes the sender-preferred action; after the other realization, the sender earns payoff only from receivers whose priors are high enough. A tighter lower bound on prior dispersion weakly raises the optimal value, and raises it strictly when this lower bound binds in Nature's least favorable distribution. A higher average prior can lower the optimal value if Nature can satisfy the higher mean with less mass above the margin of persuasion.

Related literature

This paper builds on the model-as-likelihoods approach to model persuasion introduced by Schwartzstein and Sunderam (2021).[4] Aina (2025) extends this line to ex ante menu choice mainly for a single receiver with a known prior. I keep the ex ante formulation and extend it to a public menu for a heterogeneous population, studying how partial knowledge of the cross-sectional distribution of priors changes menu design. As noted above, the same design problem also describes a single receiver whose prior is privately known and whose distribution the sender knows only through sample statistics, so the paper's analysis relaxes the known-prior assumption in Aina (2025).

The paper's focus on a common menu connects the analysis to public persuasion. Laclau and Renou (2016) analyze the restrictions prior heterogeneity imposes when a sender influences multiple receivers through a single public signal; Alonso and Câmara (2016) study a voting application; Inostroza and Pavan (2025) study public information design under worst-case adversarial coordination when receivers also possess exogenous private information.[5] While those papers study standard Bayesian persuasion where the sender commits to a signal structure, the public commitment object here is a menu of narratives, and heterogeneity enters through receivers' fit-based selection from that menu.

The narrative-menu setting also yields different comparative statics for dispersion from the Bayesian persuasion setting in Gitmez and Molavi (2022): in their model, greater polarization narrows the scope for public information manipulation, whereas here a tighter lower bound on prior variance shrinks Nature's feasible set and weakly raises the payoff guarantee.

I also contribute to robust persuasion by analyzing narrative design under an ambiguous cross-sectional distribution of priors, whereas existing work focuses on information design. Hu and Weng (2021) study uncertainty about the receiver's private information source; Dworczak and Pavan (2022) study robustness to exogenous information and adversarial strategy selection; Sapiro-Gheiler (2024) studies ambiguity in receiver preferences via a maxmin criterion over cutoff distributions; Rosenthal (2026) develops a data-driven approach in which the

[4] Another approach represents narratives as causal models via directed acyclic graphs (Spiegler, 2016; Eliaz and Spiegler, 2020; Eliaz et al., 2021).

[5] Other multi-receiver public-information work studies different forms of cross-receiver structure: Bardhi and Guo (2018) characterize sender-optimal experiments under unanimous consent with heterogeneous thresholds of doubt; a separate game-theoretic line studies information design with strategic interaction among receivers, e.g., Mathevet et al. (2020). See also Doval and Smolin (2024) on the welfare consequences of information policies in a population setting.

sender infers plausible priors from observed action distributions.[6] Kosterina (2022) is closest, studying an uncertain individual prior under a sender-designed information structure. The paper's setting differs on two counts. The instrument is different: the data-generating model is fixed, while Kosterina's sender designs the information structure. The uncertainty is also different: she hedges against the worst prior in a set, whereas here the sender is uncertain about both the prior and, beyond it, the distribution generating the prior, of which she knows only summary statistics such as the average prior and its dispersion. The population reading recasts that distribution as an unknown cross-sectional distribution of priors across an audience.

I apply the robustness criterion and the continuous moment ambiguity of Ball and Kattwinkel (2026); introducing additional structural assumptions provides a tractable way to solve the inner problem. Beyond persuasion, such moment restrictions are widely used in robust mechanism design, including selling, principal-agent, and reserve-pricing problems (Carrasco et al., 2018, 2019; Suzdaltsev, 2022). This specification further links the paper's framework to partial identification (Manski, 2003), where restrictions define an identified set for the cross-sectional distribution of priors, and sampling uncertainty is captured by estimated moment sets or confidence regions (Imbens and Manski, 2004).[7]

The paper proceeds as follows. Section 2 introduces narratives, fit-based selection, and moment ambiguity. Section 3 characterizes menu reduction, implementability, and robust payoff guarantees. Section 4 presents the political-narrative application under the population reading. Section C presents a product-review application under the single-receiver reading. Proofs are deferred to the Appendix.

2 Model

2.1 Primitives

The state space is $\Omega = \{H, L\}$ and signal space S is finite. The receiver chooses an action $a \in \{0, 1\}$, where $a = 1$ denotes the sender-preferred action.

The sender's payoff from a receiver depends on that receiver's action and the state, $U_S : \{0, 1\} \times \{H, L\} \to \mathbb{R}$. A receiver's payoff depends on his action and the state, $U_R : \{0, 1\} \times \{H, L\} \to \mathbb{R}$.

Assumption 1 (Sender's preferred action)
The sender prefers action 1 *in every state:*

$$U_S(1, H) > U_S(0, H) \qquad \textit{and} \qquad U_S(1, L) > U_S(0, L).$$

This assumption fixes persuasion as inducing action 1. It does not require the sender's payoff gain from action 1 to be the same across states.

The sender has a prior $\tilde{\mu} \in \Delta(\{H, L\})$. Each receiver has a prior $\theta \in [0, 1]$, representing his probability of state H. I identify receiver types with their prior on H and write $\Theta := [0, 1]$

[6] Alternative robustness criteria include regret minimization (Babichenko et al., 2022) and minimax-loss (Parakhonyak and Sobolev, 2026).

[7] Set-valued parameters and uncertainty are studied within a random-set framework by Molchanov and Molinari (2018).

for the type space. Types are drawn from a *cross-sectional distribution* $F \in \Delta(\Theta)$, where $\Delta(\Theta)$ is endowed with the weak topology. In the *benchmark problem*, the sender knows F; in the *robust problem*, she faces ambiguity about F, as in Section 2.3.

The *data-generating model* $\pi : \Omega \to \Delta(S)$ is the true model generating signal realizations; it specifies a conditional distribution $\pi(\cdot \mid \omega) \in \Delta(S)$ for each state ω. This model is perfectly observed by the sender, while receivers only observe the signal realization s.

Assumption 2 (Regularity)
(a) $\tilde{\mu}(H) > 0$ *and* $\tilde{\mu}(L) > 0$*; and (b) for every* $s \in S$*,* $\pi(s \mid H) > 0$ *or* $\pi(s \mid L) > 0$*.*

Assumption 2 rules out degenerate cases by ensuring both states have positive ex ante weight and every realization can occur, allowing the menu problem to be stated without irrelevant realizations.

2.2 Narratives

A *menu of narratives* is a finite set $M = \{m_1, \ldots, m_J\}$.[8] Each narrative m_j is a map $m_j : \{H, L\} \to \Delta(S)$, assigning a distribution over realizations to each state. For each integer $J \geq 1$, $\mathcal{M}_J$ denotes the set of all menus with J narratives; the *menu space* is $\mathcal{M} := \bigcup_{J \geq 1} \mathcal{M}_J$, unrestricted relative to the data-generating model π. This convention follows Schwartzstein and Sunderam (2021) and subsequent work.

What governs the realization remains π, while the menu space concerns receivers' feasible interpretations. Receivers do not observe π or construct narratives outside M: the announced menu is the only source of narratives through which they can interpret the realization.

For each realization $s \in S$ and each receiver type $\theta \in [0, 1]$, the *fit* of narrative m_j is

$$\mathrm{Fit}(s, m_j, \theta) := (1 - \theta)\, m_j(s \mid L) + \theta\, m_j(s \mid H), \tag{1}$$

the marginal likelihood of s under narrative m_j given prior θ, so higher fit means that the narrative makes the observed realization more likely for that receiver.

Given $s \in S$, $\theta \in [0, 1]$ and $M \in \mathcal{M}$, each receiver selects

$$m^*(s, \theta) \in \arg\max_{m \in M} \mathrm{Fit}(s, m, \theta).$$

As the fit of each narrative depends on the prior, after a common realization s, receivers with different priors may select different narratives. The menu thus groups types according to the narratives they select.

With a slight abuse of notation, when the realization is s and the type is θ, I write $m^*(s \mid H)$ and $m^*(s \mid L)$ for the likelihoods assigned to s in each state by the selected narrative $m^*(s, \theta)$.

Following narrative selection, whenever $\mathrm{Fit}(s, m^*(s, \theta), \theta) > 0$, the receiver applies Bayes' rule to update his belief. Writing $\hat{\mu}_s(\theta)$ for the posterior probability of state H:

$$\hat{\mu}_s(\theta) := \frac{\theta\, m^*(s \mid H)}{(1 - \theta)\, m^*(s \mid L) + \theta\, m^*(s \mid H)}. \tag{2}$$

[8] I use narrative and model interchangeably only for what the sender offers.

Given the posterior, the receiver chooses an action to maximize his expected payoff

$$a^*(s, m^*(s,\theta), \theta) \in \arg\max_{a\in\{0,1\}} \big[(1-\hat{\mu}_s(\theta))\, U_R(a, L) + \hat{\mu}_s(\theta)\, U_R(a, H)\big].$$

Write $a_M(\theta, s) := a^*(s, m^*(s,\theta), \theta)$ for the action of a type-θ receiver after realization s under menu M.

The sender's *per-type payoff*, namely type θ's contribution to her ex ante payoff, is

$$v(\theta; M) = \tilde{\mu}(H)\sum_{s\in S} \pi(s \mid H)\, U_S(a_M(\theta, s), H) + \tilde{\mu}(L)\sum_{s\in S} \pi(s \mid L)\, U_S(a_M(\theta, s), L), \tag{3}$$

where U_S is defined in Section 2.1.

Under the cross-sectional distribution F, the sender's payoff from menu M is

$$W(M, F) = \int_\Theta v(\theta; M)\, F(d\theta).$$

The timing is as follows. Types $\theta \in [0,1]$ are drawn from $F \in \Delta([0,1])$. The sender, with prior $\tilde{\mu} \in \Delta(\{H, L\})$, commits to a deterministic public menu $M \in \mathcal{M}$, and receivers observe M.[9] Nature draws $\omega \in \{H, L\}$ according to $\tilde{\mu}$ and then $s \in S$ according to $\pi(\cdot \mid \omega)$. Observing s, each receiver selects $m \in M$ maximizing $\mathrm{Fit}(s, m, \theta)$, updates his belief by (2), and chooses an action. Payoffs are realized.

This commitment rules out ex post tailoring: after s is realized, the sender cannot replace or supplement the announced menu and then ask receivers to interpret the realization only through the revised set.

Conventions Tie-breaking is sender-favorable, as in Aina (2025). If multiple narratives maximize fit, the receiver selects one that yields the highest expected payoff for the sender; if both actions maximize his utility, he chooses the sender-preferred one.

2.3 Ambiguity

In the robust problem, the sender only knows that the cross-sectional distribution F lies in $\Pi \subseteq \Delta(\Theta)$, and she evaluates each menu M by $V(M) := \inf_{F\in\Pi} W(M, F)$.[10]

Without further structure on Π, however, a worst-case criterion need not deliver robustness: payoff guarantees may vary discontinuously under small perturbations of Π. I thus impose continuous moment restrictions with an interiority condition. These restrictions capture what the sender can credibly estimate or calibrate about the distribution.

[9] The sender commits to one deterministic public menu. A public lottery over menus would deliver the mixed problem in Section A.8, which eliminates the max-min gap; I exclude that device because a one-shot public lottery over official frameworks is hard for a dispersed receiver to verify (Fréchette et al., 2022; Lin and Liu, 2024; Best and Quigley, 2024) and it is intuitive that ambiguity should matter for decision.

[10] The setting admits two interpretations. In the population interpretation, F is the cross-sectional distribution of priors across multiple receivers, and Π collects the receiver compositions the sender regards as possible. In the single-receiver interpretation, the sender faces one receiver whose prior θ is unknown and drawn from a distribution F that is only partially known through moment restrictions, and Π collects the possible distributions of the receiver's prior. Both interpretations induce the same optimization problem, so the analysis that follows applies to either. I adopt the population interpretation for exposition.

Definition 1 (Moment ambiguity set)
Fix an integer $d \geq 1$, a continuous vector-valued function $g : \Theta \to \mathbb{R}^d$, called the moment map*, and a non-empty closed convex set $Y \subseteq \operatorname{conv} g(\Theta)$, called the* moment set*. The* moment ambiguity set *generated by (g, Y) is*

$$\Pi = \mathcal{F}(g, Y) := \{F \in \Delta(\Theta) : \mathbb{E}_{\theta \sim F}[g(\theta)] \in Y\}.$$

I impose the following interiority condition on the moment set $Y \subset \operatorname{conv} g(\Theta)$.[11] The set Y is *star g-interior*: there exists $y_0 \in Y \cap \operatorname{ri}(\operatorname{conv} g(\Theta))$ such that $[y_0, y] \subseteq Y$ for every $y \in Y$. This rules out moment information whose feasible moment set has no relative-interior star center, although Y may still contain boundary points of the feasible moment region. Since the sender's summary statistics may be estimated with small errors, the payoff guarantee should not depend on exact boundary equalities alone. By Ball and Kattwinkel (2026, Propositions 5 and 6(i)), this regularity places Π in their class of continuous moment sets, which are rich and hence uniformly robust. A cross-sectional distribution that is close to Π but slightly misses the moment restrictions can be adjusted back into Π by reallocating only a vanishing amount of probability mass. For distributions near Π, any shortfall below the sender's payoff guarantee is uniformly small across admissible menus.

Example 1 (mean constraints)
Consider a mean constraint, $g(\theta) = \theta$, with $Y = [\underline{\theta}, \overline{\theta}] \subseteq [0, 1]$. The ambiguity set is $\Pi = \{F \in \Delta(\Theta) : \underline{\theta} \leq \mathbb{E}_F[\theta] \leq \overline{\theta}\}$, which collapses to a fixed mean $\Pi = \{F : \mathbb{E}_F[\theta] = \hat{\theta}\}$ when $\underline{\theta} = \overline{\theta} = \hat{\theta}$. Here Y is star g-interior when it intersects $(0, 1)$; among nonempty closed convex subsets of $[0, 1]$, this excludes only the degenerate singletons $Y = \{0\}$ and $Y = \{1\}$.

Example 2 (mean and variance constraints)
Consider a mean constraint and a lower bound on variance: $\Pi = \{F \in \Delta(\Theta) : \mathbb{E}_F[\theta] = \hat{\theta},\ \operatorname{Var}_F(\theta) \geq \underline{\sigma}^2\}$, where $0 < \hat{\theta} < 1$ and $0 < \underline{\sigma}^2 < \hat{\theta}(1 - \hat{\theta})$. This can be written as a moment-constraint set by taking $g(\theta) = (\theta, \theta^2)$ and $Y = \{(\hat{\theta}, q) : q \in [\hat{\theta}^2 + \underline{\sigma}^2, \hat{\theta}]\}$, where the lower endpoint imposes the variance lower bound and the upper endpoint follows from $\theta^2 \leq \theta$ on $[0, 1]$. The resulting Y is a vertical segment, and the strict bound $0 < \underline{\sigma}^2 < \hat{\theta}(1 - \hat{\theta})$ ensures that it is star g-interior. Notice that here Y is also closed and convex.

3 Results

Every signal realization generates one persuasion threshold. Since the state space is binary and each receiver has state-matching preference, narrative selection and action choice are cutoff-based: after each $s \in S$, the menu separates types at a threshold θ_s, distinguishing those who take the sender-preferred action from those who do not. Thus a menu matters for payoffs only through the threshold profile $(\theta_s)_{s \in S}$. The main results characterize this profile: one threshold-setting narrative per realization is enough, implementability is described by odds-coordinate halfspaces, and the robust inner problem is solved for the induced threshold payoff.

[11] Alternatively, one could allow Y to be larger and impose the condition on the relevant set $Y \cap \operatorname{conv} g(\Theta)$. I adopt the simpler formulation in which $Y \subseteq \operatorname{conv} g(\Theta)$.

3.1 Benchmark

The paper first studies the no-ambiguity problem, in which the cross-sectional distribution of priors F is known. It provides the benchmark for the robust analysis. Even when the population is known, commitment to one menu constrains persuasion across realizations. The sender would like to make each realization persuasive for as many receivers as possible. However, after every realization, receivers choose from the same set of narratives: a narrative that lowers the cutoff after one realization can also become an attractive interpretation after another realization. Hence the payoff-relevant cutoffs cannot be chosen independently.

Receiver types $\theta \in [0,1]$ are scalar priors on H, as in Section 2.1.[12]

I impose the following assumption on receiver preferences.

Assumption 3 (Receiver preferences)
The receiver has state-matching *preferences:*

$$u_H := U_R(1,H) - U_R(0,H) > 0 \qquad \textit{and} \qquad u_L := U_R(1,L) - U_R(0,L) < 0.$$

Assumptions 1 and 3 impose a simple conflict of interest: the sender always prefers action 1, regardless of the state, while the receiver prefers action 1 in state H and action 0 in state L.

Two consequences follow. The sender's payoff depends on the menu only through the induced action rule, and the receiver chooses $a = 1$ if and only if his posterior on H is at least the *action threshold* $\tau := |u_L|/(u_H + |u_L|) \in (0,1)$. Both action choice and narrative selection are then cutoffs in type space, collapsing the problem of choosing the optimal menu to the choice of persuasion thresholds defined next.

Persuasion thresholds and sender payoff structure. Under binary states, each signal realization $s \in S$ induces one endogenous persuasion threshold. For a fixed realization s, fit is affine in θ, with slope $m(s \mid H) - m(s \mid L)$. Along the upper envelope of these fit functions, selected slopes are nondecreasing in type, and sender-favorable tie-breaking selects the higher posterior at fit switches. Hence the set of types persuaded after s is an upper interval.

Definition 2 (Persuasion threshold)
Fix a menu $M \in \mathcal{M}$. For each realization $s \in S$ and type $\theta \in [0,1]$, let $m^*(s,\theta)$ be the narrative selected from M by the fit-based rule, and write $\hat{\mu}_s(\theta) := \mu_s^{m^*(s,\theta)}(\theta) \in [0,1]$ for the posterior of state H obtained by updating under the selected narrative. The persuasion threshold *after realization s under menu M is the type-space cutoff*

$$\theta_s(M) := \inf\{\theta \in [0,1] : \hat{\mu}_s(\theta) \geq \tau\}, \qquad s \in S, \tag{4}$$

with the convention $\theta_s = 1$ if the corresponding set is empty.

[12] Prior heterogeneity can also capture other forms of heterogeneity in reduced form. If, before the public signal, each receiver observes a private signal that affects his behavior only through the interim belief it induces, then a setting with a homogeneous prior and private signals is equivalent to the present formulation. Under Assumption 3, a setting with a homogeneous prior and heterogeneous action thresholds can likewise be represented, for any fixed selected narrative, by heterogeneous effective priors.

When the set defining θ_s is nonempty, the threshold θ_s is the marginal type after s. Types above θ_s are those whose selected narrative induces a posterior of at least τ, while types below θ_s are not persuaded. A lower θ_s means that the menu persuades more types after s.

Because the payoff $v(\theta; M)$ depends on M only through the threshold profile $(\theta_s(M))_{s\in S}$, define

$$\bar{u}_0 := \sum_{\omega} \tilde{\mu}(\omega) U_S(0,\omega), \qquad w_s := \sum_{\omega} \tilde{\mu}(\omega)\pi(s \mid \omega)\big[U_S(1,\omega) - U_S(0,\omega)\big], \quad s \in S.$$

By Assumptions 1 to 3, $w_s > 0$ for every $s \in S$. I call w_s the sender's *payoff weight* for realization s: it is her ex ante payoff increment from inducing one additional receiver to take action 1 after s. The constant $\bar{u}_0$ is her ex ante payoff when no receiver is persuaded; unlike the payoff weights, it does not depend on the thresholds. The profile $(\theta_s)_{s\in S}$ induces the per-type payoff

$$v(\theta; (\theta_s)_{s\in S}) = \bar{u}_0 + \sum_{s\in S} w_s \mathbf{1}\{\theta \geq \theta_s\}. \tag{5}$$

When the set defining θ_s is empty, the corresponding indicator in (5) is interpreted as the zero function. I then define

$$V((\theta_s)_{s\in S}) := \inf_{F\in\Pi} \int_{\Theta} v(\theta; (\theta_s)_{s\in S})\, F(d\theta), \tag{6}$$

and set $V(M) := V((\theta_s(M))_{s\in S})$. The benchmark problem is the singleton case $\Pi = \{F\}$, where (6) specializes to $\int v\, dF$; the robust problem in Section 3.2 uses a non-singleton Π.

For each realization with a nontrivial threshold, call a narrative *threshold-setting* if the marginal type selects it and reaches a posterior of at least τ. The selection convention in Section 2.2 is structural for the menu-reduction argument: after each realization, it implies that the types who are persuaded form an upper interval with cutoff θ_s.

Threshold-setting narratives simplify the problem by reducing the menu choice set from the whole menu space $\mathcal{M}$ to a smaller class. Keeping only one threshold-setting narrative for each signal realization, each retained narrative remains fit-maximal for its original marginal type, and, since the posterior induced by that narrative is increasing in θ, it also persuades all types above the original threshold after that realization.

Theorem 1 (Menu reduction)
Under binary states and signal space S, for every menu M, there exists a sub-menu $M' \subseteq M$ with $|M'| \leq |S|$ such that $v(\theta; M') \geq v(\theta; M)$ for every $\theta \in [0,1]$, and hence $V(M') \geq V(M)$.

The reduction shows that there is no loss in restricting attention to the narratives that set the realization-specific persuasion thresholds. An additional narrative that does not set such a threshold can matter only by changing which narrative some receivers select. It can reduce persuasion when it wins the fit comparison for some receivers while giving them a posterior below the action threshold.

Example 3 (Removing a non-threshold-setting narrative)
Let $M = \{(1/8, 3/8), (1/3, 1/3), (1, 0)\}$ and $\tau = 1/2$. After h, types on $[1/6, 1/3)$ select the

intermediate narrative from the full menu and switch to $(1,0)$ *only at* $1/3$*, so* $\theta_h(M) = 1/3$*; after* l*, the posterior under* $(1/8, 3/8)$ *crosses* τ *at* $5/12$*, giving* $\theta_l(M) = 5/12$*. The sub-menu* $M' = \{(1/8, 3/8), (1,0)\}$ *has a fit-crossing point* $3/10$*: types below* $3/10$ *select* $(1/8, 3/8)$ *and have posterior* $\theta/(3-2\theta) < 1/2$ *after* h*, while types above* $3/10$ *select* $(1,0)$ *and have posterior* 1*. Hence* $\theta_h(M') = 3/10 < 1/3 = \theta_h(M)$*, with* θ_l *unaffected since the* l*-side selection does not involve the removed intermediate narrative. The intermediate narrative wins on fit but yields lower posteriors; its presence delays persuasion to higher types. This illustrates Theorem 1: removing the narrative that was delaying persuasion after* h *lowers* θ_h *without affecting* θ_l*.*

Implementability. A threshold profile $(\theta_s)_{s\in S} \in [0,\tau]^{|S|}$ is *implementable* if some menu induces θ_s as the persuasion threshold after each realization $s \in S$. The next result characterizes exactly which threshold profiles can arise from a common menu.

The constraint is a cross-realization selection constraint. The sender uses one public menu after every realization, so a narrative that makes action 1 easier to induce after one realization may also become an attractive interpretation after another realization. Hence the realization-specific persuasion thresholds cannot be chosen independently. Implementability asks which threshold profiles can be generated while each threshold-setting narrative remains selected at its own marginal type.

Theorem 2 (Implementability)
Under binary states and signal space S*, the following statements hold.*

(i) A threshold profile $(\theta_s)_{s\in S} \in [0,\tau]^{|S|}$ *is implementable by some menu iff*

$$\sum_{s\in S} \frac{\theta_s}{1-\theta_s} \;\geq\; \tau\Big(1 + \max_{s\in S} \frac{\theta_s}{1-\theta_s}\Big). \tag{7}$$

The endpoint $\theta_s = 0$ *is read in the right limit.*

(ii) In the odds-ratio coordinates $r_s := \theta_s/(1-\theta_s) \in [0, \tau/(1-\tau)]$*, the implementable set (7) is the intersection of the* $|S|$ *halfspaces* $\{\sum_{s'\in S} r_{s'} \geq \tau(1+r_s)\}$*, one per* $s \in S$*. For any fixed ordering* $r_{(1)} \leq \cdots \leq r_{(|S|)}$*, the corresponding piece of the persuasion frontier is*

$$\sum_{i=1}^{|S|-1} r_{(i)} + (1-\tau)\, r_{(|S|)} = \tau. \tag{8}$$

Globally, the frontier is the union of the corresponding hyperplanes over all permutations of the signal labels.

In odds coordinates, the inequality has a direct interpretation. A lower persuasion threshold θ_s means stronger persuasion after realization s, and hence lower threshold odds $r_s = \theta_s/(1-\theta_s)$. The left-hand side aggregates the total threshold burden across realizations. The right-hand side is governed by the largest threshold odds, which corresponds to the tightest cross-realization selection constraint. Thus the menu can lower persuasion thresholds below the action threshold τ, but it cannot lower all of them too much at once.

Corollary 1 (Symmetric persuasion across realizations)
For $t \in [0, \tau]$, the symmetric profile $(t, \ldots, t)$ is implementable if and only if $t \geq \tau/|S|$. Consequently, narrative persuasion can symmetrically lower every persuasion threshold below the action threshold τ, but not below $\tau/|S|$.

Corollary 2 (Optimal ordering on the persuasion frontier)
Relabel the signal realizations as $s_1, \ldots, s_n$, $n = |S|$, so that $w_{s_1} \geq \cdots \geq w_{s_n}$. Every implementable threshold profile is pointwise weakly dominated in v by an implementable profile $(\theta'_{s_i})_{i=1}^n$ with

$$0 \leq \theta'_{s_1} \leq \cdots \leq \theta'_{s_n} \leq \tau$$

whose odds coordinates $r'_{s_i} := \theta'_{s_i}/(1-\theta'_{s_i})$ satisfy the frontier equation (8) under this ordering. Hence the sender's outer problem can be restricted to frontier profiles that assign weakly lower thresholds to weakly higher payoff weights.

The dominance in Corollary 2 has two steps. First, thresholds can be ordered so that weakly lower thresholds are assigned to weakly higher payoff weights. Since a lower threshold persuades weakly more types after that realization, this rearrangement weakly raises the induced payoff function. Second, scaling all threshold odds of the ordered profile down to the boundary in Theorem 2 weakly lowers every threshold, which again weakly raises the induced payoff function. The ordering step rests on a pairwise swap argument; the contraction to the frontier applies Theorem 2 together with the pointwise monotonicity of the payoff in thresholds.

The frontier represents a minimum threshold-burden condition. Under this payoff-weight ordering, lowering the threshold odds of a high-payoff realization requires assigning the remaining burden to lower-payoff realizations. This is the sense in which the sender's search can be restricted to frontier profiles with lower thresholds assigned to higher-payoff realizations.

Proposition 1 (Existence of the optimal menu)
There exists an optimal menu $M^* \in \mathcal{M}$ attaining $V(M^*) = \max_{M \in \mathcal{M}} V(M)$.

Given Proposition 1, I replace sup with max for the sender's benchmark problem and robust outer problem.

For the figures and examples, write $S = \{h, l\}$ and let (θ_h, θ_l) denote the threshold profile. Specializing Theorem 2 gives the two-dimensional implementability region. For any $\tau \in (0, 1)$, write $\alpha(\theta) := \theta(1-\tau)/((1-\theta)\tau)$. A threshold pair $(\theta_h, \theta_l) \in [0, \tau]^2$ is implementable if and only if

$$\alpha(\theta_h) + \alpha(\theta_l) \geq \frac{1 - 2\min\{\theta_h, \theta_l\}}{1 - \min\{\theta_h, \theta_l\}}. \tag{9}$$

Indeed, if $\theta_h \leq \theta_l$, then (7) becomes $r_h + (1-\tau)r_l \geq \tau$, which is equivalent to (9); the case $\theta_l \leq \theta_h$ is symmetric. Figure 1 illustrates the implementability region and its frontier.

Define the *implementability slacks*

$$\psi(\theta_h, \theta_l) := \alpha(\theta_h) + \alpha(\theta_l) - \frac{1 - 2\theta_h}{1 - \theta_h}, \qquad \phi(\theta_h, \theta_l) := \alpha(\theta_h) + \alpha(\theta_l) - \frac{1 - 2\theta_l}{1 - \theta_l}. \tag{10}$$

On $\theta_h \le \theta_l$, (9) is equivalent to $\psi \ge 0$; on $\theta_l \le \theta_h$, it is equivalent to $\phi \ge 0$. Thus

$$\{0 \le \theta_h \le \theta_l \le \tau : \psi \ge 0\}$$

is the *h-first implementability region*, and

$$\{0 \le \theta_l \le \theta_h \le \tau : \phi \ge 0\}$$

is the *l-first implementability region*. The binding portions $\psi = 0$ and $\phi = 0$ are the *h-first branch* and the *l-first branch* of the persuasion frontier. On these branches, write

$$\theta_l = \theta_l(\theta_h), \quad \theta_h \in [0, \tau/2], \qquad \theta_h = \theta_h(\theta_l), \quad \theta_l \in [0, \tau/2]. \tag{11}$$

For the figure, write $r_s := \theta_s/(1-\theta_s)$. Substituting $\alpha(\theta_s) = r_s(1-\tau)/\tau$ into $\psi = 0$ gives the h-first branch; the l-first branch is symmetric:

$$r_h + (1-\tau)\, r_l = \tau \quad (h\text{-first branch}), \qquad (1-\tau)\, r_h + r_l = \tau \quad (l\text{-first branch}). \tag{12}$$

Figure 1 plots the implementability regions in panel (a) and the linear threshold-burden frontiers in panel (b).

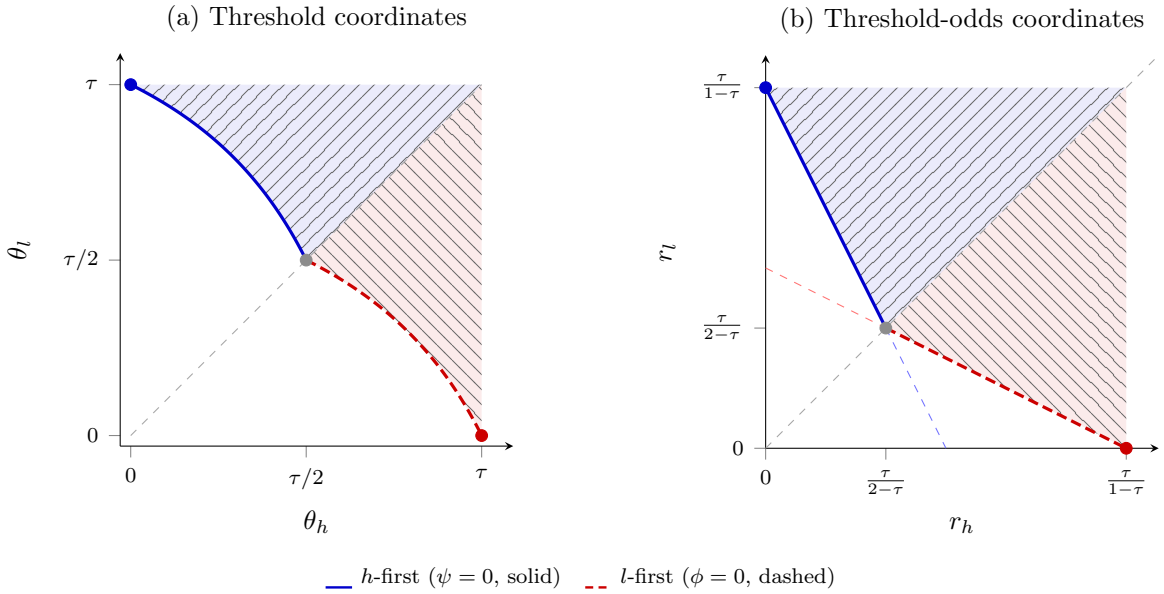


Figure 1: Implementability region and persuasion frontier restricted to $(\theta_h, \theta_l) \in [0, \tau]^2$ ($\tau = 1/2$). Panel (a) uses threshold coordinates (θ_h, θ_l); panel (b) uses threshold-odds coordinates (r_h, r_l), where $r_s = \theta_s/(1-\theta_s)$. Hatching distinguishes the h-first and l-first regions; solid and dashed curves mark their frontiers. Dots mark the junction and boundary menus.

3.2 Robustness

With a known population distribution, the sender chooses where to place the threshold burden on the frontier: the weights $(w_s)_{s \in S}$ determine the payoff from persuading receivers

after each realization, while the distribution determines the tail mass above each induced threshold. With ambiguity, this tail mass is no longer fixed. For any menu, the sender's payoff also depends on how much receiver mass Nature can keep at or below the induced positive thresholds, subject to the ambiguity set. Given any feasible threshold profile, Nature chooses a distribution that satisfies the moment restrictions and minimizes the sender's payoff. This is the inner problem. The sender evaluates each menu by its payoff guarantee, and then chooses a menu that maximizes this guarantee. This is the outer problem. As in Section 3.1, the outer choice can be restricted to frontier profiles that assign weakly lower thresholds to weakly higher payoff weights. Hence the analysis turns to the inner problem.

Payoff-guarantee equivalence. A receiver type exactly at a positive persuasion threshold cannot support a robust guarantee: under moment ambiguity, Nature can move that mass arbitrarily close to the threshold from below, where the receiver is no longer persuaded. Therefore, at a positive threshold, the payoff increment is paid only to types strictly above the threshold. At a zero threshold, the payoff increment is also paid to type $\theta = 0$, because $\Theta = [0, 1]$ gives Nature no strictly lower type from which to approach zero. This observation motivates evaluating the inner problem with the lower semi-continuous envelope of the per-type payoff:

$$\tilde{v}(\theta; (\theta_s)_{s\in S}) = \bar{u}_0 + \sum_{s\in S} w_s \Big(\mathbf{1}\{\theta > \theta_s\} + \mathbf{1}\{\theta_s = 0\}\mathbf{1}\{\theta = 0\} \Big), \qquad w_s > 0. \tag{13}$$

Receiver behavior remains as specified in Section 2.2; the envelope is used only to evaluate the inner problem and to obtain an attained worst-case distribution. For a menu M inducing thresholds $(\theta_s(M))_{s\in S}$, define

$$\widetilde{W}(M, F) := \int_\Theta \tilde{v}(\theta; (\theta_s(M))_{s\in S})\, F(d\theta), \qquad \widetilde{V}(M) := \inf_{F\in\Pi} \widetilde{W}(M, F).$$

The function $\tilde{v}$ is lower semi-continuous on Θ, so each menu has an attained payoff guarantee over Π. The next lemma shows that, under the richness of Π, this pointwise convention does not affect the payoff guarantee, the set of robust optimal menus, or the optimal value.

Lemma 1 (Payoff-guarantee equivalence)
Suppose $\Pi = \mathcal{F}(g, Y)$ *with* Y *star g-interior. For each menu* $M \in \mathcal{M}$,

$$V(M) = \inf_{F\in\Pi} W(M, F) = \inf_{F\in\Pi} \widetilde{W}(M, F) = \min_{F\in\Pi} \widetilde{W}(M, F).$$

In addition, $\arg\max_M V(M) = \arg\max_M \widetilde{V}(M)$.

Hence, for any menu M, the sender may replace $W(M, F)$ with $\widetilde{W}(M, F)$ before minimizing over Π; this leaves the payoff guarantee $\inf_{F\in\Pi} W(M, F)$ unchanged, ensures that the minimum is attained, and delivers the finite-support structure used in Proposition 2.

Corollary 3 (Robust frontier restriction)
Order the signal realizations as $s_1, \ldots, s_n$, $n = |S|$, *so that* $w_{s_1} \geq \cdots \geq w_{s_n}$. *Then*

$$\max_{M\in\mathcal{M}} V(M) = \max_{\substack{0\leq r_{s_1}\leq\cdots\leq r_{s_n}\leq \tau/(1-\tau) \\ \sum_{i=1}^{n-1} r_{s_i} + (1-\tau) r_{s_n} = \tau}} \min_{F\in\Pi} \int_0^1 \tilde{v}\big(\theta; (\theta(r_s))_{s\in S}\big)\, F(d\theta),$$

where $\theta(r) := r/(1+r)$.

Combining Proposition 1, Lemma 1, and Corollary 3, the robust problem becomes

$$\max_{M \in \mathcal{M}} \min_{F \in \Pi} \widetilde{W}(M, F),$$

with Nature's minimum attained and the sender's outer choice restricted to the persuasion frontier. For each threshold profile, the sender's payoff guarantee equals what she would earn against the least favorable $F \in \Pi$. As Π expands, the same menu must perform against a broader range of populations, so its payoff guarantee weakly decreases. It remains to characterize Nature's inner minimization problem for a fixed implementable threshold profile.

The inner problem. For each implementable threshold profile, the inner problem identifies the worst-case distribution in the ambiguity set and the associated payoff guarantee, as if Nature were placing receiver mass to minimize the sender's payoff subject to the moment restrictions. Under the maintained assumptions on Π, this problem admits a finite-dimensional dual characterization, paralleling the approach in Carrasco et al. (2018). Under $\tilde{v}$, at a positive persuasion threshold, only mass strictly above the threshold contributes to the sender's persuasion payoff; at a zero persuasion threshold, mass at type $\theta = 0$ also contributes. Thus, for a fixed threshold profile, the inner problem minimizes a weighted sum of upper-tail masses subject to the moment restrictions. The payoff guarantee depends on how those moment restrictions constrain Nature's ability to keep mass at or below the relevant persuasion thresholds. The finite-dimensional dual characterized below gives its value, and a worst-case distribution can be taken to have finite support contained in the contact set between $\tilde{v}(\cdot; (\theta_s)_{s \in S})$ and the dual minorant. Proposition A.4 records the corresponding zero-positive criterion: for a fixed threshold profile, the guarantee above $\bar{u}_0$ is strictly positive exactly when the moment restrictions rule out distributions supported weakly below the lowest positive threshold. This criterion is qualitative; the dual result below gives the level of the guarantee and the structure of least favorable distributions.

Proposition 2 (Worst-case distributions for an implementable threshold profile)
Suppose $\Pi = \mathcal{F}(g, Y)$*, where* $g : [0,1] \to \mathbb{R}^d$ *is continuous and* $Y \subseteq \operatorname{conv} g([0,1])$ *is nonempty, closed, convex, and star* g*-interior. Fix an implementable threshold profile* $(\theta_s)_{s \in S}$ *and evaluate it by the lower semi-continuous payoff* $\tilde{v}(\theta; (\theta_s)_{s \in S})$*. Then*

$$\min_{F \in \Pi} \int_0^1 \tilde{v}(\theta; (\theta_s)_{s \in S})\, F(d\theta) = \sup_{(\gamma, \lambda) \in \mathbb{R} \times \mathbb{R}^d} \Big\{ \gamma + \inf_{y \in Y} \lambda^\top y : \gamma + \lambda^\top g(\theta) \le \tilde{v}(\theta; (\theta_s)_{s \in S}) \text{ for every } \theta \in [0,1] \Big\}.$$

The minimum is attained, and at least one minimizer can be chosen with at most $d+1$ *support points. For every dual maximizer* (γ^*, λ^*)*, every worst-case distribution* F^* *satisfies* $\mathbb{E}_{F^*}[g(\theta)] \in \arg\min_{y \in Y} (\lambda^*)^\top y$ *and*

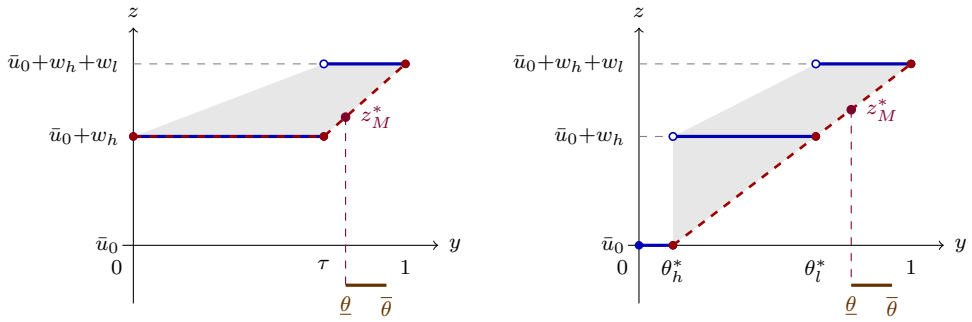
$$\operatorname{supp}(F^*) \subseteq \left\{ \theta \in [0,1] : \gamma^* + (\lambda^*)^\top g(\theta) = \tilde{v}(\theta; (\theta_s)_{s \in S}) \right\}.$$

If $g(\theta) = (\theta, \theta^2, \ldots, \theta^d)$*, then* $\gamma^* + (\lambda^*)^\top g(\theta)$ *is a polynomial of degree at most* d*.*

In Proposition 2, convexity and star g-interiority of Y give duality, and closedness gives attainment; the robustness result of Ball and Kattwinkel (2026) only requires star g-interiority. The dual problem chooses, from the affine span of the moment functions, a lower-bound function for the lower semi-continuous envelope of the sender's per-type payoff. The infinite-dimensional choice of a distribution is thereby reduced to a finite-dimensional choice of multipliers, i.e. the coefficients (γ, λ) of this affine minorant, with worst-case mass supported only on contact types. The contact set identifies the receiver types that are most relevant for a proposed menu. These locations are where Nature can place mass while meeting the observed moment restrictions and keeping the sender's payoff as low as possible. Depending on the moment restrictions and the chosen thresholds, this mass may lie far below the persuasion margin, at the margin, or above it when the restrictions require enough high-prior mass. Proposition 2 has an equivalent convexification interpretation. For d-moment restrictions, consider the closed convex hull of

$$\left\{\left(g(\theta), \tilde{v}(\theta; (\theta_s)_{s\in S})\right) : \theta \in [0,1]\right\} \subseteq \mathbb{R}^{d+1}.$$

Nature's value is obtained from the lower boundary of this set over moment vectors in Y. When $d = 1$, this lower boundary can be read from the figure. Figure 2 illustrates the case with $g(\theta) = \theta$ and $Y = [\underline{\theta}, \overline{\theta}]$. The dual constraint requires $\gamma + \lambda^\top g(\theta)$ to lie



(a) boundary pair $(0, \tau)$. (b) frontier interior pair.

Figure 2: Convexification geometry for $g(\theta) = \theta$, $Y = [\underline{\theta}, \overline{\theta}]$, and the h-first frontier. The shaded region is the closed convex hull of the graph of $\tilde{v}(\cdot; \theta_h, \theta_l)$; the dashed red curve is its lower boundary over Y. Panel (a) shows the boundary threshold pair $(0, \tau)$, and panel (b) shows a frontier interior threshold pair.

below the lower semi-continuous per-type payoff $\tilde{v}(\cdot; (\theta_s)_{s\in S})$.[13] Each positive threshold is thus priced at the value approached from the left, while a zero threshold includes type 0 in the persuaded set. Figure 3 illustrates the resulting polynomial minorant under power moments for $d = 3$. The contact-set logic is related to Carrasco et al. (2018, Lemma 2 and Proposition 3): with the first N moments of buyer valuations, the optimal transfer is the non-negative monotonic hull of a degree-N polynomial with coefficients $(\lambda_0, \ldots, \lambda_N)$, and Nature is supported on the contact set where transfer and polynomial coincide. Suzdaltsev

[13] Because $\tilde{v} \leq v$ pointwise, any affine minorant of $\tilde{v}$ also lies below the original per-type payoff v; the two functions differ only at positive persuasion thresholds.

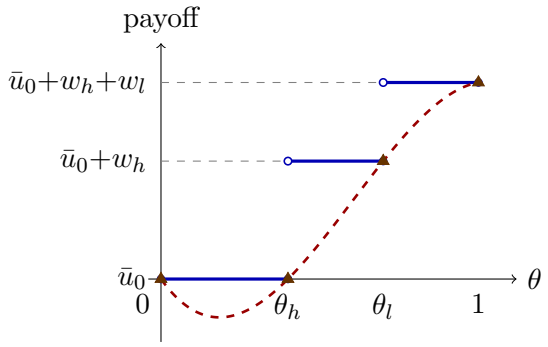


Figure 3: Dual lower bound under power moments. The dashed red polynomial $\gamma + \lambda^\top g(\theta)$ lies below the lower semi-continuous per-type payoff $\tilde{v}(\theta; \theta_h, \theta_l)$ (solid blue), and the worst-case mass concentrates at contact points (orange triangles). One worst-case distribution can be chosen with at most $d + 1$ atoms.

(2022), by contrast, faces a nonconvex ambiguity set due to the independence constraint on joint value distributions and uses an indirect argument. Finally, Ball and Kattwinkel (2026, Proposition 4) show that a finite-support saddle prior with strictly positive payoff is non-robust in a profit-maximization problem: because the mechanism is tailored to the saddle prior's atoms, small perturbations can lower payoff discontinuously. Here finite support has a different role: it is an extremal representation of the moment problem, and robustness comes from the richness of the continuous moment ambiguity set.

When robust solutions select benchmark optima. In the benchmark problem, every menu is evaluated under the same fixed distribution; in the robust problem, each menu is evaluated under its own least favorable distribution. Because a worst-case distribution can be taken to have finite support (Proposition 2), the benchmark problem evaluated at it can be under-determined: if a finite support leaves intervals of types with no mass, moving thresholds inside those intervals need not change the benchmark payoff. One might then hope that robustness acts as a tie-breaker, singling out the robust-optimal menu from within this set at its own worst-case distribution. The following example shows that this hope can fail in a strong sense: the robust-optimal menu need not be benchmark-optimal at its own worst-case distribution at all.

Example 4 (Failure of benchmark optimality)
Take $\tau = 1/2$, $\bar{u}_0 = 0$, $w_h = 2$, $w_l = 1$, *and* $\Pi = \{F \in \Delta([0,1]) : \mathbb{E}_F[\theta] = 0.4\}$. *On the h-first frontier,* $\theta_l(x) = (1-3x)/(2-4x)$ *for* $x \in [0, 1/4]$. *The boundary menu* $M_{h\text{-bd}}$ *has payoff guarantee* 2. *For any* $x \in (0, 1/4]$, *Nature can put mass* $(1-0.4)/(1-x)$ *at* x *and mass* $(0.4-x)/(1-x)$ *at* 1, *giving the frontier menu* $(x, \theta_l(x))$ *payoff at most* $3(0.4-x)/(1-x) < 2$. *Hence* $M_{h\text{-bd}}$ *is robust optimal. At* $F^* = \delta_{0.4}$, *however, any h-first frontier point* $(x, \theta_l(x))$ *with* $x \in (1/7, 1/4]$ *satisfies* $\theta_l(x) < 0.4$, *so type* 0.4 *is persuaded after both realizations and yields benchmark payoff* $w_h + w_l = 3$, *while the boundary menu* $(0, 1/2)$ *yields only* $w_h = 2$. *Hence* $M_{h\text{-}bd}$ *is not benchmark-optimal at* F^*.

The robust-optimal menu $M_{h\text{-bd}}$ is not benchmark-optimal at its own worst-case distribution $\delta_{0.4}$: it lies outside the benchmark solution set there, so robustness has not selected a

benchmark optimum. At $\delta_{0.4}$, that benchmark solution set is the continuum of h-first frontier menus that persuade type 0.4 after both realizations. These menus deliver payoff guarantees below that of $M_{h\text{-bd}}$, hence are not robust-optimal; consequently, the two solution sets are not nested either way. What robustness delivers is conditional: when the selected menu and the selected worst-case distribution are mutual best responses, a saddle point, the robust optimum is a common element of the two sets, one point of a benchmark solution set that may be a continuum. The next proposition records this, and Proposition A.5 in Appendix A.12 gives criteria for checking the saddle condition.

Proposition 3 (Saddle-point selection from benchmark optima)
If $(M^*, F^*) \in \mathcal{M} \times \Pi$ *is a saddle point of* $\widetilde{W}$*, then* $M^* \in \arg\max_{M\in\mathcal{M}} \widetilde{W}(M, F^*)$*,* $M^* \in \arg\max_{M\in\mathcal{M}} \min_{F\in\Pi} \widetilde{W}(M, F)$*, and* $M^* \in \arg\max_{M\in\mathcal{M}} \inf_{F\in\Pi} W(M, F)$*. Robustness selects an element of the* $\widetilde{W}$*-benchmark solution set at* F^**.*

Proposition 3 implies that determining whether a robust solution is also benchmark-optimal at its worst-case distribution requires checking for profitable deviations. This requires that no shift along the relevant persuasion frontier improves the sender's payoff.

3.3 Discussion

Menu-size bound. When priors are scalar and action is cutoff-based, each realization has a single payoff-relevant marginal type. Retaining one threshold-setting narrative for each realization is enough. A menu with more narratives than signal realizations can always be replaced by an optimal sub-menu with at most as many narratives as signal realizations. Theorem 1 gives this $|S|$-narrative bound. Richer narratives still affect how receivers explain a signal realization and which types select which narrative, but not in a strictly payoff-improving way.

This differs from the single-receiver size argument in Aina (2025), despite the same $|S|$ bound. There, a single receiver's prior is fixed, so each signal realization can be assigned a narrative that, for that prior, induces the target posterior and attains the maximal feasible fit conditional on that realization. With heterogeneous receivers, maximal fit is type-dependent: a narrative that maximizes fit for one prior need not maximize fit for another. The sender must control fit-based selection over the whole prior support. The same $|S|$ bound survives because the binary-state environment makes priors scalar and action cutoff-based: controlling selection over priors reduces to retaining one threshold-setting narrative per realization.

One might suspect that the finite-support representation of the worst-case distribution yields another menu-size bound: if a worst-case distribution can be chosen with at most $d+1$ support points, perhaps only $d+1$ receiver types need be controlled, and hence $d+1$ narratives suffice. This inference is not valid. The finite-support result is conditional on a fixed menu, or equivalently on a fixed threshold profile. Its support points are contact types of that menu's inner moment problem, not a menu-independent set of marginal receivers. As the menu changes, the least favorable distribution and its contact set can change as well. Without a common finite contact set that works for all feasible menus and deviations, a support-size bound for the worst-case distribution does not translate into a menu-size bound for narratives.

Beyond binary states. The binary-state assumption makes the receiver's posterior one-dimensional and, together with a monotone comparison structure in receiver behavior, yields a threshold-choice representation. With more than two states, the same one-dimensional geometry is recovered when behavior is summarized by a scalar posterior statistic. One case is a binary partition that is sufficient for receiver behavior: let H denote the event where the sender-preferred action is also receiver-preferred, let L denote its complement, and denote the action by 1. For instance, if the receiver's payoff from purchase depends only on whether product quality exceeds an adequacy cutoff q^*, the partition $H = \{q \geq q^*\}$ and $L = \{q < q^*\}$ is behaviorally sufficient whenever narrative selection and action choice factor through the induced binary posterior $(\nu(H), \nu(L))$, equivalently through the scalar $\nu(H)$.

Another case is a scalar posterior statistic. If the receiver takes action 1 whenever the posterior mean of a real-valued state exceeds a threshold, then the posterior mean plays the same role as the belief on H. In Kamenica and Gentzkow (2011, Proposition 3), sender payoffs depend on the expected state; their Proposition 6 extends to payoffs that depend only on a finite-dimensional linear statistic of posterior beliefs. Kolotilin et al. (2017) impose linear-in-state utilities, so posterior beliefs are summarized by their mean and messages can be indexed by cutoff types. Related threshold and cutoff representations also appear in Rayo and Segal (2010), Vohra et al. (2023) and Kolotilin and Zapechelnyuk (2025).

However, this condition is non-trivial. If residual distinctions change receiver action or narrative selection, the scalar index no longer suffices for receiver behavior. If the sender's ordinal comparison of menus is responsive to replacing the primitive problem by scalar thresholds, the scalar model will no longer be behavior-equivalent for sender menu choice. The cardinal results in Section 3.2 require more. Appendix A.7 gives the additional conditions under which the primitive problem collapses to thresholds.

Narrative persuasion versus Bayesian persuasion. In the binary-state setting with cutoff action studied here, both narrative persuasion and Bayesian persuasion organize persuaded types into intervals. Using the binary-signal case as an example, on the h-first branch, a boundary menu induces two intervals: types persuaded after both realizations and types persuaded only after h; a frontier-interior menu induces three intervals: types persuaded after both realizations, types persuaded only after h, and types persuaded after neither realization. The key difference is the source of the cross-realization restriction. In Bayesian persuasion, Bayes plausibility makes belief movements balance across realizations: a receiver's posterior cannot lie strictly above his prior after every realization. Narrative persuasion does not impose this balance beyond fit competition. Because receivers may select different narratives after different realizations, narrative persuasion can induce receivers with some priors to choose the sender-favorable action after all realizations.

However, this does not imply that narrative persuasion gives the sender a higher value than Bayesian persuasion, since the data-generating model used to compute the sender's expected payoff is fixed in narrative persuasion but chosen in Bayesian persuasion. Receiver welfare also has no clear ranking. For a receiver with prior θ, additional persuasion changes welfare by $\theta u_H \sum_s \pi(s \mid H)\Delta a_s(\theta) - (1-\theta)|u_L| \sum_s \pi(s \mid L)\Delta a_s(\theta)$. A given receiver gains when the prior-weighted gain in state H exceeds the prior-weighted loss in state L; if the inequality is reversed, he loses. Because this comparison varies with θ, additional persuasion

can also raise welfare for some receivers and lower it for others.

4 Application

Consider a dictator who wants citizens to support her ($a = 1$), while citizens only want to support her when she is competent ($\omega = H$). Citizens differ in their prior θ that the dictator is competent. Public news generates a favorable realization h and an unfavorable realization l, with $0 < \pi_L < \pi_H < 1$, so each realization is informative but imperfect. The action threshold is $\tau = 1/2$, as in much of the persuasion literature: a receiver takes the sender-preferred action whenever the state aligned with that action is at least as likely as the alternative (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Kamenica, 2019).

I normalize the dictator's utility when citizens choose $a = 0$, denoted by $\bar{u}_0$, to zero. This normalization is without loss because $\bar{u}_0$ enters as a constant term independent of the persuasion thresholds and the cross-sectional distribution. The parameters w_h and w_l measure her expected benefit from securing support after the favorable and unfavorable realizations, respectively: w_h is larger when the favorable realization can be broadcast to a wide audience and converted into visible support, while w_l is larger when suppressing dissent after the unfavorable realization prevents coordination, elite defection, or costly opposition.

The dictator observes citizens' priors only through coarse summaries. She knows the average prior $\hat{\theta}$ and a dispersion lower bound $\underline{\sigma}^2$. The mean is inferred from aggregate behavior, such as turnout at regime rallies, participation in street protests, or compliance with local directives; the dispersion bound comes from evidence of polarization, such as disagreement in social-media posts.[14] As preference falsification is pervasive, she seeks a payoff guarantee robust to small perturbations of these parameters.

Let $0 < \hat{\theta} < 1$ and $\underline{\sigma}^2 \in (0, \hat{\theta}(1-\hat{\theta}))$, and define the ambiguity set

$$\Pi(\hat{\theta}, \underline{\sigma}^2) := \left\{ F \in \Delta([0,1]) : \mathbb{E}_F[\theta] = \hat{\theta},\ \mathrm{Var}_F(\theta) \geq \underline{\sigma}^2 \right\}.$$

Because the dispersion lower bound rules out degenerate distributions, Nature cannot satisfy the moment restrictions by placing all citizens at the same belief. The inner problem is governed by the way the mean equality and the variance lower bound restrict Nature's ability to place citizens at or below the persuasion threshold. This threshold geometry generates the central comparative static: a higher average prior $\hat{\theta}$ need not improve the dictator's optimal value. Proposition 7 characterizes when this happens and why.

Boundary menu. In threshold coordinates, write $M_{h\text{-bd}}$ for the boundary menu with threshold pair $(0, 1/2)$ and $M_{l\text{-bd}}$ for the boundary menu with threshold pair $(1/2, 0)$. Under $M_{h\text{-bd}}$, every citizen contributes w_h, and citizens with $\theta > 1/2$ contribute an additional w_l; under $M_{l\text{-bd}}$, the roles of w_h and w_l are reversed. Equivalently, the better boundary menu earns $\max\{w_h, w_l\}$ from every citizen and an additional $\min\{w_h, w_l\}$ from citizens with $\theta > 1/2$.

The two boundary menus, $(0, 1/2)$ and $(1/2, 0)$, are symmetric. Each pairs a narrative that makes one signal realization decisive evidence of strength with a narrative that makes

[14] The lower-bound form captures a minimum degree of dispersion in citizens' priors. By contrast, an upper bound on variance would remain consistent with a nearly homogeneous population.

the other uninformative. Thus one realization is fully insured, while the other is protected only by the moment-forced mass above 1/2.

In the underlying narrative parameters, $(0,0)$ makes the unfavorable realization l uninformative, while $(1,1)$ makes the favorable realization h uninformative. The narrative $(1,0)$ says that favorable news reveals strength, as when policy success reflects the dictator's competence; $(0,1)$ says that unfavorable news reveals strength, as when enemies sabotage the dictator because she is competent.

Moment-forced tail mass under a boundary menu. The key object is the moment-forced tail mass above a persuasion threshold t. Write $\underline{T}(t)$ for the minimum, over all distributions satisfying the mean equality and variance lower bound, of the share of citizens with $\theta > t$. This is the least share of citizens the dictator can guarantee above the threshold after the relevant realization; its complement is the largest share Nature can keep at or below the margin. This threshold calculation is the specialization of Proposition 2 to $g(\theta) = (\theta, \theta^2)$ and payoff $\mathbf{1}\{\theta > t\}$: the dual is a quadratic minorant with contact in $\{0, t, 1\}$, and the worst case has two or three support points there.

Proposition 4 (Moment-forced tail mass under a boundary menu)
(i) Minimum tail mass. *For every $\hat{\theta} \in (0,1)$, $\underline{\sigma}^2 \in (0, \hat{\theta}(1-\hat{\theta}))$, and threshold $t \in (0,1)$, write*

$$\underline{T}(t) := \inf_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} F\big((t,1]\big)$$

for the minimum mass above t over the ambiguity set. Then

$$\underline{T}(t) = \max\left\{ \frac{\max\{\hat{\theta} - t,\, 0\}}{1-t},\ \frac{\underline{\sigma}^2 - \hat{\theta}(t - \hat{\theta})}{1-t},\ 0 \right\}. \tag{14}$$

The infimum is attained by a two- or three-point distribution on $\{0, t, 1\}$, matching the finite-support contact pattern in Proposition 2.
(ii) Boundary-menu value at $\tau = 1/2$. *Applying part* (i) *at $t = 1/2$, the payoff guarantees of the two boundary menus are*

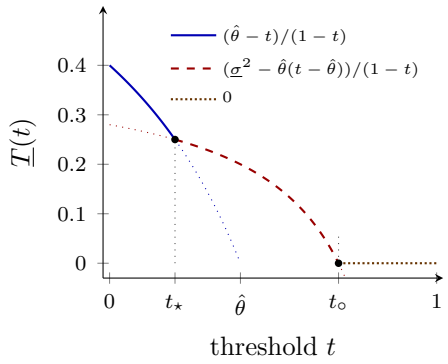
$$\min_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} \int \tilde{v}_{(0,1/2)}\, dF = w_h + w_l \cdot \underline{T}(1/2), \qquad \min_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} \int \tilde{v}_{(1/2,0)}\, dF = w_l + w_h \cdot \underline{T}(1/2).$$

When $w_h \geq w_l$, the menu $(0, 1/2)$ weakly dominates and the relevant worst-case boundary value is $w_h + w_l\, \underline{T}(1/2)$; when $w_h < w_l$, $(1/2, 0)$ dominates with value $w_l + w_h\, \underline{T}(1/2)$. The dominance is strict whenever $w_h \neq w_l$, since $\underline{T}(1/2) < 1$.

Nature places mass at the lowest prior, at the persuasion threshold, and at the highest prior: mass at t absorbs moment requirements without entering the upper tail $F((t,1])$, while mass at 1 appears only when the mean or variance bound forces mass above the threshold.

The two breakpoints at which the active piece changes are

$$t_\star := \hat{\theta} - \frac{\underline{\sigma}^2}{1-\hat{\theta}}, \qquad t_\circ := \hat{\theta} + \frac{\underline{\sigma}^2}{\hat{\theta}};$$



Worst-case tail mass $\underline{T}(t)$ at $\hat{\theta} = 0.4$, $\underline{\sigma}^2 = 0.12$

Figure 4: Worst-case tail mass $\underline{T}(t)$ for $\hat{\theta} = 0.4$ and $\underline{\sigma}^2 = 0.12$. The boundary application later evaluates this curve at $t = \tau = 1/2$.

the feasibility constraint $\underline{\sigma}^2 \in (0, \hat{\theta}(1 - \hat{\theta}))$ ensures $0 < t_\star < \hat{\theta} < t_\circ < 1$.

Three parameter regions determine which piece is active (Figure 4): the tail mass is zero, mean-forced, or variance-forced. See Appendix B.1 for details.

Proposition 5 (Boundary optimality)
For every $w_h > 0$, $w_l > 0$, every $\hat{\theta} \in (0, 1)$, and every $\underline{\sigma}^2 \in (0, \hat{\theta}(1 - \hat{\theta}))$, a boundary menu is optimal. If $w_h \geq w_l$, the optimal boundary menu is $(0, 1/2)$ with value $w_h + w_l \underline{T}(1/2)$; if $w_h < w_l$, it is $(1/2, 0)$ with value $w_l + w_h \underline{T}(1/2)$. In the knife-edge case $w_h = w_l$, both boundary menus are optimal.

Thus, in the natural case $\tau = 1/2$, the optimal menu is a boundary menu. The realization with the larger payoff weight receives the zero threshold: h if $w_h \geq w_l$, and l otherwise.

The case $w_l > w_h$ is relevant for crisis-prone authoritarian politics. A larger w_l captures situations where securing support after an unfavorable realization is more valuable: unfavorable news can threaten the regime by revealing incompetence, enabling citizens to infer regime vulnerability, and encouraging elite defection. Literature on authoritarian survival emphasizes these asymmetric dangers: information manipulation is valuable because it can forestall coordinated uprisings (Edmond, 2013); censorship tends to target collective-action potential rather than criticism as such (King et al., 2013); preference falsification can generate sudden protest cascades once adverse events become common knowledge (Kuran, 1991); and modern "informational autocrats" survive by sustaining a competence image that unfavorable news threatens (Guriev and Treisman, 2019; Rozenas and Stukal, 2019). Proposition 5 then selects the boundary menu $(1/2, 0)$. This corresponds to, for instance, a common authoritarian narrative in which protests, unrest, or policy failures are framed as sabotage by domestic opponents backed by foreign enemies. It resonates with the alternative-reality account of propaganda in Szeidl and Szucs (2025), in which criticism or adverse evidence validates a populist narrative where attacks from hostile elites serve as proof that the politician is effectively challenging the corrupt establishment on behalf of the people.

A tighter variance lower bound shrinks $\Pi(\hat{\theta}, \underline{\sigma}^2)$ and restricts Nature's ability to remove citizens from persuaded tails. This yields the following monotonicity result.

Proposition 6 (Monotonicity in the variance lower bound)
(i) Weak monotonicity. *For any dictator primitives* $(w_h, w_l, \hat{\theta})$ *and any two feasible variance lower bounds* $0 < \underline{\sigma}_1^2 \leq \underline{\sigma}_2^2 < \hat{\theta}(1-\hat{\theta})$*, the payoff guarantee of every menu is weakly higher under the tighter constraint:*

$$\min_{F \in \Pi(\hat{\theta}, \underline{\sigma}_1^2)} \int \tilde{v}\, dF \;\leq\; \min_{F \in \Pi(\hat{\theta}, \underline{\sigma}_2^2)} \int \tilde{v}\, dF \quad \forall\, M.$$

Consequently, the optimal value $\max_M \min_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} \int \tilde{v}\, dF$ *is weakly increasing in* $\underline{\sigma}^2$.
(ii) Strict monotonicity. *On the variance-binding region for* $\underline{T}(1/2)$*, the optimal value satisfies*

$$\max_M \min_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} \int \tilde{v}\, dF \;=\; \max\{w_h, w_l\} + 2\min\{w_h, w_l\}\big(\underline{\sigma}^2 - \hat{\theta}(1/2 - \hat{\theta})\big),$$

and on each connected component of this region the optimal value is strictly increasing in $\underline{\sigma}^2$ *with derivative* $2\min\{w_h, w_l\} > 0$.

Greater observed dispersion among citizens' priors can raise the dictator's optimal value even though she seeks uniform political support: evidence of high enough polarization can rule out the least favorable population in which receiver mass is concentrated at the persuasion threshold and contributes no upper-tail payoff.

In Gitmez and Molavi (2022)'s Bayesian persuasion setting, dispersion is a property of the known population: a more diverse society makes public persuasion harder because the sender must appeal to more skeptical receivers without losing support elsewhere. Here, dispersion enters as a restriction on Nature's feasible set: a higher lower bound on prior dispersion shrinks $\Pi(\hat{\theta}, \underline{\sigma}^2)$. By Proposition 6, this weakly raises the optimal value.

The mean restriction has a subtler effect. It can raise the average prior by moving mass from lower-prior types toward the persuasion threshold, yet this shift can lower the worst-case mass strictly above the threshold.

Proposition 7 (Non-monotonicity in $\hat{\theta}$)
Fix $w_h > 0$*,* $w_l > 0$*, and* $\underline{\sigma}^2 \in (0, 1/4)$*, and vary* $\hat{\theta}$ *over* $\{\hat{\theta} \in (0,1) : \hat{\theta}(1-\hat{\theta}) > \underline{\sigma}^2\}$*. At* $\tau = 1/2$*, the optimal value is*

$$\max_M \min_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} \int \tilde{v}\, dF = \max\{w_h, w_l\} + \min\{w_h, w_l\} \max\{2\max\{\hat{\theta} - 1/2, 0\},\; 2(\underline{\sigma}^2 - \hat{\theta}(1/2 - \hat{\theta})),\; 0\}.$$

It fails to be monotone in $\hat{\theta}$ *if and only if* $\underline{\sigma}^2 < 3/16$*. Its minimum equals* $\max\{w_h, w_l\}$ *whenever* $\underline{\sigma}^2 \leq 1/16$*, and equals* $\max\{w_h, w_l\} + \min\{w_h, w_l\}(2\underline{\sigma}^2 - 1/8)$ *at* $\hat{\theta} = 1/4$ *whenever* $1/16 < \underline{\sigma}^2 < 3/16$*. When* $\underline{\sigma}^2 \geq 3/16$*, the optimal value is strictly increasing in* $\hat{\theta}$.

In the variance-binding region, Nature can raise the average prior by moving mass to $1/2$, where citizens are more favorable but still outside the strict upper tail. This lowers the worst-case upper-tail mass. Here $2(\underline{\sigma}^2 - \hat{\theta}(1/2 - \hat{\theta}))$ determines $\underline{T}(1/2)$; once the slack below the threshold is exhausted, further increases in $\hat{\theta}$ require more mass above $1/2$. See Figure 5.

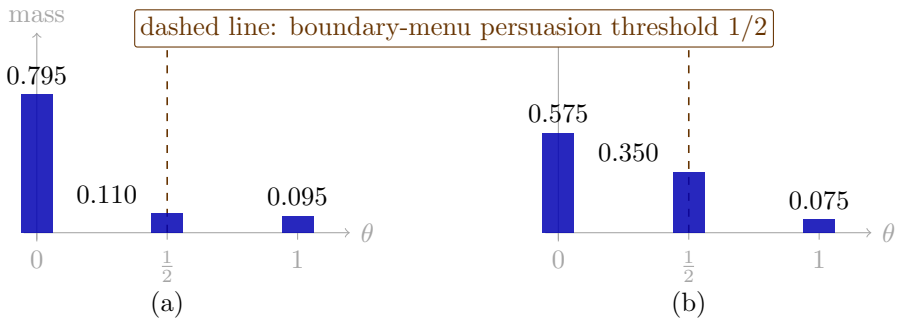


Figure 5: Worst-case distributions with $\underline{\sigma}^2 = 0.10$ and $\tau = 1/2$. Raising the mean from $\hat{\theta} = 0.15$ to $\hat{\theta} = 0.25$ lowers worst-case mass above $1/2$ from 9.5% to 7.5%.

This nonmonotonic pattern is possible only when the variance lower bound is not too high. With a high lower bound, Nature must keep enough mass far from the threshold throughout the feasible range of $\hat{\theta}$; increasing $\hat{\theta}$ then requires more mass on the high-prior side, and the optimal value is increasing. Different restrictions determine worst-case tail mass in different parameter regions: within the variance-binding region, a higher mean can lower the optimal value; within the mean-binding region, it uniformly helps.

By affecting the dictator's optimal value through the minimum tail mass above the persuasion threshold, ambiguity can also reshape the cross-realization allocation of persuasion.

Robust concentration versus benchmark spread. The boundary-menu result means that ambiguity leads the dictator to concentrate persuasion on the realization with the larger payoff weight, relying on the moment restrictions to force some citizens above the unfavorable-realization threshold. Whether this allocation differs from the benchmark problem at the worst-case distribution F^* depends on a saddle-point condition in Appendix B.5. When it holds, $M_{h\text{-bd}}$ is benchmark-optimal at F^* if $w_h \geq w_l$, and $M_{l\text{-bd}}$ is benchmark-optimal at F^* if $w_l > w_h$. When it fails, a benchmark optimum at F^* can be chosen in the frontier interior, and ambiguity then changes both the optimal value and the allocation of persuasion across realizations.

Beyond $\tau = 1/2$. The same threshold geometry extends to arbitrary action thresholds. Formula (14) applies to any $\tau \in (0, 1)$, and a boundary menu with threshold τ has value

$$\max\{w_h, w_l\} + \min\{w_h, w_l\}\, \underline{T}(\tau)$$

whenever that boundary menu is optimal. On the variance-binding piece,

$$\underline{T}(\tau) = \frac{\underline{\sigma}^2 - \hat{\theta}(\tau - \hat{\theta})}{1 - \tau},$$

so the boundary value has $\hat{\theta}$-derivative

$$\frac{\min\{w_h, w_l\}(2\hat{\theta} - \tau)}{1 - \tau}.$$

Thus the point at which the mean effect switches sign is $\hat{\theta} = \tau/2$; the turning point 1/4 in Proposition 7 is the specialization to $\tau = 1/2$.

This negative mean effect is not merely a boundary-payoff calculation. If $\hat{\theta} < \tau/2$, then $\hat{\theta} < 1/2$, so the primitive boundary-optimality condition in Appendix B.6 is automatically satisfied for the boundary menu associated with the larger payoff weight. Hence, whenever the variance piece is active below $\tau/2$, raising the average prior lowers the actual robust value. For a fixed variance lower bound, this decreasing low-mean interval is nonempty exactly when

$$\underline{\sigma}^2 < \frac{\tau(2-\tau)}{4}.$$

Outside the boundary-optimal regions, however, frontier-interior menus can dominate; the exact general-τ value then requires a case-by-case comparison along the persuasion frontier. The main text therefore keeps the symmetric threshold $\tau = 1/2$, where boundary optimality is global and the comparative statics are sharp.

5 Conclusion

Robust narrative persuasion broadens the scope of persuasion. The signal space and the data-generating model may be outside the sender's control, governed by forces such as physical processes, other actors, or institutional rules. Information design is then unavailable. When interpretation remains malleable, however, narratives still allow the sender to pursue robust persuasion. The sender does not know the full distribution of priors, so she cannot solve a distribution-specific narrative persuasion problem. When she observes only a point estimate or confidence region for moments of that distribution, the sender can instead formulate the narrative persuasion problem over the associated ambiguity set. Moreover, under the paper's formulation, small errors in the moment information used to construct the ambiguity set do not make the resulting payoff guarantee drop discontinuously.

The analysis yields restrictions on behavior across signal realizations. If receivers take the sender-preferred action whenever they regard its aligned state as more likely, the optimal menu frames the realization with the higher sender payoff weight as diagnostic of the sender-preferred state. Moreover, for any strictly positive action threshold, no menu can make every receiver take the sender-preferred action after both signal realizations; if nearly all receivers do so after all events, forces outside narrative selection are likely at work. Each narrative covers the entire signal space, so adopting it shifts a receiver's belief after every realization; because public menus are chosen before the realization is known, persuading more receivers after one event comes at the cost of persuading fewer after the other. Without such cross-event consistency, ex post explanations can be more flexible and event-specific. The optimal menu also uses a limited narrative family. In persistent campaigns, similar narratives, for instance opponent sabotage, explain similar events. These restrictions illustrate the behavioral implications of robust narrative persuasion.

A Proofs for the results

A.1 Preliminaries

Definition A.1 (Lower semi-continuous envelope)
For a bounded function $f\colon \Theta \to \mathbb{R}$, the lower semi-continuous envelope of f is

$$\operatorname{lsc}(f)(\theta) \;:=\; \liminf_{\theta' \to \theta} f(\theta'), \qquad \theta \in \Theta.$$

Equivalently, $\operatorname{lsc}(f)$ is the largest lower semi-continuous function that is pointwise no greater than f.

Lemma A.1 (Properties of narrative selection)
Fix a realization $s \in S$. For any narrative m, write

$$\eta_s(m) := m(s \mid H) - m(s \mid L).$$

Take two distinct narratives m_i and m_j. If $\eta_s(m_i) = \eta_s(m_j)$, then their fit difference after s is constant in θ. If $\eta_s(m_i) < \eta_s(m_j)$, then

$$\operatorname{Fit}(s, m_j, \theta) - \operatorname{Fit}(s, m_i, \theta) = \big(m_j(s \mid L) - m_i(s \mid L)\big) + \big(\eta_s(m_j) - \eta_s(m_i)\big)\theta$$

is strictly increasing in θ. Hence any two narratives cross at most once after s. Along the upper envelope of fit functions after s, the selected value of $\eta_s(m)$ is non-decreasing in type.

Proof of Lemma A.1. Direct subtraction from

$$\operatorname{Fit}(s, m, \theta) = m(s \mid L) + \eta_s(m)\theta$$

gives the displayed difference. Equal slopes make the difference constant. If $\eta_s(m_i) < \eta_s(m_j)$, the difference has strictly positive slope, so the pair crosses at most once and only from m_i to m_j as θ increases. Applying this pairwise crossing property along the upper envelope gives monotone selected slopes. □

Corollary A.1 (Monotone posteriors and action profiles)
For any menu $M \in \mathcal{M}$, the posterior $\hat{\mu}_s(\theta) := \mu_s^{m^*(s,\theta)}(\theta)$ is non-decreasing in θ for each $s \in S$, under the zero-fit completion in Definition A.2. Consequently, the induced action $a_\theta(s)$ is non-decreasing in θ for each $s \in S$.

Proof of Corollary A.1. Fix $s \in S$. If every narrative in M is s-null in the sense of Definition A.2, then the completed posterior after s is constant and the result is immediate. Otherwise the maximum fit after s is positive on $(0, 1)$.

For a fixed narrative m, the posterior

$$\mu_s^m(\theta) = \frac{\theta m(s \mid H)}{(1-\theta)m(s \mid L) + \theta m(s \mid H)}$$

is weakly increasing in θ wherever the denominator is positive. Between switching points the selected narrative is fixed, so the selected posterior is weakly increasing there.

At an interior switching point from m_a to m_b as θ increases, Lemma A.1 gives $\eta_s(m_b) > \eta_s(m_a)$. Fit equality at θ^* gives

$$m_b(s \mid L) - m_a(s \mid L) = -\big(\eta_s(m_b) - \eta_s(m_a)\big)\theta^*$$

and hence

$$m_b(s \mid H) - m_a(s \mid H) = \big(\eta_s(m_b) - \eta_s(m_a)\big)(1 - \theta^*) \geq 0.$$

Since the two narratives have the same positive fit at θ^*, the posterior under m_b is weakly higher than the posterior under m_a at the switch, with a strict increase whenever $\theta^* \in (0, 1)$ and the slopes differ. Sender-favorable tie-breaking selects the upward value at switches. Boundary zero-fit cases are completed by one-sided limits in Definition A.2. Therefore $\hat{\mu}_s$ is non-decreasing, and $a_\theta(s) = \mathbf{1}\{\hat{\mu}_s(\theta) \geq \tau\}$ is non-decreasing as well. □

Lemma A.2 (Threshold formula)
Fix $s \in S$. If narrative m is adopted after s, whenever the posterior is well-defined,

$$\mu_s^m(\theta) \geq \tau \quad \textit{iff} \quad \theta \geq \frac{\tau m(s \mid L)}{(1-\tau)m(s \mid H) + \tau m(s \mid L)}.$$

This threshold is continuous in the likelihood pair $(m(s \mid H), m(s \mid L))$ wherever the denominator is positive.

Proof of Lemma A.2. The inequality

$$\frac{\theta m(s \mid H)}{\theta m(s \mid H) + (1-\theta)m(s \mid L)} \geq \tau$$

rearranges to the displayed threshold formula whenever the denominator is positive. Continuity is immediate on that domain. □

Definition A.2 (Zero-fit completion)
Fix $s \in S$. A narrative is s-null if $m(s \mid H) = m(s \mid L) = 0$. If every narrative in a menu is s-null, then the realization has zero fit for every type under every offered narrative; I complete the posterior after s by $\hat{\mu}_s(\theta) = 0$ for all θ, so the induced persuasion threshold is $\theta_s = 1$.

If the menu contains at least one narrative that is not s-null, then the maximum fit after s is positive for every interior type. A zero-fit selected narrative can arise only at a boundary belief $\theta \in \{0, 1\}$. At such a boundary, nearby interior types have positive fit under the selected upper-envelope segment, and Bayes' rule has a one-sided limit along those types. I extend (2) to the boundary belief by that limit. This completion preserves Lemmas A.1 and A.2 and Corollary A.1 and leaves the threshold payoff representation unchanged.

Lemma A.3 (Pointwise threshold monotonicity)
For any two threshold profiles $(\theta_s)_{s\in S}, (\theta'_s)_{s\in S} \in [0,1]^{|S|}$ with $\theta'_s \leq \theta_s$ for every $s \in S$, and any $\theta \in [0, 1]$,

$$v(\theta; (\theta'_s)_{s\in S}) \geq v(\theta; (\theta_s)_{s\in S}) \quad \textit{and} \quad \tilde{v}(\theta; (\theta'_s)_{s\in S}) \geq \tilde{v}(\theta; (\theta_s)_{s\in S}).$$

Hence $\int \tilde{v}(\theta; (\theta'_s)_{s\in S})\, F(d\theta) \geq \int \tilde{v}(\theta; (\theta_s)_{s\in S})\, F(d\theta)$ for every $F \in \Delta([0, 1])$, and taking the infimum over any common non-empty subset of $\Delta([0, 1])$ preserves the inequality.

Proof of Lemma A.3. For v, $v(\theta;(\theta_s)_{s\in S}) = \bar{u}_0 + \sum_{s\in S} w_s \mathbf{1}\{\theta \geq \theta_s\}$ with $w_s > 0$, so lowering any threshold weakly enlarges its super-level set, giving the pointwise inequality for v.

For $\tilde{v}$, applying Definition A.1 to each indicator gives

$$\tilde{v}(\theta;(\theta_s)_{s\in S}) = \bar{u}_0 + \sum_{s\in S} w_s\big(\mathbf{1}\{\theta > \theta_s\} + \mathbf{1}\{\theta_s = 0\}\mathbf{1}\{\theta = 0\}\big),$$

because the lower semi-continuous envelope of $\mathbf{1}\{\theta \geq \theta_s\}$ on $[0,1]$ equals $\mathbf{1}\{\theta > \theta_s\}$ when $\theta_s > 0$ and equals 1 when $\theta_s = 0$. Lowering a threshold enlarges the corresponding open upper set and, at $\theta = 0$, can only add the endpoint correction. This gives the pointwise inequality for $\tilde{v}$. Integration and infimum over a fixed feasible set preserve pointwise order. □

Lemma A.4 (Compactness and attainment for moment ambiguity sets)
Let Θ be a compact metric space, $g\colon \Theta \to \mathbb{R}^d$ continuous, and $Y \subseteq \mathbb{R}^d$ closed. Set $\Pi := \{F \in \Delta(\Theta) : \mathbb{E}_F[g(\theta)] \in Y\}$ and assume $\Pi \neq \varnothing$. Then:

(a) Π is weakly compact in $\Delta(\Theta)$.

(b) For every bounded lower semi-continuous $f\colon \Theta \to \mathbb{R}$, $F \mapsto \int_\Theta f\,dF$ is weakly lower semi-continuous on $\Delta(\Theta)$.

(c) $\inf_{F\in\Pi} \int_\Theta f\,dF$ is attained whenever f is bounded and lower semi-continuous.

Proof of Lemma A.4. $\Delta(\Theta)$ is weakly compact by Aliprantis and Border (2006, Theorem 15.11). Continuity of g on the compact Θ makes $F \mapsto \mathbb{E}_F[g(\theta)]$ weakly continuous, so Π is the pre-image of a closed set under a weakly continuous map, hence weakly closed; a weakly closed subset of a weakly compact set is weakly compact, giving (a). Part (b) is Aliprantis and Border (2006, Theorem 15.5). (c) follows from the Weierstrass theorem: a weakly lower semi-continuous function on a non-empty weakly compact set attains its infimum. □

A.2 Proof of Theorem 1

Proof. For each $s \in S$, if no type is persuaded after s under M, choose an arbitrary narrative from M; otherwise let $\theta_s(M)$ be the persuasion threshold and retain a narrative fit-maximal for type $\theta_s(M)$ after s and inducing a posterior of at least τ there. Let $M' \subseteq M$ be the set of retained narratives, so $|M'| \leq |S|$.

Fix s. If no type is persuaded under M, there is nothing to prove. Otherwise the retained narrative remains fit-maximal for $\theta_s(M)$ within M' because $M' \subseteq M$, so sender-favorable tie-breaking gives a posterior of at least τ for type $\theta_s(M)$ under M'. Monotonicity of the posterior then extends persuasion under M' to every $\theta \geq \theta_s(M)$, which is exactly the realization-s upper interval of types induced to take action 1 under M.

Summing the per-realization inclusions over s and using $w_s > 0$ gives $v(\theta;M') \geq v(\theta;M)$ for every $\theta \in [0,1]$; the infimum over $F \in \Pi$ gives $V(M') \geq V(M)$. □

A.3 Proof of Theorem 2

Proof. Part (i). The restriction to $[0,\tau]^{|S|}$ records the part of the threshold problem relevant for the sender: any threshold above τ can be lowered to τ, weakly expanding persuasion.

Sufficiency. For each $s \in S$, set $m_s(s \mid H) = 1$, $m_s(s' \mid H) = 0$ for $s' \neq s$, and

$$m_s(s \mid L) = \frac{\theta_s(1-\tau)}{(1-\theta_s)\tau}.$$

Choose the remaining L-row entries with $0 \leq m_s(s' \mid L) \leq \min\{1, \theta_{s'}/(\tau(1-\theta_{s'}))\}$ summing to $1 - m_s(s \mid L)$. Such a row exists because (7) gives the required slack:

$$1 - \frac{\theta_s(1-\tau)}{(1-\theta_s)\tau} \leq \sum_{s' \neq s} \min\Big\{1, \frac{\theta_{s'}}{\tau(1-\theta_{s'})}\Big\}.$$

At signal s, m_s reaches the action threshold at θ_s (with $\theta_s = 0$ read in the right limit), while every $m_{s'}$ with $s' \neq s$ has fit $(1-\theta_s)m_{s'}(s \mid L) \leq \theta_s/\tau$. Hence m_s is fit-maximal at (s, θ_s), and the induced threshold is θ_s.

Necessity. Let m_s be selected at (s, θ_s) in some implementing menu. The posterior condition gives

$$m_s(s \mid L) \leq \frac{\theta_s(1-\tau)}{(1-\theta_s)\tau} m_s(s \mid H) \leq \frac{\theta_s(1-\tau)}{(1-\theta_s)\tau}.$$

Fix $j \in S$. For $s \neq j$, fit-maximality of m_s at (s, θ_s) gives

$$\theta_s\, m_j(s \mid H) + (1-\theta_s)\, m_j(s \mid L) \leq \theta_s\, m_s(s \mid H) + (1-\theta_s)\, m_s(s \mid L) \leq \theta_s/\tau,$$

so $m_j(s \mid L) \leq \theta_s/(\tau(1-\theta_s))$. Summing the L-row of m_j,

$$1 \leq \frac{\theta_j(1-\tau)}{(1-\theta_j)\tau} + \sum_{s \neq j} \frac{\theta_s}{\tau(1-\theta_s)}.$$

Multiplying by τ and maximizing over j yields (7).

Part (ii). Fix an ordering $r_{(1)} \leq \cdots \leq r_{(|S|)}$. The maximum in (7) is then attained at $r_{(|S|)}$, so the binding inequality $\sum_{s \in S} r_s = \tau(1 + r_{(|S|)})$ rearranges to $\sum_{i=1}^{|S|-1} r_{(i)} + (1-\tau)\, r_{(|S|)} = \tau$. The remaining $|S| - 1$ inequalities are weakly slack under this ordering. Permuting labels gives the global frontier. □

Lemma A.5 (τ-capping dominance)
For any menu $M \in \mathcal{M}$, the capped profile $(\bar\theta_s)_{s \in S}$ with $\bar\theta_s := \min\{\theta_s(M), \tau\}$ is implementable, and

$$v(\theta; (\bar\theta_s)_{s \in S}) \geq v(\theta; M), \qquad \tilde v(\theta; (\bar\theta_s)_{s \in S}) \geq \tilde v(\theta; M)$$

for every $\theta \in [0,1]$.

Proof of Lemma A.5. If $\theta_s(M) \leq \tau$ for every $s \in S$, then $(\theta_s(M))_{s \in S} = (\bar\theta_s)_{s \in S}$ is already implementable and there is nothing to prove.

Otherwise fix $s^* \in S$ with $\theta_{s^*}(M) \geq \tau$, so $\bar{\theta}_{s^*} = \tau$ and $\bar{r}_{s^*} := \bar{\theta}_{s^*}/(1-\bar{\theta}_{s^*}) = \tau/(1-\tau)$. Since $\bar{\theta}_s \leq \tau$ for every $s \in S$ by construction, $\bar{r}_s \leq \tau/(1-\tau)$ for every $s \in S$, so $\max_{s\in S} \bar{r}_s = \tau/(1-\tau)$. Using $\bar{r}_s \geq 0$ for $s \neq s^*$,

$$\sum_{s\in S} \bar{r}_s \geq \bar{r}_{s^*} = \frac{\tau}{1-\tau} = \tau\Big(1 + \frac{\tau}{1-\tau}\Big) = \tau\Big(1 + \max_{s\in S} \bar{r}_s\Big).$$

This is exactly (7) for $(\bar{\theta}_s)_{s\in S} \in [0,\tau]^{|S|}$, so the sufficiency direction of Theorem 2 gives a menu implementing $(\bar{\theta}_s)_{s\in S}$.

Since $\bar{\theta}_s = \min\{\theta_s(M), \tau\} \leq \theta_s(M)$ for every $s \in S$, Lemma A.3 gives $v(\theta; (\bar{\theta}_s)_{s\in S}) \geq v(\theta; (\theta_s(M))_{s\in S}) = v(\theta; M)$ for every $\theta \in [0,1]$, and likewise for $\tilde{v}$. □

A.4 Proof of Corollary 2

Proof. Let $(\theta_s)_{s\in S}$ be any implementable profile, so $\theta_s \in [0,\tau]$ for every $s \in S$. By symmetry of (7), relabel the threshold coordinates so that lower thresholds are assigned to weakly larger payoff weights: after ordering $w_{s_1} \geq \cdots \geq w_{s_n}$, take $0 \leq \theta_{s_1} \leq \cdots \leq \theta_{s_n} \leq \tau$. This rearrangement weakly raises pointwise payoff by the standard pairwise swap argument: if $w_i \geq w_j$ but $\theta_i > \theta_j$, swapping the two thresholds changes payoff by $(w_i - w_j)\mathbf{1}\{\theta_j \leq \theta < \theta_i\} \geq 0$.

Write $r_{s_i} := \theta_{s_i}/(1-\theta_{s_i})$. Since the ordered profile is implementable, Theorem 2 gives

$$L := \sum_{i=1}^{n-1} r_{s_i} + (1-\tau) r_{s_n} \geq \tau.$$

If $L = \tau$, the profile is already on the persuasion frontier. If $L > \tau$, set $\lambda := \tau/L \in (0,1)$ and define $r'_{s_i} := \lambda r_{s_i}$ for every i. The ordering is preserved, $r'_{s_i} \leq r_{s_i}$ for every i, and the new profile lies on the persuasion frontier:

$$\sum_{i=1}^{n-1} r'_{s_i} + (1-\tau) r'_{s_n} = \tau.$$

Converting back by $\theta'_{s_i} = r'_{s_i}/(1+r'_{s_i})$ gives an implementable profile by Theorem 2. Since every threshold weakly falls, Lemma A.3 gives pointwise weak dominance. The maximum can therefore be restricted to the persuasion frontier. □

A.5 Proof of Proposition 1

Proof for Proposition 1. By Lemma A.5, every menu is pointwise weakly dominated in v by one inducing an implementable threshold profile in $[0,\tau]^{|S|}$, so the sender can restrict attention to that set without loss. In odds coordinates this set is closed by (7) (Theorem 2); hence it is compact. For each $F \in \Pi$, the map

$$(\theta_s)_{s\in S} \mapsto \int v(\theta; (\theta_s)_{s\in S})\, F(d\theta)$$

is upper semi-continuous: for any convergent sequence of threshold profiles, apply Aliprantis and Border (2006, Lemma 11.20) to the bounded nonnegative functions obtained by subtracting the finite sum of upper semi-continuous indicators in (5) from $\bar{u}_0 + \sum_{s \in S} w_s$. Therefore V, as the infimum over $F \in \Pi$ of upper semi-continuous functions on a compact set, is upper semi-continuous (Aliprantis and Border, 2006, Lemma 2.41). The Weierstrass theorem gives an optimal threshold profile, and Theorem 2 gives a finite menu implementing it. □

A.6 Default-model extension

This extension lets receivers choose from $M \cup \{\pi\}$, where π is the fixed data-generating model. The sender still chooses only M. For a menu M, let $\theta_s^\pi(M)$ denote the persuasion threshold after realization $s \in S$ under this enlarged choice set. Define

$$\underline{\theta}_s^\pi := \frac{\tau\, \pi(s \mid L)}{1 - \tau\, \pi(s \mid H) + \tau\, \pi(s \mid L)}, \qquad s \in S. \tag{A.1}$$

Proposition A.1 (Reduction and existence under the default model)
Under the default model:

(i) *for every menu M and every realization $s \in S$, the induced persuasion set is an upper set, hence is represented by a threshold $\theta_s^\pi(M)$;*

(ii) *every menu is pointwise weakly dominated by a sub-menu of M with at most $|S|$ narratives;*

(iii) *an optimal menu exists.*

Proof. Apply Lemma A.1 and Corollary A.1 to the enlarged menu $M \cup \{\pi\}$. This gives monotone selected posteriors and upper persuasion sets after every realization.

For each $s \in S$, if no type is persuaded after s, retain no narrative for that realization. If the threshold is set by a sender-provided narrative, retain one such narrative. If the threshold is set by π, retain no sender-provided narrative for that realization. If no narrative from M has been retained after this step, keep an arbitrary element of M. The retained menu has at most $|S|$ narratives. After adjoining the fixed default model π, the threshold-setting argument used in Theorem 1 applies realization by realization: the retained threshold-setting object, whether it is sender-provided or π, is still available, and its posterior is non-decreasing in type. Hence every type persuaded under $M \cup \{\pi\}$ remains persuaded under the retained sub-menu together with π.

For existence, Proposition A.2 below identifies the feasible threshold profiles under the default model as the intersection of the compact implementability set in Theorem 2 with the closed coordinatewise lower bounds in (A.1). The upper semi-continuity argument in the proof of Proposition 1 therefore applies on this compact feasible set, and Proposition A.2 supplies an implementing menu for an optimal profile. □

Proposition A.2 (Default-model implementability)
A threshold profile $(\theta_s)_{s \in S} \in [0, \tau]^{|S|}$ is implementable under the default model if and only if it is implementable in the problem without the default model and

$$\theta_s \geq \underline{\theta}_s^\pi \qquad \text{for every } s \in S. \tag{A.2}$$

Consequently, whenever $\pi(s \mid L) > 0$ for some realization s, the default-model feasible set rules out profiles with $\theta_s < \underline{\theta}_s^{\pi}$.

Proof. Necessity. If $(\theta_s)_{s\in S}$ is induced by $M \cup \{\pi\}$, then it is implementable without the default model by the menu $M \cup \{\pi\}$, since π is itself a stochastic kernel from Ω to $\Delta(S)$ and sender menus are unrestricted.

Fix $s \in S$. If π is selected at the marginal type θ_s and persuades, then θ_s is at least the persuasion threshold of π after s:

$$\frac{\tau\,\pi(s \mid L)}{(1-\tau)\pi(s \mid H) + \tau\,\pi(s \mid L)}.$$

This number is weakly above $\underline{\theta}_s^{\pi}$, because its denominator is weakly smaller than the denominator in (A.1). Hence (A.2) holds.

If a sender-provided narrative m_s is selected at the marginal type and persuades, then

$$\mathrm{Fit}(s, m_s, \theta_s) \le \frac{\theta_s}{\tau},$$

because posterior persuasion implies

$$\theta_s m_s(s \mid H) \ge \tau\ \mathrm{Fit}(s, m_s, \theta_s)$$

and $m_s(s \mid H) \le 1$. Since m_s must weakly beat the default model at the same type,

$$(1-\theta_s)\pi(s \mid L) + \theta_s \pi(s \mid H) \le \frac{\theta_s}{\tau}.$$

Rearranging gives $\theta_s \ge \underline{\theta}_s^{\pi}$.

Sufficiency. Take a profile that satisfies Theorem 2 and (A.2). Use the constructive menu from the proof of Theorem 2: for each $s \in S$, the s-threshold narrative m_s has $m_s(s \mid H) = 1$ and reaches posterior τ at θ_s. Its fit at (s, θ_s) is θ_s/τ. The bound (A.2) is exactly

$$\frac{\theta_s}{\tau} \ge (1-\theta_s)\pi(s \mid L) + \theta_s \pi(s \mid H),$$

so m_s weakly beats the default model at the marginal type.

For $\theta < \theta_s$, the same constructed narrative has fit strictly above θ/τ after realization s. If the default model persuaded such a type, then

$$\mathrm{Fit}(s, \pi, \theta) \le \frac{\theta\,\pi(s \mid H)}{\tau} \le \frac{\theta}{\tau},$$

so the default model could not be selected against m_s. The sender-provided narratives already implement the threshold profile without the default model by Theorem 2; adding π therefore does not create persuasion below any θ_s.

It remains to check that adding π does not destroy persuasion above θ_s. The default model itself reaches posterior τ after s at

$$\frac{\tau\,\pi(s \mid L)}{(1-\tau)\pi(s \mid H) + \tau\pi(s \mid L)}$$

when the denominator is positive. If θ is weakly above this value, selection of π is harmless because π persuades. Consider lower values of θ in the interval starting at θ_s. The constructed narrative has fit

$$\theta + (1-\theta)\frac{\theta_s(1-\tau)}{(1-\theta_s)\tau},$$

while the default model has fit

$$(1-\theta)\pi(s \mid L) + \theta\pi(s \mid H).$$

Both are affine in θ. The lower bound (A.2) gives the constructed narrative weakly higher fit at θ_s. To check the other endpoint, compare the two fits at the type where the default model itself reaches posterior τ. A direct calculation gives

$$\begin{aligned}\frac{\partial}{\partial\theta_s}&\{\text{constructed fit minus default-model fit at that type}\}\\ &= -\frac{\pi(s \mid H)(1-\tau)^2}{\tau(1-\theta_s)^2(\tau\pi(s \mid H) - \tau\pi(s \mid L) - \pi(s \mid H))} \geq 0,\end{aligned}$$

as a function of θ_s. This derivative is non-negative because $\tau\pi(s \mid H) - \tau\pi(s \mid L) - \pi(s \mid H) = -((1-\tau)\pi(s \mid H) + \tau\pi(s \mid L)) < 0$. At $\theta_s = \underline{\theta}_s^{\pi}$, the constructed fit minus default-model fit at that type equals

$$\frac{\tau\pi(s \mid L)(\pi(s \mid H) - 1)^2}{(\tau\pi(s \mid H) - 1)(\tau\pi(s \mid H) - \tau\pi(s \mid L) - \pi(s \mid H))} \geq 0,$$

with the zero-probability boundary cases obtained by continuity. Hence the constructed narrative weakly beats the default model at the default model's own persuasion threshold whenever (A.2) holds. Since both fit functions are affine, the constructed narrative weakly beats the default model throughout the interval between θ_s and the default model's own persuasion threshold. Thus the default model cannot displace the threshold narrative on the part of the upper interval where the default model would not persuade. Persuasion is preserved for every $\theta \geq \theta_s$, and the profile is implementable under the default model. □

For any threshold profile implementable under the default model, the inner minimization problem is unchanged after restricting the feasible threshold profiles by (A.2). Hence the payoff-guarantee equivalence, dual characterization, finite-support worst-case distribution, and contact-set characterization apply on the restricted feasible set.

A.7 Beyond binary states

This subsection records the formal condition behind the scope discussion in Section 3.3.

Proposition A.3 (Scalar-statistic threshold sufficiency)
Let $\mathcal{Q}$ be a primitive state space, let $X = [\underline{x}, \bar{x}] \subset \mathbb{R}$ be an ordered scalar index space, and let $T : \Delta(\mathcal{Q}) \to X$ be a posterior statistic. For each menu M and realization s, let $Z_s^M(\nu)$ be the scalar posterior index generated by the narrative selected after s from M when the primitive posterior is $\nu \in \Delta(\mathcal{Q})$. Suppose:

(i) the induced posterior index factors through T: for every M and s, there exists a map $z_s^M : X \to X$ such that

$$Z_s^M(\nu) = z_s^M(T(\nu)) \quad \text{for every } \nu \in \Delta(\mathcal{Q});$$

(ii) z_s^M is non-decreasing for every M and s;

(iii) the receiver takes action 1 if and only if $z_s^M(x) \geq \kappa$ for a fixed cutoff κ;

(iv) the sender's ordinal comparison of menus depends on primitive posterior beliefs only through the induced scalar-index thresholds;

(v) for the sender payoff representation below, the sender's payoff from action 1 after realization s is $w_s > 0$, and residual features of the primitive posterior do not enter her payoff.

Then the primitive model is behavior-equivalent to a scalar-index model with receiver index $x = T(\nu)$: conditions (i)-(iii) *give receiver behavior equivalence, and condition* (iv) *gives sender menu-choice equivalence. For every menu M and realization s, the set $\{x : z_s^M(x) \geq \kappa\}$ is an upper interval. Let*

$$\theta_s(M) := \inf\{x : z_s^M(x) \geq \kappa\},$$

with the convention that $\theta_s(M) = \bar{x}^+$ when the set is empty, where $\bar{x}^+$ is an artificial point above $\bar{x}$ and $\mathbf{1}\{x \geq \bar{x}^+\} = 0$ for every $x \in X$. If condition (v) *also holds, the sender's per-type payoff is*

$$\bar{u}_0 + \sum_{s \in S} w_s \mathbf{1}\{x \geq \theta_s(M)\}.$$

Thus receiver behavior is represented by the threshold profile $(\theta_s(M))_{s \in S}$; under condition (iv)*, the same profile represents sender menu choice; and under condition* (v)*, it also represents the sender's cardinal payoff-relevant part of the menu.*

Proof. Condition (i) says that primitive posteriors with the same statistic T produce the same induced posterior index after every realization and menu. Hence the primitive posterior ν can be replaced, for receiver behavior purposes, by the scalar index $x = T(\nu)$. Condition (iii) then gives the action rule in the scalar-index model.

By condition (ii), for each M and s the super-level set $\{x : z_s^M(x) \geq \kappa\}$ is an upper interval. Its lower endpoint is $\theta_s(M)$ under the stated convention, so the induced action after s is $\mathbf{1}\{x \geq \theta_s(M)\}$. This proves the receiver-side threshold representation. Condition (iv) makes the scalar-index threshold profile sufficient for the sender's menu choice. Condition (v) rules out residual payoff terms that distinguish primitive posteriors with the same scalar statistic. Under that additional payoff restriction, the sender's payoff is exactly the displayed threshold payoff, and the menu is payoff-relevant only through $(\theta_s(M))_{s \in S}$. □

Binary partition. Let $H \subseteq \mathcal{Q}$, write $L = \mathcal{Q} \setminus H$, and set $T(\nu) = \nu(H)$. If a seller wants the receiver to buy and the receiver buys whenever product quality q exceeds an adequacy cutoff q^*, the natural partition is $H = \{q \geq q^*\}$ and $L = \{q < q^*\}$. This partition is receiver-behavior-relevant when conditions (i)-(iii) in Proposition A.3 hold for $T(\nu) = \nu(H)$: action

choice and the selected posterior index factor through the partition-induced coarsening, so receiver behavior is invariant to redistributions of posterior probability within H and within L. A primitive sufficient condition is that receiver utility differences and narrative likelihoods are measurable with respect to $\sigma(\{H, L\})$. Behavior equivalence including sender menu choice additionally requires condition (iv). The threshold-payoff representation additionally requires condition (v) for the sender objective.

Scalar posterior statistic. If the primitive state is real-valued and the receiver takes action 1 whenever the posterior mean exceeds a cutoff, take $T(\nu) = \int q \, d\nu(q)$. The posterior mean then plays the same role as the belief on H, provided that the selected posterior index factors through T. Sender behavior equivalence additionally requires the sender's menu ordering to depend only on the scalar-index thresholds; payoff-guarantee equivalence further requires payoff increments to depend on ν only through $T(\nu)$. This is the scalar-statistic case discussed in Section 3.3. If any condition in Proposition A.3 fails, the scalar index need not summarize the design problem: the action rule may fail to be an upper interval, the induced posterior index may fail to be monotone in the index, the sender's menu ordering may use residual posterior features, or the sender's payoff may use posterior features not captured by the thresholds.

A.8 Proof of Lemma 1

Proof of Lemma 1. Fix M and $F \in \Pi$. Applying Ball and Kattwinkel (2026, Lemma 1) to the bounded nonnegative payoff $v(\cdot; M) - \bar{u}_0$ yields a weakly convergent sequence $\{F_n\} \subset \Delta(\Theta)$ with $\widetilde{W}(M, F) \geq \liminf_{n\to\infty} W(M, F_n)$.

Richness of Π implies uniform robustness by their Proposition 5, so $\liminf_n W(M, F_n) \geq V(M)$, since $\inf_{G\in\Pi} W(M, G) = V(M)$. Thus $\widetilde{W}(M, F) \geq V(M)$ for every $F \in \Pi$, and hence $\inf_{F\in\Pi} \widetilde{W}(M, F) \geq V(M)$. The reverse inequality follows from $\mathrm{lsc}(v(\cdot, M)) \leq v(\cdot, M)$ pointwise. Therefore $\inf_{F\in\Pi} \widetilde{W}(M, F) = V(M)$.

The payoff $\tilde{v}(\cdot, M)$ is bounded and lower semi-continuous in θ, so Lemma A.4 gives an attaining $F^* \in \Pi$. Hence $\widetilde{V}(M) = V(M)$ for every menu M, and the maximizers and maximum values of the two objective functions coincide by Proposition 1. □

Proposition A.4 (Positive payoff guarantee above $\bar{u}_0$)
Suppose $\Pi = \mathcal{F}(g, Y)$, where $g : [0, 1] \to \mathbb{R}^d$ is continuous, $Y \subseteq \mathrm{conv}\, g([0, 1])$ is closed, and $\Pi \neq \varnothing$. Fix an implementable threshold profile $(\theta_s)_{s\in S}$ and let

$$\Delta(\theta) := \tilde{v}(\theta; (\theta_s)_{s\in S}) - \bar{u}_0.$$

If $\theta_s = 0$ for some $s \in S$, then

$$\inf_{F\in\Pi} \int_0^1 \Delta(\theta)\, F(d\theta) > 0.$$

If $\theta_s > 0$ for every $s \in S$ and $\underline{\theta} := \min_{s\in S} \theta_s$, then

$$\inf_{F\in\Pi} \int_0^1 \Delta(\theta)\, F(d\theta) > 0 \quad \Longleftrightarrow \quad Y \cap \mathrm{conv}\, g([0, \underline{\theta}]) = \varnothing.$$

Proof. If $\theta_s = 0$ for some s, then the lower semi-continuous payoff convention in (13) gives

$$\mathbf{1}\{\theta > \theta_s\} + \mathbf{1}\{\theta_s = 0\}\mathbf{1}\{\theta = 0\} = 1$$

for every $\theta \in [0,1]$. Since $w_s > 0$, $\Delta(\theta) \geq w_s$ pointwise, and the infimum is at least w_s.

Now suppose $\theta_s > 0$ for every s and write $\underline{\theta} = \min_s \theta_s$. Since all w_s are strictly positive,

$$\Delta(\theta) = \sum_{s \in S} w_s \mathbf{1}\{\theta > \theta_s\}$$

is strictly positive exactly on $(\underline{\theta}, 1]$. Hence

$$\inf_{F \in \Pi} \int_0^1 \Delta(\theta)\, F(d\theta) > 0 \quad \Longleftrightarrow \quad \inf_{F \in \Pi} F((\underline{\theta}, 1]) > 0.$$

It remains to classify the right-hand condition.

Let $A = (\underline{\theta}, 1]$. If $Y \cap \operatorname{conv} g([0, \underline{\theta}]) \neq \varnothing$, choose $y \in Y \cap \operatorname{conv} g([0, \underline{\theta}])$. By the definition of convex hull on the compact interval, there is $F \in \Delta([0, \underline{\theta}])$ with $\mathbb{E}_F[g(\theta)] = y$. Then $F \in \Pi$ and $F(A) = 0$, so $\inf_{G \in \Pi} G(A) = 0$.

Conversely, suppose $\inf_{F \in \Pi} F(A) = 0$. Choose $F_n \in \Pi$ with $F_n(A) \to 0$. Since $\Delta([0,1])$ is weakly compact, a subsequence converges weakly to some $F^* \in \Delta([0,1])$. The set A is open in the relative topology of $[0,1]$, so Portmanteau's theorem gives

$$F^*(A) \leq \liminf_n F_n(A) = 0.$$

Thus F^* is supported on $[0, \underline{\theta}]$. Continuity of g gives

$$\mathbb{E}_{F_n}[g(\theta)] \to \mathbb{E}_{F^*}[g(\theta)]$$

along the convergent subsequence. Since each moment vector lies in the closed set Y, the limit belongs to Y. Since F^* is supported on $[0, \underline{\theta}]$, the same limit also belongs to $\operatorname{conv} g([0, \underline{\theta}])$. Therefore $Y \cap \operatorname{conv} g([0, \underline{\theta}]) \neq \varnothing$. □

Public randomization. For a public randomization $\nu \in \Delta(\mathcal{M})$, let $\widetilde{W}(\nu, F)$ be the ν-average of $\widetilde{W}(M, F)$ over $\mathcal{M}$. If the sender could publicly randomize over menus and the realized menu were observed before receivers act, the mixed problem for $\widetilde{W}$ would satisfy a minimax identity. Since Y is convex, $\Pi = \mathcal{F}(g, Y)$ is convex; by Lemma A.4(a), it is weakly compact. For each M, $F \mapsto \widetilde{W}(M, F)$ is bounded and weakly lower semi-continuous by Lemma A.4(b), affine in F, and jointly Borel measurable under the finite-menu parameterization and the fixed tie-breaking rule. The mixed-strategy minimax theorem of Alpern and Gal (1988) therefore gives

$$\sup_{\nu \in \Delta(\mathcal{M})} \inf_{F \in \Pi} \widetilde{W}(\nu, F) = \inf_{F \in \Pi} \sup_{\nu \in \Delta(\mathcal{M})} \widetilde{W}(\nu, F) = \inf_{F \in \Pi} \sup_{M \in \mathcal{M}} \widetilde{W}(M, F),$$

where the second equality follows because $\widetilde{W}(\nu, F)$ is linear in ν. Thus public randomization eliminates the max-min gap relative to the distribution-specific benchmark; deterministic public menus are the environment in which the cost of ambiguity studied here arises.

A.9 Proof of Corollary 3

Proof of Corollary 3. By Lemma 1, the robust objective can be evaluated with $\tilde{v}$. By Lemma A.5, every menu $M \in \mathcal{M}$ is pointwise weakly dominated in $\tilde{v}$ by a menu inducing an implementable threshold profile in $[0,\tau]^{|S|}$, so the outer maximum over $\mathcal{M}$ can be restricted to implementable threshold profiles. Corollary 2 then gives, for every such implementable threshold profile, a persuasion frontier profile that weakly lowers every threshold after assigning lower thresholds to weakly higher payoff weights. By Lemma A.3, this raises $\tilde{v}$ pointwise in θ, and therefore weakly raises the robust payoff guarantee after taking the infimum over $F \in \Pi$. Combined with attainment in Proposition 1, this gives the max-min identity in Corollary 3.

When $S = \{h, l\}$ and $w_h \geq w_l$, the persuasion frontier is exactly the h-first branch $\psi(\theta_h, \theta_l) = 0$ with $0 \leq \theta_h \leq \theta_l \leq \tau$. The case $w_l > w_h$ follows by exchanging h and l and replacing ψ by ϕ. □

A.10 Proof of Proposition 2

Proof of Proposition 2. By Definition A.1, $\tilde{v}(\cdot; (\theta_s)_{s\in S})$ is bounded and lower semi-continuous on the compact $[0,1]$.

Star g-interiority of Y gives the required relative-interior condition: combine a point in $Y \cap \operatorname{ri}\operatorname{conv} g([0,1])$ with any point of $\operatorname{ri} Y$ and apply Rockafellar (1970, Theorem 6.1). The compact support, continuity of g, closed convexity of Y, and bounded lower semi-continuity of $\tilde{v}$ therefore yield the displayed strong dual in Proposition 2. Primal attainment follows from weak compactness and lower semi-continuity by Lemma A.4. The support bound $|\operatorname{supp} F^*| \leq d+1$ follows from Winkler (1988, Theorem 2.1).

Complementary slackness gives F^*-a.s. equality $\tilde{v}(\theta; (\theta_s)_{s\in S}) = \gamma^* + (\lambda^*)^\top g(\theta)$, so the support of F^* lies in the closed contact set. When $g(\theta) = (\theta, \theta^2, \ldots, \theta^d)$, the dual minorant is a polynomial of degree at most d. □

A.11 Proof of Proposition 3

Proof of Proposition 3. The saddle inequalities $\widetilde{W}(M, F^*) \leq \widetilde{W}(M^*, F^*) \leq \widetilde{W}(M^*, F)$ for all $M \in \mathcal{M}$, $F \in \Pi$ give $M^* \in \arg\max_M \widetilde{W}(M, F^*)$ and $\widetilde{W}(M^*, F^*) = \min_F \widetilde{W}(M^*, F)$. Hence

$$\min_F \widetilde{W}(M, F) \leq \widetilde{W}(M, F^*) \leq \min_F \widetilde{W}(M^*, F) \quad \text{for every } M \in \mathcal{M},$$

so $M^* \in \arg\max_M \min_F \widetilde{W}(M, F)$. Lemma 1 replaces $\widetilde{W}$ by W inside the infimum, giving $M^* \in \arg\max_M \inf_F W(M, F)$. □

A.12 Saddle-point criterion for benchmark selection

Proposition A.5 (Saddle-point criterion for benchmark selection)
Let M^* induce threshold profile $\theta^* = (\theta^*_s)_{s\in S}$, and fix $F^* \in \arg\min_{F\in\Pi} \widetilde{W}(M^*, F)$. Then

(M^*, F^*) *is a saddle point of* $\widetilde{W}$ *on* $\mathcal{M} \times \Pi$ *if and only if*

$$\int \tilde{v}(\theta; (\theta_s)_{s\in S})\, F^*(d\theta) \le \int \tilde{v}(\theta; \theta^*)\, F^*(d\theta) \tag{A.3}$$

for every threshold profile $(\theta_s)_{s\in S}$ *for which there exists an enumeration* $s_1, \dots, s_n$ *of* S, $n = |S|$, *satisfying*

$$w_{s_1} \ge \cdots \ge w_{s_n}, \qquad 0 \le \theta_{s_1} \le \cdots \le \theta_{s_n} \le \tau,$$

and

$$\sum_{i=1}^{n-1} \frac{\theta_{s_i}}{1-\theta_{s_i}} + (1-\tau)\frac{\theta_{s_n}}{1-\theta_{s_n}} = \tau.$$

Whenever (A.3) *holds, Proposition 3 implies that* M^* *is benchmark-optimal at* F^*.

Proof of Proposition A.5. Since F^* minimizes $\widetilde{W}(M^*, \cdot)$ over Π, the right saddle inequality

$$\widetilde{W}(M^*, F^*) \le \widetilde{W}(M^*, F) \qquad \text{for every } F \in \Pi$$

already holds. It remains to characterize the left saddle inequality. By Corollary 2 and Lemma A.3, every menu is pointwise weakly dominated in $\tilde{v}$ by a persuasion frontier profile of the form displayed in the proposition. Integrating against F^* preserves pointwise dominance. Hence

$$\widetilde{W}(M, F^*) \le \widetilde{W}(M^*, F^*) \qquad \text{for every } M \in \mathcal{M}$$

holds if and only if every persuasion frontier profile is non-improving at F^*, which is exactly (A.3). The final statement follows from Proposition 3. □

B Proofs for applications

The proofs below specialize the frontier restriction and the moment-problem representation to the political-narrative application in Section 4.

B.1 Proof of Proposition 4

Proof of Proposition 4. Part (i). Fix $t \in (0,1)$. Any quadratic minorant $q(\theta) = \gamma + \lambda_1\theta + \lambda_2\theta^2$ with $\lambda_2 \ge 0$ and $q(\theta) \le \mathbf{1}[\theta > t]$ gives

$$F((t,1]) = \mathbb{E}_F[\mathbf{1}[\theta > t]] \ge \mathbb{E}_F[q(\theta)] \ge \gamma + \lambda_1\hat{\theta} + \lambda_2(\hat{\theta}^2 + \underline{\sigma}^2).$$

Among affine minorants, $q(t) \le 0$ and $q(1) \le 1$, so $q(\hat{\theta}) \le (\hat{\theta} - t)/(1-t)$ when $\hat{\theta} > t$, with equality at $q(\theta) = (\theta - t)/(1-t)$; when $\hat{\theta} \le t$, the affine bound is nonpositive. Among quadratic minorants with roots 0 and t, $q(\theta) = \lambda\theta(\theta - t)$ is feasible exactly when $0 \le \lambda \le 1/(1-t)$, giving the bound $(\underline{\sigma}^2 - \hat{\theta}(t - \hat{\theta}))/(1-t)$ when this term is positive and no positive bound otherwise. Taking the larger lower bound gives the right-hand side of (14).

The bound is tight. If $\hat{\theta} \le t$ and $\underline{\sigma}^2 \le \hat{\theta}(t - \hat{\theta})$, the distribution $(1 - \hat{\theta}/t)\delta_0 + (\hat{\theta}/t)\delta_t$ is feasible and has tail mass 0. If $\hat{\theta} > t$ and $\underline{\sigma}^2 \le (\hat{\theta} - t)(1 - \hat{\theta})$, the distribution $((1-\hat{\theta})/(1-$

$t))\delta_t + ((\hat{\theta} - t)/(1-t))\delta_1$ is feasible and has tail mass $(\hat{\theta} - t)/(1-t)$. In the remaining case, use $a\delta_0 + b\delta_t + c\delta_1$, where

$$b = \frac{\hat{\theta}(1-\hat{\theta}) - \underline{\sigma}^2}{t(1-t)}, \qquad c = \frac{\underline{\sigma}^2 - \hat{\theta}(t-\hat{\theta})}{1-t}, \qquad a = 1 - b - c.$$

A direct calculation gives mean $\hat{\theta}$, variance $\underline{\sigma}^2$, and tail mass c. Nonnegativity follows from $\underline{\sigma}^2 < \hat{\theta}(1-\hat{\theta})$, the remaining-case inequality, and $a = (\underline{\sigma}^2 + (t-\hat{\theta})(1-\hat{\theta}))/t \geq 0$. Thus each piece of (14) is attained.

Part (ii). At $\tau = 1/2$, the boundary menu $(0, 1/2)$ pays w_h to every type and w_l to types above $1/2$, giving payoff guarantee $w_h + w_l \underline{T}(1/2)$ by part (i); the value at $(1/2, 0)$ is symmetric. Their difference $(w_h - w_l)(1 - \underline{T}(1/2))$ has $\underline{T}(1/2) < 1$ since each piece of (14) at $t = 1/2$ is strictly below one under $\underline{\sigma}^2 < \hat{\theta}(1-\hat{\theta})$. Hence $(0, 1/2)$ weakly dominates iff $w_h \geq w_l$; when $w_h < w_l$ the inequality reverses and the symmetric boundary value is $w_l + w_h \underline{T}(1/2)$ at $(1/2, 0)$. □

B.2 Proof of Proposition 5

Proof of Proposition 5. We prove the case $w_h \geq w_l$; the case $w_h < w_l$ follows by interchanging h and l. By Corollary 3, it is enough to compare the boundary menu $(0, 1/2)$ with h-first frontier menus. Every frontier interior menu has a strictly positive lower threshold, so the Bernoulli distribution $(1-\hat{\theta})\delta_0 + \hat{\theta}\delta_1$ is feasible and gives payoff $(w_h + w_l)\hat{\theta}$. Hence every frontier interior menu has payoff guarantee at most $(w_h + w_l)\hat{\theta}$.

The boundary value is $w_h + w_l \underline{T}(1/2)$. If $\hat{\theta} \leq 1/2$, then $w_h + w_l \underline{T}(1/2) \geq w_h \geq (w_h + w_l)/2 \geq (w_h + w_l)\hat{\theta}$. If $\hat{\theta} > 1/2$, then Proposition 4 gives $\underline{T}(1/2) \geq 2\hat{\theta} - 1$, and therefore

$$w_h + w_l \underline{T}(1/2) - (w_h + w_l)\hat{\theta} \geq (w_h - w_l)(1 - \hat{\theta}) \geq 0.$$

Thus $(0, 1/2)$ weakly dominates every frontier interior menu. Together with Proposition 4, it is an optimal menu when $w_h \geq w_l$; the symmetric boundary menu is optimal when $w_h < w_l$. □

B.3 Proof of Proposition 6

Proof of Proposition 6. Part (i) follows from set inclusion: $\underline{\sigma}_1^2 \leq \underline{\sigma}_2^2$ implies $\Pi(\hat{\theta}, \underline{\sigma}_2^2) \subseteq \Pi(\hat{\theta}, \underline{\sigma}_1^2)$, and taking the infimum of a fixed integrand over the smaller set cannot decrease its value; the supremum over menus preserves the inequality. For Part (ii), on the variance-binding term Proposition 4 gives $\underline{T}(1/2) = 2(\underline{\sigma}^2 - \hat{\theta}(1/2 - \hat{\theta}))$ with $\underline{\sigma}^2$-derivative 2, so by Proposition 4 and Proposition 5 the optimal value has derivative $2\min\{w_h, w_l\} > 0$. □

B.4 Proof of Proposition 7

Proof of Proposition 7. At $t = 1/2$, Proposition 4 gives

$$\underline{T}(1/2) = \max\{2\max\{\hat{\theta} - 1/2, 0\}, 2(\underline{\sigma}^2 - \hat{\theta}(1/2 - \hat{\theta})), 0\}.$$

By Proposition 5, the optimal value equals $\max\{w_h, w_l\} + \min\{w_h, w_l\}\underline{T}(1/2)$ throughout. Since $\min\{w_h, w_l\} > 0$, the optimal value has the same monotonicity and minimizers as $\underline{T}(1/2)$.

The variance-binding term has derivative $4\hat{\theta} - 1$. The feasible mean interval has lower endpoint $(1-\sqrt{1-4\underline{\sigma}^2})/2$, which is below $1/4$ exactly when $\underline{\sigma}^2 < 3/16$; hence the decreasing portion exists exactly in that case. At $\hat{\theta} = 1/4$, the variance-binding term equals $2\underline{\sigma}^2 - 1/8$. Therefore the minimum is $\max\{w_h, w_l\}$ if $\underline{\sigma}^2 \leq 1/16$, is $\max\{w_h, w_l\} + \min\{w_h, w_l\}(2\underline{\sigma}^2 - 1/8)$ if $1/16 < \underline{\sigma}^2 < 3/16$, and the value is strictly increasing in $\hat{\theta}$ when $\underline{\sigma}^2 \geq 3/16$. □

B.5 Derivation of the saddle-point condition

Suppose $w_h \geq w_l$, so $M_{h\text{-bd}} = (0, 1/2)$ is robust optimal. By Proposition 4(i), the payoff guarantee $w_h + w_l\,\underline{T}(1/2)$ is attained by a distribution F^* supported on $\{0, 1/2, 1\}$ with probability masses

$$p_0 = 1 - 2\hat{\theta} + \underline{T}(1/2), \qquad p_{1/2} = 2\big(\hat{\theta} - \underline{T}(1/2)\big), \qquad p_1 = \underline{T}(1/2).$$

Specializing Proposition A.5 to $S = \{h, l\}$, the boundary menu is benchmark-optimal at F^* iff every h-first frontier profile $(x, \theta_l(x))$, $x \in (0, 1/4]$, is non-improving. Along the h-first frontier at $\tau = 1/2$,

$$\tilde{v}(\theta; x, \theta_l(x)) - \tilde{v}(\theta; 0, 1/2) = w_l \mathbf{1}\{\theta_l(x) < \theta \leq 1/2\} - w_h \mathbf{1}\{0 \leq \theta \leq x\}.$$

The least favorable distribution has support $\{0, 1/2, 1\}$. Hence every $x \in (0, 1/4]$ is non-improving iff

$$w_l \cdot 2\big(\hat{\theta} - \underline{T}(1/2)\big) \leq w_h\big(1 - 2\hat{\theta} + \underline{T}(1/2)\big). \tag{B.1}$$

The reverse strict inequality gives a strictly profitable frontier deviation. The case $w_l > w_h$ is symmetric.

B.6 Boundary optimality under general action thresholds

The political-narrative application of the main paper takes $\tau = 1/2$, where a boundary menu is globally optimal (Proposition 5) and the optimal-value comparative statics of Propositions 6 and 7 apply throughout. At arbitrary $\tau \in (0, 1)$, the tail-mass identity (14) of Proposition 4 continues to determine the payoff of any boundary menu. The following proposition therefore does not claim a complete general-τ characterization. Instead, it gives primitive sufficient conditions for boundary optimality and records the comparative statics that remain sharp on those regions. In particular, the low-mean negative effect of raising the average prior survives for every τ because it lies inside an automatically boundary-optimal region.

Proposition B.1 (Boundary optimality under general action thresholds)
Fix $\tau \in (0, 1)$, $0 < \hat{\theta} < 1$, and $0 < \underline{\sigma}^2 < \hat{\theta}(1 - \hat{\theta})$, with the mean-variance ambiguity set $\Pi(\hat{\theta}, \underline{\sigma}^2)$ of Proposition 4. Let $\underline{T}(\tau)$ denote the value of (14) at $t = \tau$.

(i) *The boundary menus have payoff guarantees*

$$V_h^{\text{bd}} = w_h + w_l\,\underline{T}(\tau), \qquad V_l^{\text{bd}} = w_l + w_h\,\underline{T}(\tau).$$

(ii) *Suppose $w_h \geq w_l$. The boundary menu $M_{h\text{-bd}}$ is optimal whenever*

$$\frac{w_h}{w_l} \geq \frac{\min\{\hat{\theta}, \tau\}}{1 - \min\{\hat{\theta}, \tau\}}. \tag{B.2}$$

If (B.2) fails and the variance piece of $\underline{T}(\tau)$ is active, $M_{h\text{-bd}}$ is still optimal whenever

$$\underline{\sigma}^2 \geq (1 - \hat{\theta})\Big[\hat{\theta} - (1 - \tau)\frac{w_h}{w_l}\Big]. \tag{B.3}$$

In particular, when $\hat{\theta} \leq 1/2$ a boundary menu is optimal for every feasible $\underline{\sigma}^2$: $M_{h\text{-bd}}$ if $w_h \geq w_l$, $M_{l\text{-bd}}$ if $w_l > w_h$.

(iii) *The optimal value is weakly increasing in $\underline{\sigma}^2$. On any boundary-optimal region where the variance piece of $\underline{T}(\tau)$ is active, it is strictly increasing with derivative $\min\{w_h, w_l\}/(1 - \tau)$.*

(iv) *Let $C := \max\{w_h, w_l\}$ and $m := \min\{w_h, w_l\}$. At every feasible point satisfying*

$$\hat{\theta} < \frac{\tau}{2} \qquad \text{and} \qquad \underline{\sigma}^2 > \hat{\theta}(\tau - \hat{\theta}), \tag{B.4}$$

the relevant boundary menu is optimal, the variance piece of $\underline{T}(\tau)$ is active, and the optimal value satisfies

$$V^* = C + \frac{m\big(\underline{\sigma}^2 - \hat{\theta}(\tau - \hat{\theta})\big)}{1 - \tau}, \qquad \frac{\partial V^*}{\partial \hat{\theta}} = \frac{m(2\hat{\theta} - \tau)}{1 - \tau} < 0.$$

Thus raising the average prior lowers the actual robust value on this low-mean variance-binding region.

For fixed $\underline{\sigma}^2$, this decreasing region is nonempty if and only if

$$\underline{\sigma}^2 < \frac{\tau(2 - \tau)}{4}. \tag{B.5}$$

More precisely, with $\hat{\theta}_{\text{lo}} := \frac{1 - \sqrt{1 - 4\underline{\sigma}^2}}{2}$, the optimal value is strictly decreasing on

$$\Big(\hat{\theta}_{\text{lo}},\, \bar{\theta}_\tau(\underline{\sigma}^2)\Big), \qquad \bar{\theta}_\tau(\underline{\sigma}^2) := \begin{cases} \dfrac{\tau - \sqrt{\tau^2 - 4\underline{\sigma}^2}}{2}, & \underline{\sigma}^2 < \tau^2/4, \\[2ex] \tau/2, & \underline{\sigma}^2 \geq \tau^2/4. \end{cases}$$

At $\tau = 1/2$, Proposition 5 makes the boundary-optimal region cover the entire feasible mean range, so (B.5) becomes $\underline{\sigma}^2 < 3/16$ with turning point $\hat{\theta} = 1/4$, and Proposition 7 is recovered.

Proof. Part (i). The boundary menu $(0, \tau)$ yields per-type payoff $\bar{u}_0 + w_h + w_l \mathbf{1}\{\theta > \tau\}$ under the lower semi-continuous envelope, so its payoff guarantee over $\Pi(\hat{\theta}, \underline{\sigma}^2)$ is $w_h + w_l \underline{T}(\tau)$ by Proposition 4. The formula for V_l^{bd} follows by interchanging h and l. Since $\underline{T}(\tau) \leq 1$, $V_h^{\text{bd}} - V_l^{\text{bd}} = (w_h - w_l)(1 - \underline{T}(\tau)) \geq 0$ when $w_h \geq w_l$.

Part (ii). We prove the case $w_h \geq w_l$; the case $w_l > w_h$ is symmetric. By Corollary 3, it suffices to compare $M_{h\text{-bd}}$ with h-first frontier interior menus. Every interior menu has

positive lower threshold, so the Bernoulli distribution $(1-\hat{\theta})\delta_0+\hat{\theta}\delta_1$ lies in $\Pi(\hat{\theta},\underline{\sigma}^2)$ (variance $\hat{\theta}(1-\hat{\theta})>\underline{\sigma}^2$) and gives expected payoff $(w_h+w_l)\hat{\theta}$, bounding the payoff guarantee of every interior menu from above by $(w_h+w_l)\hat{\theta}$. Thus $M_{h\text{-bd}}$ is optimal whenever

$$w_h + w_l\,\underline{T}(\tau) \geq (w_h+w_l)\hat{\theta}. \tag{B.6}$$

If $\hat{\theta}\leq\tau$, (B.6) simplifies via $\underline{T}(\tau)\geq 0$ to $w_h\geq(w_h+w_l)\hat{\theta}$, equivalent to $w_h/w_l\geq\hat{\theta}/(1-\hat{\theta})$. If $\hat{\theta}>\tau$, (14) gives $\underline{T}(\tau)\geq(\hat{\theta}-\tau)/(1-\tau)$, and substituting yields $w_h/w_l\geq\tau/(1-\tau)$ for (B.6). The two cases combine into (B.2).

When (B.2) fails and the variance piece of $\underline{T}(\tau)$ is active, $\underline{T}(\tau)=(\underline{\sigma}^2-\hat{\theta}(\tau-\hat{\theta}))/(1-\tau)$, and substituting into (B.6) and using $\hat{\theta}(\tau-\hat{\theta})+(1-\tau)\hat{\theta}=\hat{\theta}(1-\hat{\theta})$ yields (B.3). For $\hat{\theta}\leq 1/2$, $\min\{\hat{\theta},\tau\}\leq 1/2$ gives $\min\{\hat{\theta},\tau\}/(1-\min\{\hat{\theta},\tau\})\leq 1\leq w_h/w_l$, so (B.2) holds; the case $w_l>w_h$ gives $M_{l\text{-bd}}$ symmetrically.

Part (iii). Weak monotonicity in $\underline{\sigma}^2$ follows from set inclusion: $\underline{\sigma}_1^2\leq\underline{\sigma}_2^2$ gives $\Pi(\hat{\theta},\underline{\sigma}_2^2)\subseteq\Pi(\hat{\theta},\underline{\sigma}_1^2)$, so every menu's payoff guarantee is weakly larger at $\underline{\sigma}_2^2$, and the supremum over menus preserves the inequality. On the variance-active boundary-optimal region, differentiating part (i) gives $\partial_{\underline{\sigma}^2}V_h^{\text{bd}}=w_l/(1-\tau)$ and $\partial_{\underline{\sigma}^2}V_l^{\text{bd}}=w_h/(1-\tau)$; whichever boundary is optimal, the active coefficient is $\min\{w_h,w_l\}$.

Part (iv). If (B.4) holds, then $\hat{\theta}<\tau/2<1/2$, so part (ii) implies boundary optimality for the boundary menu associated with the larger payoff weight. Since $\hat{\theta}<\tau$, the mean-forced term in (14) is zero, and the strict inequality $\underline{\sigma}^2>\hat{\theta}(\tau-\hat{\theta})$ makes the variance term active:

$$\underline{T}(\tau)=\frac{\underline{\sigma}^2-\hat{\theta}(\tau-\hat{\theta})}{1-\tau}.$$

Substituting this expression into the boundary value in part (i) gives the displayed value with coefficient $m=\min\{w_h,w_l\}$, and differentiating gives $m(2\hat{\theta}-\tau)/(1-\tau)<0$.

For fixed $\underline{\sigma}^2$, feasibility requires $\hat{\theta}\in(\hat{\theta}_{\text{lo}},\hat{\theta}_{\text{hi}})$, where

$$\hat{\theta}_{\text{lo}}=\frac{1-\sqrt{1-4\underline{\sigma}^2}}{2},\qquad \hat{\theta}_{\text{hi}}=\frac{1+\sqrt{1-4\underline{\sigma}^2}}{2}.$$

A feasible point satisfying $\hat{\theta}<\tau/2$ exists if and only if $\hat{\theta}_{\text{lo}}<\tau/2$, which is equivalent to (B.5). On $[0,\tau/2]$, the function $\hat{\theta}(\tau-\hat{\theta})$ is increasing. Hence the additional variance-activity requirement $\underline{\sigma}^2>\hat{\theta}(\tau-\hat{\theta})$ holds exactly below $(\tau-\sqrt{\tau^2-4\underline{\sigma}^2})/2$ when $\underline{\sigma}^2<\tau^2/4$, and holds throughout $\hat{\theta}<\tau/2$ when $\underline{\sigma}^2\geq\tau^2/4$. This yields the interval displayed in the statement. The specialization $\tau=1/2$ gives $\tau(2-\tau)/4=3/16$ and turning point $\hat{\theta}=1/4$. □

Remark on complete general-τ characterization. Outside the sufficient boundary-optimality regions above, frontier-interior menus may dominate. When $w_h\geq w_l$, the h-first frontier can be parameterized by $x=\theta_h\in[0,\tau/2]$ with

$$\theta_l(x)=\frac{\tau-(1+\tau)x}{1-2x}.$$

For each fixed $x>0$, Nature's problem is a finite moment problem over the four endpoints $\{0,x,\theta_l(x),1\}$. The active support, and hence the formula for the inner value, changes with

$(\hat{\theta}, \underline{\sigma}^2, x, w_h/w_l)$. A complete primitive classification therefore requires comparing several frontier-interior cases against the boundary value. This is why the proposition records sufficient conditions and the low-mean comparative static rather than presenting a full general-τ menu classification.

C Product reviews and an ambiguous consumer prior

This application reads the model through the single-receiver interpretation introduced in Section 2. A seller wants a consumer to buy a product ($a = 1$), while the consumer buys only when quality is high ($\omega = H$). The consumer's prior θ that quality is high is privately known, and the seller knows only that it is drawn from a distribution F whose mean lies in an interval $[\underline{\theta}, \overline{\theta}]$. Public reviews generate h or l, with $0 < \pi_L < \pi_H < 1$. I normalize the seller's utility from no purchase to 0, so $\bar{u}_0 = 0$, and let w_h and w_l denote her expected revenue from a purchase induced after a positive and a negative review.

Let $0 < \underline{\theta} \leq \overline{\theta} < 1$ and define

$$\Pi([\underline{\theta}, \overline{\theta}]) := \left\{F \in \Delta([0,1]) : \underline{\theta} \leq \mathbb{E}_F[\theta] \leq \overline{\theta}\right\}.$$

The question is which menu achieves the highest payoff guarantee: a boundary menu targeting one realization, or an interior menu that spreads persuasion across both realizations.

Proposition C.1 (Menu selection under interval mean ambiguity)
Assume $0 < \underline{\theta} \leq \overline{\theta} < 1$ and $\tau \in (0,1)$.

(i) *For every menu M,*

$$\min_{F \in \Pi([\underline{\theta}, \overline{\theta}])} \int \tilde{v}(\theta; M)\, dF = \min_{\substack{F \in \Delta([0,1]) \\ \mathbb{E}_F[\theta] = \underline{\theta}}} \int \tilde{v}(\theta; M)\, dF.$$

Thus the interval-mean robust problem coincides with the exact-mean problem at the lower endpoint $\underline{\theta}$. This is the specialization of Proposition 2 to $g(\theta) = \theta$, $Y = [\underline{\theta}, \overline{\theta}]$: the non-decreasing per-type payoff $\tilde{v}(\cdot; M)$ forces the worst-case mean to the lower endpoint.

(ii) *Suppose $w_h \geq w_l$. The boundary menu $M_{h\text{-bd}}$ with threshold pair $(0, \tau)$ is optimal iff*

$$\frac{w_h}{w_h + w_l} \geq \begin{cases} 2\underline{\theta} - \tau, & \underline{\theta} \leq \tau, \\ \tau, & \underline{\theta} > \tau. \end{cases}$$

When it is optimal, its value is $w_h + w_l \max\{(\underline{\theta} - \tau)/(1 - \tau), 0\}$. When this condition fails, an optimal h-first frontier interior menu has threshold pair

$$\theta_h = \frac{\tau - \frac{w_h}{w_h + w_l}}{2\left(1 - \frac{w_h}{w_h + w_l}\right)}, \qquad \theta_l = \frac{\tau + \frac{w_h}{w_h + w_l}}{2},$$

and its value is $(w_h + w_l)(\underline{\theta} - \theta_h)/(1 - \theta_h)$.

(iii) *If $w_l \geq w_h$, the symmetric statement holds after interchanging h and l together with w_h and w_l: $M_{l\text{-bd}}$ replaces $M_{h\text{-bd}}$, the cutoff uses $w_l/(w_h+w_l)$, when it is optimal its value is $w_l + w_h \max\{(\underline{\theta}-\tau)/(1-\tau), 0\}$, and the optimal interior menu is the corresponding l-first frontier menu with value obtained from* (ii) *by exchanging w_h and w_l.*

Part (i) follows because $\tilde{v}(\cdot; M)$ is non-decreasing in type for every menu M. The seller is rewarded only by what the data rules out from below: tightening $\overline{\theta}$ has no value here, while raising $\underline{\theta}$ can raise the guarantee. Parts (ii) and (iii) compare concentrating persuasion at one endpoint with spreading it along the frontier. A boundary menu is optimal when the insured realization is valuable enough, or when the lower-endpoint mean is low enough that Nature can keep little mass above the other threshold. When the condition fails, the seller uses both review realizations: lowering the less protected threshold is worth raising the fully insured one, and the optimal interior pair is where the one-moment lower envelope switches from the middle step to the chord that reaches the top step (see Figure 2).

Proof of Proposition C.1. For every menu M, the lower semi-continuous payoff $\tilde{v}(\theta; M)$ is non-decreasing in θ. Take any $F \in \Pi([\underline{\theta}, \overline{\theta}])$ with $\mathbb{E}_F[\theta] > \underline{\theta}$ and replace it by $(\underline{\theta}/\mathbb{E}_F[\theta])F + (1 - \underline{\theta}/\mathbb{E}_F[\theta])\delta_0$. The new distribution has mean $\underline{\theta}$ and remains in $\Pi([\underline{\theta}, \overline{\theta}])$. Since $\tilde{v}(0; M) \leq \tilde{v}(\theta; M)$ for every θ, this replacement weakly lowers the expected payoff. Hence it is without loss to restrict Nature's minimization to distributions with mean $\underline{\theta}$. The reverse inequality is immediate because the exact-mean set is contained in $\Pi([\underline{\theta}, \overline{\theta}])$.

This proves part (i). For parts (ii) and (iii), we prove the case $w_h \geq w_l$; the case $w_l \geq w_h$ follows by interchanging h and l together with w_h and w_l.

By Corollary 3, it is enough to optimize over the h-first frontier. At a frontier point (θ_h, θ_l) with $0 < \theta_h < \theta_l \leq \tau$, the one-moment convexification of the three-step per-type payoff (the one-dimensional exact-mean specialization of Proposition 2) gives the following exact-mean value. If $(\theta_l - \theta_h)/(1 - \theta_h) \geq w_h/(w_h + w_l)$, the middle step lies on the lower convex envelope: the value is 0 for $\underline{\theta} \leq \theta_h$, is $w_h(\underline{\theta} - \theta_h)/(\theta_l - \theta_h)$ for $\theta_h < \underline{\theta} \leq \theta_l$, and is $w_h + w_l(\underline{\theta} - \theta_l)/(1 - \theta_l)$ for $\underline{\theta} > \theta_l$. If $(\theta_l - \theta_h)/(1 - \theta_h) \leq w_h/(w_h + w_l)$, the middle step is bypassed by the chord from $(\theta_h, 0)$ to $(1, w_h + w_l)$: the value is 0 for $\underline{\theta} \leq \theta_h$ and is $(w_h + w_l)(\underline{\theta} - \theta_h)/(1 - \theta_h)$ for $\underline{\theta} > \theta_h$. At equality the two descriptions agree.

Along the frontier, $(\theta_l - \theta_h)/(1 - \theta_h)$ decreases from τ at the boundary to 0 at $\theta_h = \theta_l = \tau/2$, and the equality $(\theta_l - \theta_h)/(1 - \theta_h) = w_h/(w_h + w_l)$ has the displayed interior solution whenever $w_h/(w_h + w_l) < \tau$. The boundary value at $(0, \tau)$ is $w_h + w_l \max\{(\underline{\theta} - \tau)/(1 - \tau), 0\}$. Comparing it with the frontier formula gives the boundary condition in part (ii): for $\underline{\theta} \leq \tau$ it is $w_h/(w_h + w_l) \geq 2\underline{\theta} - \tau$, and for $\underline{\theta} > \tau$ it is $w_h/(w_h + w_l) \geq \tau$. When the condition fails, the frontier value is maximized at the kink above, giving the displayed pair and value. □

Corollary C.1 (Comparative statics under interval mean ambiguity)
Assume $0 < \underline{\theta} \leq \overline{\theta} < 1$ and $\tau \in (0, 1)$.

(i) *The optimal value and the optimal-menu set do not depend on $\overline{\theta}$.*

(ii) *The optimal value is weakly increasing in $\underline{\theta}$.*

(iii) *Suppose $w_h \geq w_l$. If $w_h/(w_h + w_l) \geq \tau$, then $M_{h\text{-bd}}$ is optimal for every $\underline{\theta}$. If $w_h/(w_h + w_l) < \tau$, then $M_{h\text{-bd}}$ is optimal exactly when $\underline{\theta} \leq (w_h/(w_h + w_l) + \tau)/2$, and an interior h-first frontier menu is optimal above that cutoff. The case $w_l \geq w_h$ is symmetric.*

(iv) *In the h-first interior region, increasing* $w_h/(w_h + w_l)$ *lowers* θ_h *and raises* θ_l*, while increasing* τ *raises both thresholds. The l-first interior region is symmetric.*

Corollary C.1 describes how the lower mean bound and payoff weights move the robust menu. Raising $\underline{\theta}$ weakly improves the payoff guarantee because Nature's least favorable prior distribution must be more optimistic about quality. If the larger payoff share is at least τ, the seller insures the high-payoff review realization for every $\underline{\theta}$; otherwise this boundary menu remains optimal only at low $\underline{\theta}$, and the seller moves into the frontier interior once the data rules out sufficiently pessimistic prior distributions. In the interior region, increasing the payoff share of the more valuable realization lowers that realization's threshold and raises the other, while a larger τ raises both thresholds.

Proof of Corollary C.1. Part (i) follows from Proposition C.1(i): for each menu M, the payoff guarantee equals the exact-mean problem at $\underline{\theta}$, which contains no $\overline{\theta}$.

For part (ii), take $0 < \underline{\theta}_1 < \underline{\theta}_2 < 1$. For any menu M and any $F \in \Delta([0,1])$ with $\mathbb{E}_F[\theta] = \underline{\theta}_2$, the distribution $(\underline{\theta}_1/\underline{\theta}_2)F + (1 - \underline{\theta}_1/\underline{\theta}_2)\delta_0$ has mean $\underline{\theta}_1$. Since $\tilde{v}(\theta; M)$ is non-decreasing in θ, the expected payoff under this new distribution is weakly below the expected payoff under F. Hence the exact-mean payoff guarantee for each fixed menu is weakly increasing in $\underline{\theta}$. Maximizing over menus preserves the inequality.

For part (iii), suppose $w_h \geq w_l$. The condition in Proposition C.1(ii) is equivalent to $\underline{\theta} \leq (w_h/(w_h+w_l)+\tau)/2$ on $\underline{\theta} \leq \tau$ and to $w_h/(w_h+w_l) \geq \tau$ on $\underline{\theta} > \tau$. When $w_h/(w_h+w_l) \geq \tau$, the boundary optimality condition is satisfied for every $\underline{\theta}$, so $M_{h\text{-bd}}$ is always optimal. When $w_h/(w_h + w_l) < \tau$, $(w_h/(w_h + w_l) + \tau)/2 < \tau$ pins down the boundary-optimality range; beyond it, Proposition C.1(ii) gives the h-first frontier interior menu.

For part (iv), in the h-first interior region the thresholds are those displayed in Proposition C.1(ii). Differentiating with respect to $w_h/(w_h+w_l)$ gives $(\tau-1)/(2(1-w_h/(w_h+w_l))^2) < 0$ for θ_h and $1/2 > 0$ for θ_l. Differentiating with respect to τ gives $1/(2(1-w_h/(w_h+w_l))) > 0$ for θ_h and $1/2 > 0$ for θ_l. □

This differs from Carrasco et al. (2018), where a first-moment restriction alone does not guarantee strictly positive worst-case revenue. Here, because the receiver's type lies in the compact set $[0,1]$ and payoffs are bounded step functions of persuasion thresholds, the same kind of first-moment information limits how much mass Nature can move below the relevant thresholds and can therefore secure a strictly positive payoff above $\bar{u}_0$.

References

Aina, Chiara, "Tailored Stories," 2025. Preprint, Universitat Pompeu Fabra.

Aliprantis, Charalambos D. and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed., Springer, 2006.

Alonso, Ricardo and Odilon Câmara, "Persuading Voters," *American Economic Review*, 2016, *106* (11), 3590–3605.

Alpern, Steve and Shmuel Gal, "A Mixed-Strategy Minimax Theorem without Compactness," *SIAM Journal on Control and Optimization*, 1988, *26* (6), 1357–1361.

Babichenko, Yakov, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi, "Regret-minimizing Bayesian persuasion," *Games and Economic Behavior*, 2022, *136*, 226–248.

Ball, Ian and Deniz Kattwinkel, "Robust Robustness," 2026. Preprint, arXiv:2408.16898v4 (4 May 2026).

Bardhi, Arjada and Yingni Guo, "Modes of persuasion toward unanimous consent," *Theoretical Economics*, 2018, *13* (3), 1111–1150.

Bergemann, Dirk and Stephen Morris, "Information Design: A Unified Perspective," *Journal of Economic Literature*, 2019, *57* (1), 44–95.

Best, James and Daniel Quigley, "Persuasion for the Long Run," *Journal of Political Economy*, 2024, *132* (5), 1740–1791.

Carrasco, Vinicius, Vitor Farinha Luz, Nenad Kos, Matthias Messner, Paulo Monteiro, and Humberto Moreira, "Optimal selling mechanisms under moment conditions," *Journal of Economic Theory*, 2018, *177*, 245–279.

_ , _ , Paulo K. Monteiro, and Humberto Moreira, "Robust mechanisms: the curvature case," *Economic Theory*, 2019, *68* (1), 203–222.

Doval, Laura and Alex Smolin, "Persuasion and Welfare," *Journal of Political Economy*, 2024, *132* (7), 2451–2487.

Dworczak, Piotr and Alessandro Pavan, "Preparing for the Worst but Hoping for the Best: Robust (Bayesian) Persuasion," *Econometrica*, 2022, *90* (5), 2017–2051.

Edmond, Chris, "Information Manipulation, Coordination, and Regime Change," *Review of Economic Studies*, 2013, *80* (4), 1422–1458.

Eliaz, Kfir and Ran Spiegler, "A Model of Competing Narratives," *American Economic Review*, 2020, *110* (12), 3786–3816.

_ , _ , and Yair Weiss, "Cheating with Models," *American Economic Review: Insights*, 2021, *3* (4), 417–434.

Fréchette, Guillaume R., Alessandro Lizzeri, and Jacopo Perego, "Rules and Commitment in Communication: An Experimental Analysis," *Econometrica*, 2022, *90* (5), 2283–2318.

Gitmez, A. Arda and Pooya Molavi, "Informational Autocrats, Diverse Societies," 2022. Preprint, arXiv:2203.12698; CEPR Discussion Paper No. 18903.

Guriev, Sergei and Daniel Treisman, "Informational Autocrats," *Journal of Economic Perspectives*, 2019, *33* (4), 100–127.

Hu, Ju and Xi Weng, "Robust persuasion of a privately informed receiver," *Economic Theory*, 2021, *72* (3), 909–953.

Imbens, Guido W. and Charles F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2004, *72* (6), 1845–1857.

Inostroza, Nicolas A. and Alessandro Pavan, "Adversarial coordination and public information design," *Theoretical Economics*, 2025, *20* (2), 763–813.

Jain, Atulya, "Informing agents amidst biased narratives," *Job Mark. Pap., HEC Paris, Paris*, 2023.

Kamenica, Emir, "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 2019, *11* (1), 249–272.

_ **and Matthew Gentzkow**, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

King, Gary, Jennifer Pan, and Margaret E. Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, 2013, *107* (2), 326–343.

Kolotilin, Anton and Andriy Zapechelnyuk, "Persuasion meets delegation," *Econometrica*, 2025, *93* (1), 195–228.

_ **, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li**, "Persuasion of a Privately Informed Receiver," *Econometrica*, 2017, *85* (6), 1949–1964.

Kosterina, Svetlana, "Persuasion with unknown beliefs," *Theoretical Economics*, 2022, *17* (3), 1075–1107.

Kuran, Timur, "Now Out of Never: The Element of Surprise in the East European Revolution of 1989," *World Politics*, 1991, *44* (1), 7–48.

Laclau, Marie and Ludovic Renou, "Public Persuasion," 2016. Preprint, PSE Working Paper, HAL hal-01285218.

Lin, Xiao and Ce Liu, "Credible Persuasion," *Journal of Political Economy*, 2024, *132* (7), 2228–2273.

Manski, Charles F., *Partial Identification of Probability Distributions* Springer Series in Statistics, New York, NY: Springer, 2003.

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

Molchanov, Ilya and Francesca Molinari, *Random Sets in Econometrics*, Vol. 60 of *Econometric Society Monographs*, Cambridge University Press, 2018.

Parakhonyak, Alexei and Anton Sobolev, "Persuasion without Priors," *Economic Theory*, 2026, *81* (3), 625–659.

Rayo, Luis and Ilya Segal, "Optimal Information Disclosure," *Journal of Political Economy*, 2010, *118* (5), 949–987.

Rockafellar, R. Tyrrell, *Convex Analysis* number 28. In 'Princeton Mathematical Series.', Princeton, NJ: Princeton University Press, 1970.

Rosenthal, Maxwell, "Data-Driven Persuasion," *Journal of Mathematical Economics*, August 2026, *125.*

Rozenas, Arturas and Denis Stukal, "How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television," *Journal of Politics*, 2019, *81* (3), 982–996.

Sapiro-Gheiler, Eitan, "Persuasion with ambiguous receiver preferences," *Economic Theory*, 2024, *77* (4), 1173–1218.

Schwartzstein, Joshua and Adi Sunderam, "Using Models to Persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Spiegler, Ran, "Bayesian Networks and Boundedly Rational Expectations," *Quarterly Journal of Economics*, 2016, *131* (3), 1243–1290.

Suzdaltsev, Alex, "Distributionally robust pricing in independent private value auctions," *Journal of Economic Theory*, 2022, *206*, 105555.

Szeidl, Adam and Ferenc Szucs, "A model of populism as a conspiracy theory," *American Economic Review*, 2025, *115* (9), 3214–3247.

Vohra, Akhil, Juuso Toikka, and Rakesh Vohra, "Bayesian persuasion: Reduced form approach," *Journal of Mathematical Economics*, 2023, *107*, 102863.

Winkler, Gerhard, "Extreme Points of Moment Sets," *Mathematics of Operations Research*, 1988, *13* (4), 581–587.