Robust Narrative Persuasion

Song Li
Universitat Autònoma de Barcelona | Barcelona School of Economics

Working Paper • July 2026

Motivation

The research problem

Environment

One sender faces receivers with heterogeneous priors. A fixed information structure generates a public realization s. The sender does not choose the information; she chooses a public menu of likelihood narratives for interpreting s. The population distribution of priors is not known: aggregate evidence only implies $F \in \Pi$.

The research problem

Environment

One sender faces receivers with heterogeneous priors. A fixed information structure generates a public realization s. The sender does not choose the information; she chooses a public menu of likelihood narratives for interpreting s. The population distribution of priors is not known: aggregate evidence only implies $F \in \Pi$.

Research question

Which public menu of likelihood narratives maximizes the sender's worst-case payoff over $F \in \Pi$, when each receiver selects the narrative that best fits his prior?

The same public event, different interpretations

The same public event, different interpretations

Motivation

The same public event can support different interpretations.
People choose whichever interpretation best explains what they observe, given what they already believe.

▸ HK mapping ▸ Menu of narratives ▸ More examples

From aggregate evidence to robust narrative design

Persuasion works receiver by receiver: after each event, the payoff is driven by the people who are nearly won over, so the sender's decisive unknown is the cross-sectional distribution of prior beliefs F.

From aggregate evidence to robust narrative design

Persuasion works receiver by receiver: after each event, the payoff is driven by the people who are nearly won over, so the sender's decisive unknown is the cross-sectional distribution of prior beliefs F.

What the sender actually observes

- an average, as a point or a range: polls, turnout, ratings;
- dispersion, or higher moments: polarization, review spread;
- **not** the full distribution F.

From aggregate evidence to robust narrative design

Persuasion works receiver by receiver: after each event, the payoff is driven by the people who are nearly won over, so the sender's decisive unknown is the cross-sectional distribution of prior beliefs F.

What the sender actually observes

- an average, as a point or a range: polls, turnout, ratings;
- dispersion, or higher moments: polarization, review spread;
- **not** the full distribution F.

The ambiguity set

- a **set** of distributions are consistent with the moments;
- maxmin approach: the worst case;
- the payoff secured there is the menu's **payoff guarantee**.

From aggregate evidence to robust narrative design

Persuasion works receiver by receiver: after each event, the payoff is driven by the people who are nearly won over, so the sender's decisive unknown is the cross-sectional distribution of prior beliefs F.

What the sender actually observes

- an average, as a point or a range: polls, turnout, ratings;
- dispersion, or higher moments: polarization, review spread;
- **not** the full distribution F.

The ambiguity set

- a **set** of distributions are consistent with the moments;
- maxmin approach: the worst case;
- the payoff secured there is the menu's **payoff guarantee**.

Robust menu design

*Which public menu of narratives maximizes the **payoff guarantee**?*

Moment sets

Contributions and roadmap

Part	Contribution
Model	Public narrative-menu design for a heterogeneous-prior population known through aggregate moments.
Reduction	Fit-based model adoption reduces to implementable persuasion-threshold profiles.
Frontier	A single public menu links persuasion thresholds across signal realizations.
Robustness	Each fixed threshold profile gives a moment problem over the prior distribution.
Application	Boundary optimality, a dispersion-monotone value, and a non-monotone average-prior effect in the dictatorship application.

Contributions and roadmap

Part	Contribution
Model	Public narrative-menu design for a heterogeneous-prior population known through aggregate moments.
Reduction	Fit-based model adoption reduces to implementable persuasion-threshold profiles.
Frontier	A single public menu links persuasion thresholds across signal realizations.
Robustness	Each fixed threshold profile gives a moment problem over the prior distribution.
Application	Boundary optimality, a dispersion-monotone value, and a non-monotone average-prior effect in the dictatorship application.

Main message

Robust narrative design chooses one public menu of likelihood narratives for a heterogeneous population whose prior distribution is only partially known.

Related literature

- **Models as likelihoods / Narrative persuasion:** Schwartzstein & Sunderam (2021); Aina (2025).
  The starting point is likelihood narratives and fit-based adoption. This paper studies one public menu for a heterogeneous population.
- **Public persuasion with heterogeneous priors:** Laclau & Renou (2016); Alonso & Câmara (2016); Inostroza & Pavan (2025); Gitmez & Molavi (2022).
  Those papers design signals for heterogeneous receivers. Here the commitment object is a narrative menu, and priors affect model adoption.
- **Robust persuasion / ambiguity:** Kosterina (2022); Dworczak & Pavan (2022); Ball & Kattwinkel (2026); Carrasco et al. (2018).
  Moment-duality is used to evaluate payoff guarantees generated by the narrative threshold frontier.

Model

A narrative is a likelihood model; receivers self-select

Narrative

A likelihood model $m : \Omega \to \Delta(S)$ assigning probabilities to realizations in each state.

A narrative is a likelihood model; receivers self-select

Narrative

A likelihood model $m : \Omega \to \Delta(S)$ assigning probabilities to realizations in each state.

Selection by fit

$$\mathrm{Fit}(s, m, \theta) = \sum_{\omega \in \Omega} \theta(\omega) m(s \mid \omega).$$

Receivers choose from M the narrative that makes the observed realization most likely under their prior.

A narrative is a likelihood model; receivers self-select

Narrative

A likelihood model $m : \Omega \to \Delta(S)$ assigning probabilities to realizations in each state.

Selection by fit

$$\mathrm{Fit}(s, m, \theta) = \sum_{\omega \in \Omega} \theta(\omega) m(s \mid \omega).$$

Receivers choose from M the narrative that makes the observed realization most likely under their prior.

Why heterogeneity matters

- same s, different selected narratives;
- persuasion is type-by-type;
- mass near the margin drives payoffs.

Setup and timing

Primitives

- States $\Omega = \{H, L\}$; realization $s \in S$.
- Actions $a \in \{0, 1\}$; payoffs U_S, U_R.
- Sender has a state-uniform preferred action: $U_S(1, \omega) > U_S(0, \omega)$.
- Type $\theta = \Pr(H) \in \Theta = [0, 1]$.
- Population of priors: F; sender's prior $\tilde{\mu} \in \Delta(\Omega)$.
- Data-generating model $\pi(\cdot \mid \omega)$ is fixed; with $\tilde{\mu}$, it evaluates payoffs.
- Finite public menu $M \subseteq \mathcal{M}$.

π governs true signal frequencies;
$m \in M$ governs interpretation.

Setup and timing

Primitives

- States $\Omega = \{H, L\}$; realization $s \in S$.
- Actions $a \in \{0, 1\}$; payoffs U_S, U_R.
- Sender has a state-uniform preferred action: $U_S(1, \omega) > U_S(0, \omega)$.
- Type $\theta = \Pr(H) \in \Theta = [0, 1]$.
- Population of priors: F; sender's prior $\tilde{\mu} \in \Delta(\Omega)$.
- Data-generating model $\pi(\cdot \mid \omega)$ is fixed; with $\tilde{\mu}$, it evaluates payoffs.
- Finite public menu $M \subseteq \mathcal{M}$.

π governs true signal frequencies;
$m \in M$ governs interpretation.

Timing

1. Types are distributed according to F
2. Sender commits to one public menu M
3. Nature draws $\omega \sim \tilde{\mu}$, then $s \sim \pi(\cdot \mid \omega)$
4. Each type selects the best-fitting narrative $m^*(s, \theta)$
5. Receiver updates within m^* and chooses an action

Fit-based selection, then Bayes within

Receiver behavior after realization s

$$m^*(s,\theta) \in \arg\max_{m \in M} \sum_{\omega} \theta(\omega) m(s \mid \omega), \qquad \Pr(H \mid s, m^*) = \frac{\theta\, m^*(s \mid H)}{\theta\, m^*(s \mid H) + (1-\theta) m^*(s \mid L)}.$$

Fit-based selection, then Bayes within

Receiver behavior after realization s

$$m^*(s,\theta) \in \arg\max_{m\in M} \sum_{\omega} \theta(\omega) m(s \mid \omega), \qquad \Pr(H \mid s, m^*) = \frac{\theta\, m^*(s \mid H)}{\theta\, m^*(s \mid H) + (1-\theta) m^*(s \mid L)}.$$

Two cutoffs

Action rule: $a = 1 \iff \Pr(H \mid s, m^*) \geq \tau$, where $\tau = \frac{|u_L|}{u_H + |u_L|}$ is the **action threshold**.

θ_s is the **persuasion threshold**: the marginal prior type just persuaded after realization s.

$u_H = U_R(1, H) - U_R(0, H) > 0,$
$u_L = U_R(1, L) - U_R(0, L) < 0.$

Fit-based selection, then Bayes within

Receiver behavior after realization s

$$m^*(s,\theta) \in \arg\max_{m\in M} \sum_{\omega} \theta(\omega) m(s \mid \omega), \qquad \Pr(H \mid s, m^*) = \frac{\theta\, m^*(s \mid H)}{\theta\, m^*(s \mid H) + (1-\theta) m^*(s \mid L)}.$$

Two cutoffs

Action rule: $a = 1 \iff \Pr(H \mid s, m^*) \geq \tau$, where $\tau = \frac{|u_L|}{u_H + |u_L|}$ is the **action threshold**.

θ_s is the **persuasion threshold**: the marginal prior type just persuaded after realization s.

$u_H = U_R(1, H) - U_R(0, H) > 0,$
$u_L = U_R(1, L) - U_R(0, L) < 0.$

Affine fit and sender-favorable tie-breaking make the persuaded set an **upper interval**.

A public menu can split a heterogeneous crowd after the same realization. The selected m determines the receiver's posterior; the fixed π determines expected payoffs.

Ambiguity and the sender's robust problem

From the model, a menu reduces to a per-type payoff $v(\theta; M)$; the sender does not know the population F, only its moments. The admissible populations form a **continuous moment set**:

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

Ambiguity and the sender's robust problem

From the model, a menu reduces to a per-type payoff $v(\theta; M)$; the sender does not know the population F, only its moments. The admissible populations form a **continuous moment set**:

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

Outer problem

Choose M to maximize

$$V(M) = \inf_{F \in \Pi} \int_\Theta v(\theta; M)\, F(d\theta).$$

Inner problem

For a fixed M, choose F to minimize

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

Ambiguity and the sender's robust problem

From the model, a menu reduces to a per-type payoff $v(\theta; M)$; the sender does not know the population F, only its moments. The admissible populations form a **continuous moment set**:

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

Outer problem

Choose M to maximize

$$V(M) = \inf_{F \in \Pi} \int_\Theta v(\theta; M)\, F(d\theta).$$

Inner problem

For a fixed M, choose F to minimize

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

Two senses of robustness (key assumption: star g-interior Y)

Robust $=$ worst-case over Π *and* stable to perturbing Π. Star g-interior rules out knife-edge moment information: a relative-interior star center makes Π a *rich, continuous* moment set (Ball & Kattwinkel 2026), so small misspecification of (g, Y) cannot drop $V(M)$ discontinuously.

Results

Threshold reduction: one setter per realization

Theorem (Menu reduction [Thm. 1]**)**

For every menu M there is a sub-menu $M' \subseteq M$ with $|M'| \leq |S|$ such that $v(\theta; M') \geq v(\theta; M)$ for every θ; hence

$$\sup_{M \in \mathcal{M}} V(M) = \sup_{\substack{M \in \mathcal{M} \\ |M| \leq |S|}} V(M).$$

Threshold reduction: one setter per realization

Theorem (Menu reduction [Thm. 1]**)**

For every menu M there is a sub-menu $M' \subseteq M$ with $|M'| \leq |S|$ such that $v(\theta; M') \geq v(\theta; M)$ for every θ; hence

$$\sup_{M \in \mathcal{M}} V(M) = \sup_{\substack{M \in \mathcal{M} \\ |M| \leq |S|}} V(M).$$

- The payoff-relevant object is the first type persuaded after each realization.
- Keep, for each realization, one narrative that sets this threshold; it persuades all higher types too.
- Extra narratives can delay persuasion: they win on fit but generate a posterior below τ.
- The robust public-menu problem can therefore be searched over at most $|S|$ threshold-setting narratives.

▸ Proof idea

Implementability is bounded by a persuasion frontier

Menu design reduces to a **threshold profile** $(\theta_s)_{s\in S}$: payoff depends *only* on them, and a lower θ_s persuades more types after s.

Theorem (Implementability [Thm. 2]**)**

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

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

Figure: two-signal illustration, $S=\{h,l\}$, $\tau=1/2$.

Implementability is bounded by a persuasion frontier

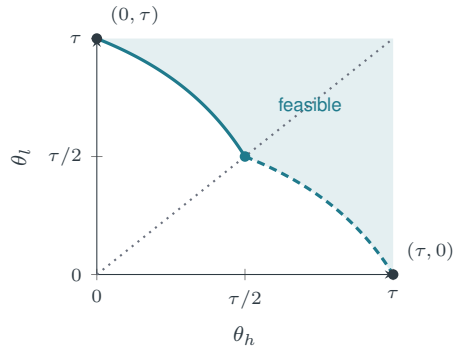
Menu design reduces to a **threshold profile** $(\theta_s)_{s\in S}$: payoff depends *only* on them, and a lower θ_s persuades more types after s.

Theorem (Implementability [Thm. 2]**)**

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

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

Figure: two-signal illustration, $S=\{h,l\}$, $\tau=1/2$.



Implementability is bounded by a persuasion frontier

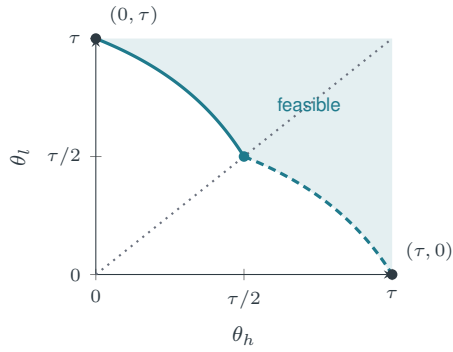
Menu design reduces to a **threshold profile** $(\theta_s)_{s\in S}$: payoff depends *only* on them, and a lower θ_s persuades more types after s.

Theorem (Implementability [Thm. 2]**)**

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

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

Figure: two-signal illustration, $S=\{h,l\}$, $\tau=1/2$.



Each m_s must reach τ at its own marginal pair (s,θ_s) and cannot strictly beat m_t at another marginal pair (t,θ_t). Thus the threshold profile must satisfy a joint feasibility constraint: lowering one threshold raises the burden somewhere else.

▸ Bayes comparison

Simplifying notations

A menu affects the sender only through the **threshold profile** $(\theta_s)_{s\in S}$. Ignoring cutoff atoms, the per-type payoff is a weighted count of the realizations after which type θ is persuaded:

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

For robust evaluation below, positive cutoff atoms are handled by the lower-semicontinuous envelope $\tilde{v}$.

Simplifying notations

A menu affects the sender only through the **threshold profile** $(\theta_s)_{s\in S}$. Ignoring cutoff atoms, the per-type payoff is a weighted count of the realizations after which type θ is persuaded:

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

For robust evaluation below, positive cutoff atoms are handled by the lower-semicontinuous envelope $\tilde{v}$. Each realization's weight w_s assembles the three primitives:

$$w_s = \sum_{\omega\in\{H,L\}} \underbrace{\tilde{\mu}(\omega)}_{\text{sender prior}}\ \underbrace{\pi(s\mid\omega)}_{\substack{\text{data-generating}\\ \text{model}}}\ \underbrace{\big[U_S(1,\omega)-U_S(0,\omega)\big]}_{\text{sender's persuasion gain}} > 0.$$

Simplifying notations

A menu affects the sender only through the **threshold profile** $(\theta_s)_{s\in S}$. Ignoring cutoff atoms, the per-type payoff is a weighted count of the realizations after which type θ is persuaded:

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

For robust evaluation below, positive cutoff atoms are handled by the lower-semicontinuous envelope $\tilde{v}$. Each realization's weight w_s assembles the three primitives:

$$w_s = \sum_{\omega\in\{H,L\}} \underbrace{\tilde{\mu}(\omega)}_{\text{sender prior}} \underbrace{\pi(s\mid\omega)}_{\substack{\text{data-generating}\\ \text{model}}} \underbrace{\big[U_S(1,\omega)-U_S(0,\omega)\big]}_{\text{sender's persuasion gain}} > 0.$$

- w_s is realization s's state-averaged value of moving a type from $a=0$ to $a=1$.
- The data-generating model π sets the *relative* weights across realizations.
- Baseline $\bar{u}_0 = \sum_\omega \tilde{\mu}(\omega)\,U_S(0,\omega)$ is the payoff when no one is taking the sender-preferred action.

Burden sharing in threshold-odds coordinates

Order realizations by payoff weights, $w_{s_1} \geq \cdots \geq w_{s_n}$, and write $r_{s_i} = \theta_{s_i}/(1-\theta_{s_i})$. The sender can restrict attention to sorted profiles on the *linear threshold-burden frontier* [Cor. 2]

$$0 \leq r_{s_1} \leq \cdots \leq r_{s_n}, \qquad \sum_{i=1}^{n-1} r_{s_i} + (1-\tau)\, r_{s_n} = \tau.$$

- Lower thresholds are assigned to higher-payoff realizations.
- Lowering one threshold means some other threshold must carry more burden.

Robustness: fixed menus induce moment problems

For an implementable threshold profile, robust evaluation is a moment problem over F [Prop. 2]. Write $\tilde{v}$ for the lower-semicontinuous payoff: at positive thresholds, cutoff atoms are not counted as persuaded.

Robustness: fixed menus induce moment problems

For an implementable threshold profile, robust evaluation is a moment problem over F [Prop. 2]. Write $\tilde{v}$ for the lower-semicontinuous payoff: at positive thresholds, cutoff atoms are not counted as persuaded.

Under the **star g-interior** set above (with Y closed and convex), strong duality holds and the dual chooses a minorant of the threshold payoff:

$$\gamma + \lambda^\top g(\theta) \leq \tilde{v}(\theta) \qquad \forall \theta \in [0,1].$$

- A least favorable F^* sits on the **contact set**, where the minorant equals $\tilde{v}$.
- With d moments, one can take F^* with **at most $d+1$ atoms**.

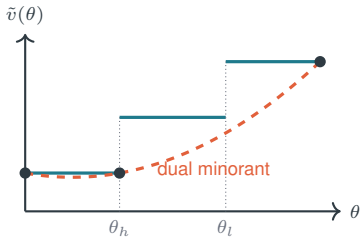
Robustness: fixed menus induce moment problems

For an implementable threshold profile, robust evaluation is a moment problem over F [Prop. 2]. Write $\tilde{v}$ for the lower-semicontinuous payoff: at positive thresholds, cutoff atoms are not counted as persuaded.

Under the **star g-interior** set above (with Y closed and convex), strong duality holds and the dual chooses a minorant of the threshold payoff:

$$\gamma + \lambda^\top g(\theta) \le \tilde{v}(\theta) \qquad \forall \theta \in [0,1].$$

- A least favorable F^* sits on the **contact set**, where the minorant equals $\tilde{v}$.
- With d moments, one can take F^* with **at most $d+1$ atoms**.



Quadratic minorant (mean–variance moments): contact at $\{0, \theta_h, 1\}$; contact types (dots) support a least favorable population.

The finite-support step is the moment-problem implication; the paper uses it for payoff guarantees generated by narrative menus.

▸ LSC payoff ▸ Dual ▸ Convexification

Application

A dictator persuading a polarized public

- A dictator wants support ($a = 1$); citizens support only if she is competent ($\omega = H$).
- Public news yields favorable h or unfavorable l; let $\pi_H = \pi(h \mid H)$ and $\pi_L = \pi(h \mid L)$, with $0 < \pi_L < \pi_H < 1$.
- Action threshold $\tau = 1/2$.
- She observes only the **average prior** $\hat{\theta}$ (rallies, turnout) and a **dispersion lower bound** $\underline{\sigma}^2$ (polarization in posts).

A dictator persuading a polarized public

- A dictator wants support ($a = 1$); citizens support only if she is competent ($\omega = H$).
- Public news yields favorable h or unfavorable l; let $\pi_H = \pi(h \mid H)$ and $\pi_L = \pi(h \mid L)$, with $0 < \pi_L < \pi_H < 1$.
- Action threshold $\tau = 1/2$.
- She observes only the **average prior** $\hat{\theta}$ (rallies, turnout) and a **dispersion lower bound** $\underline{\sigma}^2$ (polarization in posts).

Ambiguity set

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

w_h: convert favorable news into visible support. w_l: suppress dissent after bad news.

$w_l > w_h$ is the crisis-prone regime case: unfavorable news enables coordinated unrest and elite defection (Edmond 2013; King, Pan & Roberts 2013; Guriev & Treisman 2019).

The optimal menu insures one realization

Proposition (Boundary optimality [Prop. 5]**)**

For every $w_h, w_l > 0$, $\hat{\theta} \in (0,1)$, *and* $\underline{\sigma}^2 \in (0, \hat{\theta}(1-\hat{\theta}))$, *a boundary menu is optimal. If* $w_h \geq w_l$, *it is* $(0, \frac{1}{2})$ *with value* $w_h + w_l \underline{T}(\frac{1}{2})$, *where* $\underline{T}(t) := \inf_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} F((t,1])$ *is the strict upper-tail guarantee; the case* $w_l > w_h$ *is symmetric.*

The optimal menu insures one realization

Proposition (Boundary optimality [Prop. 5]**)**

For every $w_h, w_l > 0$, $\hat{\theta} \in (0,1)$, *and* $\underline{\sigma}^2 \in (0, \hat{\theta}(1-\hat{\theta}))$, *a boundary menu is optimal. If* $w_h \geq w_l$, *it is* $(0, \frac{1}{2})$ *with value* $w_h + w_l \underline{T}(\frac{1}{2})$, *where* $\underline{T}(t) := \inf_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} F((t,1])$ *is the strict upper-tail guarantee; the case* $w_l > w_h$ *is symmetric.*

After favorable news h

$\theta_h = 0$: every citizen is persuaded. The higher-payoff realization is **fully insured**.

The optimal menu insures one realization

Proposition (Boundary optimality [Prop. 5]**)**

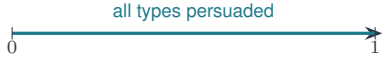
For every $w_h, w_l > 0$, $\hat{\theta} \in (0,1)$, and $\underline{\sigma}^2 \in (0, \hat{\theta}(1-\hat{\theta}))$, a boundary menu is optimal. If $w_h \geq w_l$, it is $(0, \frac{1}{2})$ with value $w_h + w_l \underline{T}(\frac{1}{2})$, where $\underline{T}(t) := \inf_{F \in \Pi(\hat{\theta}, \underline{\sigma}^2)} F((t,1])$ is the strict upper-tail guarantee; the case $w_l > w_h$ is symmetric.

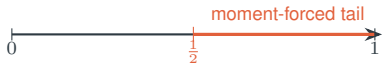
After favorable news h

$\theta_h = 0$: every citizen is persuaded. The higher-payoff realization is **fully insured**.

After unfavorable news l

$\theta_l = \frac{1}{2}$: only citizens strictly above the margin are persuaded.





With $w_l > w_h$, the optimal menu is $(\frac{1}{2}, 0)$. This resonates with the alternative-reality account of propaganda in Szeidl & Szűcs (2025): criticism or adverse evidence validates a populist narrative in which attacks from hostile elites serve as proof that the politician is challenging the corrupt establishment on behalf of the people.

▸ Saddle point

Moment restrictions determine the guaranteed tail mass

Proposition (Moment-forced tail mass [Prop. 4])

For $t \in (0,1)$,

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

attained by a two- or three-point distribution on $\{0, t, 1\}$.

Moment restrictions determine the guaranteed tail mass

Proposition (Moment-forced tail mass [Prop. 4])

For $t \in (0,1)$,

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

attained by a two- or three-point distribution on $\{0, t, 1\}$.

Three pieces govern the guarantee:

- mean-forced tail mass;
- variance-forced tail mass;
- zero guaranteed tail mass.

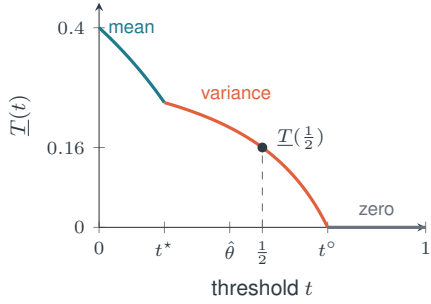
Moment restrictions determine the guaranteed tail mass

Proposition (Moment-forced tail mass [Prop. 4])

For $t \in (0,1)$,

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

attained by a two- or three-point distribution on $\{0, t, 1\}$.



$\hat{\theta} = 0.4,\ \underline{\sigma}^2 = 0.12.$

Three pieces govern the guarantee:

- mean-forced tail mass;
- variance-forced tail mass;
- zero guaranteed tail mass.

Least favorable mass sits on 0, t, and 1: mass at t absorbs moment requirements without entering the strict upper tail; the sender is paid only above the threshold.

Dispersion helps: monotone in the variance bound

Proposition (Monotonicity in $\underline{\sigma}^2$ [Prop. 6])

A tighter variance lower bound shrinks $\Pi(\hat{\theta}, \underline{\sigma}^2)$, so every menu's payoff guarantee, and hence the optimal value, is ***weakly increasing*** *in $\underline{\sigma}^2$. On the variance-binding region,*

$$V^* = \max\{w_h, w_l\} + 2\min\{w_h, w_l\}\big(\underline{\sigma}^2 - \hat{\theta}(\tfrac{1}{2} - \hat{\theta})\big),$$

strictly increasing with slope $2\min\{w_h, w_l\} > 0$.

Dispersion helps: monotone in the variance bound

Proposition (Monotonicity in $\underline{\sigma}^2$ [Prop. 6])

A tighter variance lower bound shrinks $\Pi(\hat{\theta}, \underline{\sigma}^2)$, so every menu's payoff guarantee, and hence the optimal value, is ***weakly increasing*** *in $\underline{\sigma}^2$. On the variance-binding region,*

$$V^* = \max\{w_h, w_l\} + 2\min\{w_h, w_l\}\big(\underline{\sigma}^2 - \hat{\theta}(\tfrac{1}{2} - \hat{\theta})\big),$$

strictly increasing with slope $2\min\{w_h, w_l\} > 0$.

Mechanism

Evidence of polarization rules out the least favorable populations that concentrate mass at the persuasion threshold; Nature must leave more mass in the paid upper tail.

Dispersion helps: monotone in the variance bound

Proposition (Monotonicity in $\underline{\sigma}^2$ [Prop. 6])

A tighter variance lower bound shrinks $\Pi(\hat{\theta}, \underline{\sigma}^2)$, so every menu's payoff guarantee, and hence the optimal value, is ***weakly increasing*** *in $\underline{\sigma}^2$. On the variance-binding region,*

$$V^* = \max\{w_h, w_l\} + 2\min\{w_h, w_l\}\big(\underline{\sigma}^2 - \hat{\theta}(\tfrac{1}{2} - \hat{\theta})\big),$$

strictly increasing with slope $2\min\{w_h, w_l\} > 0$.

Mechanism

Evidence of polarization rules out the least favorable populations that concentrate mass at the persuasion threshold; Nature must leave more mass in the paid upper tail.

Contrast with public Bayesian persuasion (Gitmez & Molavi 2022): there a more diverse (known) population makes public persuasion harder; here dispersion *evidence* restricts Nature and weakly raises the guarantee.

A higher average prior might hurt

Proposition (Non-monotonicity in $\hat{\theta}$ [Prop. 7])

At $\tau = \frac{1}{2}$, varying feasible $\hat{\theta}$ with $\underline{\sigma}^2 < \hat{\theta}(1-\hat{\theta})$,

$$V^* = \max\{w_h, w_l\} + \min\{w_h, w_l\}\, \underline{T}(\tfrac{1}{2}),$$

*which **fails to be monotone** in $\hat{\theta}$ iff $\underline{\sigma}^2 < 3/16$.*

A higher average prior might hurt

Proposition (Non-monotonicity in $\hat{\theta}$ [Prop. 7])

At $\tau = \frac{1}{2}$, varying feasible $\hat{\theta}$ with $\underline{\sigma}^2 < \hat{\theta}(1-\hat{\theta})$,

$$V^* = \max\{w_h, w_l\} + \min\{w_h, w_l\}\, \underline{T}(\tfrac{1}{2}),$$

*which **fails to be monotone** in $\hat{\theta}$ iff $\underline{\sigma}^2 < 3/16$.*

Mechanism

In the variance-binding region, the least favorable distribution can raise the mean by moving mass to $\frac{1}{2}$: more favorable on average, but still outside the strict upper tail.

A higher average prior might hurt

Proposition (Non-monotonicity in $\hat{\theta}$ [Prop. 7])

At $\tau = \frac{1}{2}$, varying feasible $\hat{\theta}$ with $\underline{\sigma}^2 < \hat{\theta}(1-\hat{\theta})$,

$$V^* = \max\{w_h, w_l\} + \min\{w_h, w_l\}\, \underline{T}(\tfrac{1}{2}),$$

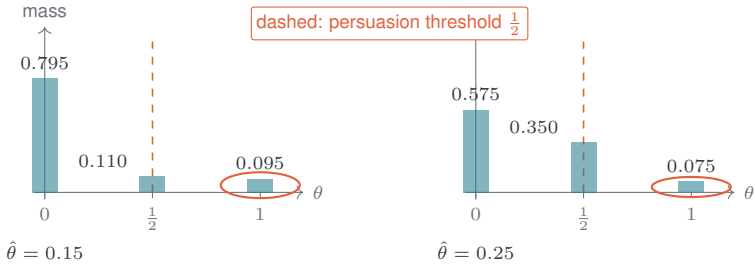
*which **fails to be monotone** in $\hat{\theta}$ iff $\underline{\sigma}^2 < 3/16$.*

Mechanism

In the variance-binding region, the least favorable distribution can raise the mean by moving mass to $\frac{1}{2}$: more favorable on average, but still outside the strict upper tail.

$$\hat{\theta} \uparrow \implies \text{more mass at the margin} \implies T(\tfrac{1}{2}) \downarrow .$$

Illustration: guaranteed mass above $\frac{1}{2}$ falls



Worst-case three-point distributions with $\underline{\sigma}^2 = 0.10$, $\tau = 1/2$. Raising $\hat{\theta}$ from 0.15 to 0.25 lowers mass above $\frac{1}{2}$ from 9.5% to 7.5% (Fig. 5 in the paper).

Discussion

Beyond binary states

The binary model is a scalar-index model. The same threshold geometry extends whenever receiver behavior is governed by a **scalar cutoff index**.

Beyond binary states

The binary model is a scalar-index model. The same threshold geometry extends whenever receiver behavior is governed by a **scalar cutoff index**.

Binary partition of states. Example: $H = \{q \geq q^*\}$, $L = H^c$ for a quality cutoff. Behavior-equivalent if action choice, narrative selection, and the sender's menu ordering depend on quality beliefs only through $\Pr(H)$. Payoff guarantees additionally require payoff differences to be summarized by $\Pr(H)$. Then $\Pr(H)$ is the scalar receiver type θ.

Beyond binary states

The binary model is a scalar-index model. The same threshold geometry extends whenever receiver behavior is governed by a **scalar cutoff index**.

Binary partition of states. Example: $H = \{q \geq q^*\}$, $L = H^c$ for a quality cutoff. Behavior-equivalent if action choice, narrative selection, and the sender's menu ordering depend on quality beliefs only through $\Pr(H)$. Payoff guarantees additionally require payoff differences to be summarized by $\Pr(H)$. Then $\Pr(H)$ is the scalar receiver type θ.

Scalar posterior statistic. If action choice is cutoff-based in $E_\mu[\omega]$, the posterior mean replaces θ. KG (2011, Proposition 3; Appendix Proposition 6) covers expected-state payoffs and linear-statistic extensions. Kolotilin et al. (2017) impose linear-in-state utilities, so messages map to cutoff types.

Residual distinctions that affect action choice, narrative selection, menu ordering, or payoff increments push beliefs into a higher-dimensional simplex.

▸ Further results

Conclusion

Conclusion

Main message

Robust narrative persuasion studies public narrative-menu design when the data-generating model is fixed and the cross-sectional distribution of receivers' priors is only partially known.

Conclusion

Main message

Robust narrative persuasion studies public narrative-menu design when the data-generating model is fixed and the cross-sectional distribution of receivers' priors is only partially known.

- **Public-menu design:** the fit-based adoption problem reduces to threshold profiles for a heterogeneous-prior population; the same menu links thresholds across realizations.

Conclusion

Main message

Robust narrative persuasion studies public narrative-menu design when the data-generating model is fixed and the cross-sectional distribution of receivers' priors is only partially known.

- **Public-menu design:** the fit-based adoption problem reduces to threshold profiles for a heterogeneous-prior population; the same menu links thresholds across realizations.
- **Robust guarantee:** combining the threshold frontier with moment ambiguity yields a least-favorable moment problem over prior distributions.

Conclusion

Main message

Robust narrative persuasion studies public narrative-menu design when the data-generating model is fixed and the cross-sectional distribution of receivers' priors is only partially known.

- **Public-menu design:** the fit-based adoption problem reduces to threshold profiles for a heterogeneous-prior population; the same menu links thresholds across realizations.
- **Robust guarantee:** combining the threshold frontier with moment ambiguity yields a least-favorable moment problem over prior distributions.
- **Application:** in the two-signal dictatorship application, a boundary menu is optimal, dispersion evidence helps, and a higher average prior can lower the sender's value.

Testable implication: with a positive action threshold, no public menu persuades every receiver after every realization; near-universal persuasion signals forces beyond narrative selection.

References


Aina, Chiara. 2025. *Tailored Stories*.


Ball, Ian, and D. Kattwinkel. 2026. *Robust Robustness*.


Gitmez, A. Arda, and Pooya Molavi. 2022. *Informational Autocrats, Diverse Societies*.


Kamenica, Emir, and Matthew Gentzkow. 2011. Bayesian Persuasion. *AER*.


Paul, Christopher, and Miriam Matthews. 2016. *The RAND "Firehose of Falsehood" Model*.


Schwartzstein, Joshua, and Adi Sunderam. 2021. Using Models to Persuade. *AER*.


Winkler, Gerhard. 1988. Extreme Points of Moment Sets. *Mathematics of Operations Research*.

Thank you!

Backup Slides

Backup: Hong Kong 2019

province. The US has compared China to a 'thuggish regime'. Photograph: STR/AFP/Getty Images

A US official has described China as a "thuggish regime" for disclosing personal details about a US diplomat who met student leaders involved in Hong Kong's pro-democracy movement.

The denunciation came as the US became the latest country to issue a travel alert to the territory on Thursday, and Hong Kong's police force brought out of retirement a senior officer who led the police response to the 2014 Occupy movement.

China's Hong Kong office asked the US on Thursday to explain reports in Communist party-controlled media that American diplomats were in contact with leaders of protests that have convulsed Hong Kong for nine weeks.

Hong Kong newspaper Ta Kung Pao published a photograph of a US diplomat, who it identified as Julie Eadeh of the consulate's political section, talking to student leaders including Joshua Wong in the lobby of a luxury hotel.

The photograph appeared under the headline "Foreign Forces Intervene", continuing a theme of previous protests from Beijing officials, who have blamed Hong Kong's unrest on "black hands" from the US.

Source: The Guardian, Aug. 9, 2019.

Back

Backup: adverse evidence can confirm a narrative

Adverse evidence can confirm a narrative. If a narrative says that intellectual elites conspire against the populist, then elite criticism can *strengthen* support among receptive receivers. Szeidl and Szucs (2025, AER).

This gives intuition for "bad news can help" and shows why reducing the same public event to a single binary signal, under one fixed interpretation, misses narrative selection.

Complements the main motivation frame on the "firehose of falsehood": high-volume, multichannel propaganda with weak consistency discipline. Source: Paul and Matthews (2016, RAND PE-198-OSD).

Back

Backup: a menu of narratives (firehose of falsehood)

RAND "firehose of falsehood" Paul & Matthews (2016), PE-198-OSD

Key feature: **no commitment to consistency**. Contradictory narratives are broadcast simultaneously.

Public realization	Narrative A	Narrative B
Troops in Ukraine	no Russian soldiers present	volunteers protecting Russians
MH17 shootdown	Ukrainian jet responsible	Western/rebel provocation
Crimea referendum	sovereign democratic choice	defense against NATO expansion

Each narrative resonates with a different audience segment. The sender does not pick one story; she runs a menu.

Pattern

Public communication often looks like a **menu of narratives** rather than a single Bayes-plausible signal.

Adverse evidence Back

Backup: the same event, more dual interpretations

Public realization	Interpretation A	Interpretation B
Election result	legitimate mandate	manipulated outcome
Product review	credible quality evidence	paid reviews
Forensic evidence	reliable proof of guilt	flawed procedure
Protest report	genuine public grievance	foreign interference

Each row is a public realization that supports competing likelihood explanations; receivers select by fit given their priors.

Back

Backup: moment ambiguity sets

General form

$$\Pi = \mathcal{F}(g, Y) := \{F \in \Delta([0,1]) : \mathbb{E}_F[g(\theta)] \in Y\}.$$

Mean bounds

$$g(\theta) = \theta, \qquad Y = [\underline{\theta}, \overline{\theta}].$$

Polls or aggregate ratings bound the average prior.

Mean and dispersion

$$g(\theta) = (\theta, \theta^2),$$
$$Y = \{(\hat{\theta}, q) : q \geq \hat{\theta}^2 + \underline{\sigma}^2\}.$$

Polarization bounds how concentrated the population can be.

The star g-interior condition rules out guarantees that rely only on exact boundary equalities.

Back

Backup: why the lower semicontinuous payoff appears

Robust evaluation for a threshold profile

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

- At a positive threshold, mass exactly at θ_s is not counted as persuaded; types just below θ_s are not persuaded either.
- At a zero threshold, type 0 has no lower neighbor in $[0,1]$.
- Under star g-interior moment ambiguity,

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

Back

Backup: narrative vs. Bayesian persuasion

In Bayesian persuasion, *Bayes plausibility* forces belief movements to balance: posteriors cannot beat the prior after *every* realization. Narrative persuasion **breaks this balance**: a receiver can choose *different* narratives after different realizations, so the same prior can yield favorable posteriors after *all* realizations.

Corollary 1 (symmetric)

The symmetric pair (t, t) is implementable iff

$$t \geq \tau/2.$$

Both thresholds can fall below τ, but not below $\tau/2$.

Two caveats: breaking balance is not a free lunch

(i) Not necessarily higher value: the data-generating model is *fixed* here but *chosen* in Bayesian persuasion, so the two senders' values are not directly ranked. **(ii) Welfare is ambiguous:** extra persuasion shifts a type-θ receiver's welfare by $\theta u_H \sum_s \pi(s \mid H)\Delta a_s - (1-\theta)|u_L|\sum_s \pi(s \mid L)\Delta a_s$, which changes sign across θ.

(Breaking balance only enlarges the feasible threshold set beyond Bayes-plausibility.)

Back

Backup: the dual program

For an implementable threshold profile $(\theta_s)_{s\in S}$, duality (Rockafellar 1970) gives a finite-dimensional characterization:

$$\min_{F\in\Pi}\int_0^1 \tilde{v}\big(\theta;(\theta_s)_{s\in S}\big)\,F(d\theta) = \sup_{(\gamma,\lambda)}\Big\{\gamma + \inf_{y\in Y}\lambda^\top y : \gamma + \lambda^\top g(\theta) \le \tilde{v}(\theta)\ \forall\theta\Big\}.$$

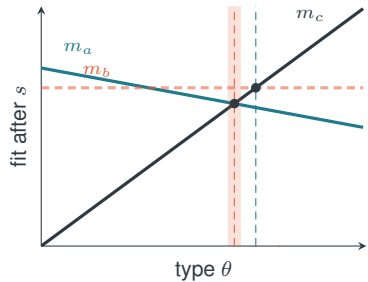
- The dual picks an *affine minorant* of the per-type payoff from the moment functions.
- A worst-case F^* is supported on the **contact set** where minorant $=$ payoff: $\le d+1$ **atoms** for d moments (Winkler 1988).
- Caveat: the contact set is *menu-dependent*, so the $d{+}1$-atom worst case does **not** imply that $d{+}1$ narratives suffice.

Back

Backup: proof sketch of two-narrative sufficiency

- Any two narratives cross *at most once*; persuaded types form an *upper set*.
- For each s, retain a narrative that is fit-maximal at the marginal type $\theta_s(M)$ and persuades there.
- Monotonicity then preserves all types above $\theta_s(M)$.
- If the retained pair persuades lower types as well, the threshold strictly falls.

Pointwise dominance for every $F \in \Pi$ gives $V(M') \geq V(M)$.



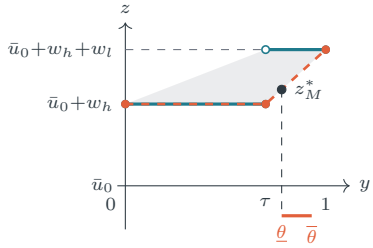
- model switching point (fit curves cross) persuasion threshold θ_s

Keep $m_a = (\frac{1}{8}, \frac{3}{8})$ and $m_c = (1, 0)$; remove $m_b = (\frac{1}{3}, \frac{1}{3})$, so $\theta_s(M') < \theta_s(M)$.

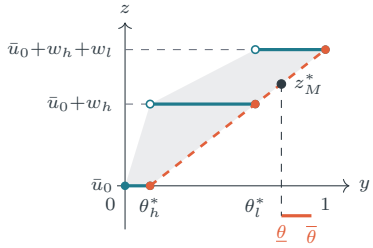
Back

Backup: convexification geometry

The inner value is read off the *lower boundary* of the closed convex hull of $\{(g(\theta), \tilde{v}(\theta; \theta_h, \theta_l)) : \theta \in [0,1]\} \subseteq \mathbb{R}^{d+1}$ over moment vectors in Y.



(a) boundary pair $(0, \tau)$.



(b) frontier interior pair.

Shaded: $\overline{\text{conv}}\{(y, \tilde{v}(\cdot))\}$;

dashed highlight = lower boundary over Y; contact points carry worst-case mass. Power moments: minorant is a degree-d polynomial.

Back

Backup: saddle-point selection (Proposition 3)

When does robustness select a benchmark optimum?

If (M^*, F^*) is a saddle point, then M^* is benchmark-optimal at F^* among menus evaluated under that fixed distribution.

Example 4 shows this can fail: $M_{h\text{-bd}}$ is robust optimal but not benchmark-optimal at $F^* = \delta_{0.4}$.

- Robustness evaluates each deviation against *its own* worst case.
- The selected menu and worst-case F must be **mutual best responses**.
- Checking benchmark optimality at F^* requires verifying no profitable shift along the relevant persuasion frontier.

Back

Backup: further results I

- **Product-quality application (Prop. C.1).** Interval-mean ambiguity $\Pi([\underline{\theta}, \overline{\theta}])$ collapses to the *lower* endpoint; a boundary menu is optimal when the insured realization is valuable enough, else an interior menu spreads persuasion.
- **Default model (Props. A.1–A.2).** If the data-generating model π stays available, the implementable region *shrinks* (one additional algebraic constraint per signal realization).
- **Beyond $\tau = \frac{1}{2}$ (Prop. B.1).** Same forces: value increasing in $\underline{\sigma}^2$ and non-monotone in $\hat{\theta}$ on the boundary-optimal variance-binding region. Sufficient conditions for a boundary menu include small τ, $\hat{\theta} \leq \frac{1}{2}$, high $\underline{\sigma}^2$, or a sufficiently dominant payoff weight on one realization.

Back