Identity-Based Political Communication

Helios Herrera* and Ravideep Sethi†

August 20, 2026

Abstract

Despite seemingly damning news, support for some political leaders remains remarkably resilient and polarized. We develop a model of political communication in which a leader designs public messages to maximize support in an electorate composed of pro-regime partisans, anti-regime partisans, and independents. Citizens are rational Bayesian updaters, but partisans endogenously select media environments that sustain identity-consistent beliefs. Anticipating these responses, the leader optimally trades off retaining loyal supporters against persuading independents.

We show that this interaction endogenously generates persistent polarization and limited responsiveness to fundamentals, even in the absence of heterogeneous priors or behavioral biases. Equilibrium communication takes one of three forms—pro-regime retention, independent targeting, or full revelation—depending on the composition of the electorate; when the leader can also underplay good news, a fourth regime, partisan targeting, appears whenever the prior exceeds one quarter. The model provides a rational benchmark for resilient political support and identifies the conditions under which propaganda is effective or self-defeating.

JEL codes: D72, D83, D90

Keywords: Populism, Propaganda, Political Identity

*CEPR and Department of Economics, University of Warwick, Coventry CV4 7AL, UK. Email: h.herrera@warwick.ac.uk

†Division of Quantitative Analysis of Markets and Organizations, David Eccles School of Business, University of Utah, Salt Lake City, UT 84111, USA. Email: ravi@eccles.utah.edu

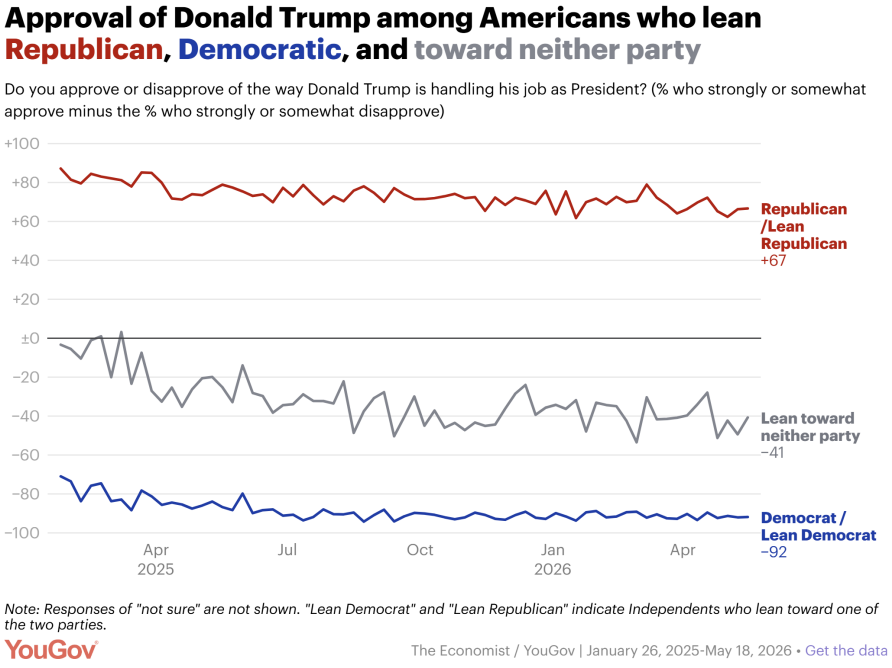
1 Introduction

1.1 Motivation: Polarization Without Movement

Political support for leaders is often sharply divided and remarkably stable within groups. Leaders facing weak economic performance, policy failures, or adverse shocks frequently retain strong backing from their supporters while remaining deeply unpopular among their opponents. What is striking is not only the degree of polarization, but its persistence: support appears largely unresponsive to new information, even when that information is widely observed.

Figure 1 illustrates this pattern using recent survey data.

Figure 1: Net approval of Donald Trump by partisan identity. *Notes:* Net approval is the percentage who strongly or somewhat approve minus the percentage who strongly or somewhat disapprove. Partisan series combine identifiers with leaners; the Independent series comprises independents who lean toward neither party. *Source:* YouGov, "Trump's Job Performance, Iran, the Economy, and More: May 15–18, 2026 Economist/YouGov Poll," May 19, 2026; weekly Economist/YouGov surveys, January 26, 2025–May 18, 2026.



Approval is consistently high among pro-regime supporters and low among anti-regime citizens, with little convergence over time. Aggregate movements are driven primarily by independents, while partisan groups remain anchored. This combination of polarization and stability is difficult to reconcile with standard models of belief updating,

in which agents process common information and update toward the truth. Instead, beliefs appear insulated from new information. Similar patterns arise across countries and political systems.

1.2 Mechanism: Identity-Preserving Communication

A key feature of modern politics is endogenous media choice. Individuals sort across information sources in ways that reflect political identity, generating systematically different information environments across groups (e.g., Gentzkow and Shapiro, 2011; Flaxman et al., 2016).

We study how a strategic leader communicates in such an environment. A pro-regime leader designs a public signal to maximize support, anticipating that citizens differ not in priors or updating rules, but in their choice of information. The electorate consists of pro-regime partisans, anti-regime partisans, and independents. All agents are rational Bayesian updaters and share a common prior.

Partisans choose information structures to maximize the likelihood of holding identity-consistent beliefs, while independents process information non-strategically. Anticipating these responses, the leader selects messages that are filtered, reinforced, or counteracted across groups.

We refer to this interaction as *identity-preserving communication*: the leader tailors messages to audiences whose information environments already protect identity, while citizens select media that limit exposure to unfavorable updates.

1.3 Contribution

The model provides a rational account of persistent polarization in political support. Differences across groups arise endogenously from identity-driven media choice, rather than from heterogeneous priors or behavioral biases. In equilibrium, support is ordered as

$$\text{pro-regime} \;>\; \text{independent} \;>\; \text{anti-regime},$$

and this ordering emerges from optimal behavior.

We characterize the leader's optimal communication strategy. When the prior is unfavorable, equilibrium messaging falls into three regimes. In a *pro-regime retention* regime, the leader preserves support among loyalists while extracting limited persuasion from independents. In an *independent targeting* regime, the leader adopts more informative communication, trading off partisan support for gains among persuadable voters. Finally, in a *full revelation* regime, manipulation collapses: countervailing information neutralizes

persuasion, and truthful communication becomes optimal. A fourth regime, *partisan targeting*, arises once the leader can also underplay good news, and only when the prior exceeds one quarter.

1.4 Summary of Results

Our main results are as follows:

- **Polarization without biases.** Persistent and polarized support arises in equilibrium even though all agents share a common prior and update according to Bayes' rule. The cross-sectional ordering
$$S^P > S^I > S^A$$
emerges endogenously from identity-driven information choice.

- **Optimal communication is regime-dependent.** The leader's strategy varies with the composition of the electorate. When pro-partisans are dominant, the leader preserves support through minimally informative messages. When the electorate is balanced, the leader increases informativeness to persuade independents. When anti-partisans are sufficiently numerous, countervailing information eliminates the gains from persuasion and full revelation becomes optimal.

- **Limited responsiveness to fundamentals.** Aggregate approval is largely stable because changes in fundamentals primarily affect independents, while partisan support remains anchored. This generates polarization without movement: persistent cross-group differences alongside muted aggregate fluctuations.

- **Endogenous limits to persuasion.** The effectiveness of communication is constrained by countervailing information choices. As persuasion intensifies, opposition groups optimally offset its impact. When access to such countervailing information is restricted, persuasion becomes more effective and more aggressive communication is sustained.

1.5 Discussion

The analysis highlights how media fragmentation shapes political outcomes. Propaganda is most effective when reinforced by identity-consistent media and least effective when opposing groups can access countervailing information. Persistent polarization therefore arises not from biased updating, but from the interaction of strategic communication and endogenous media choice.

The framework also explains why aggregate approval is relatively stable: changes in fundamentals primarily affect independents, while partisan beliefs remain anchored. As a result, large and persistent differences across groups coexist with limited aggregate movement.

More broadly, the model shows that polarized and resilient support can arise under fully rational behavior, without relying on heterogeneous priors or behavioral biases, just on identity preferences.

2 A leading example

We illustrate the model's three forces—pro-retention, independent persuasion, and opposition counter-design—through a single comparative-statics exercise. Fix the prior $w = 0.45$ and the share of independents $p_I = 0.40$. We sweep the partisan coalition by setting $p_P \in [0, 0.60]$ with $p_A = 0.60 - p_P$, holding the share of independents constant. Three values of p_P display the regimes.

Figure 2: Optimal communication as the coalition tilts from anti- to pro-regime, with $w = 0.45$ and $p_I = 0.40$ fixed.

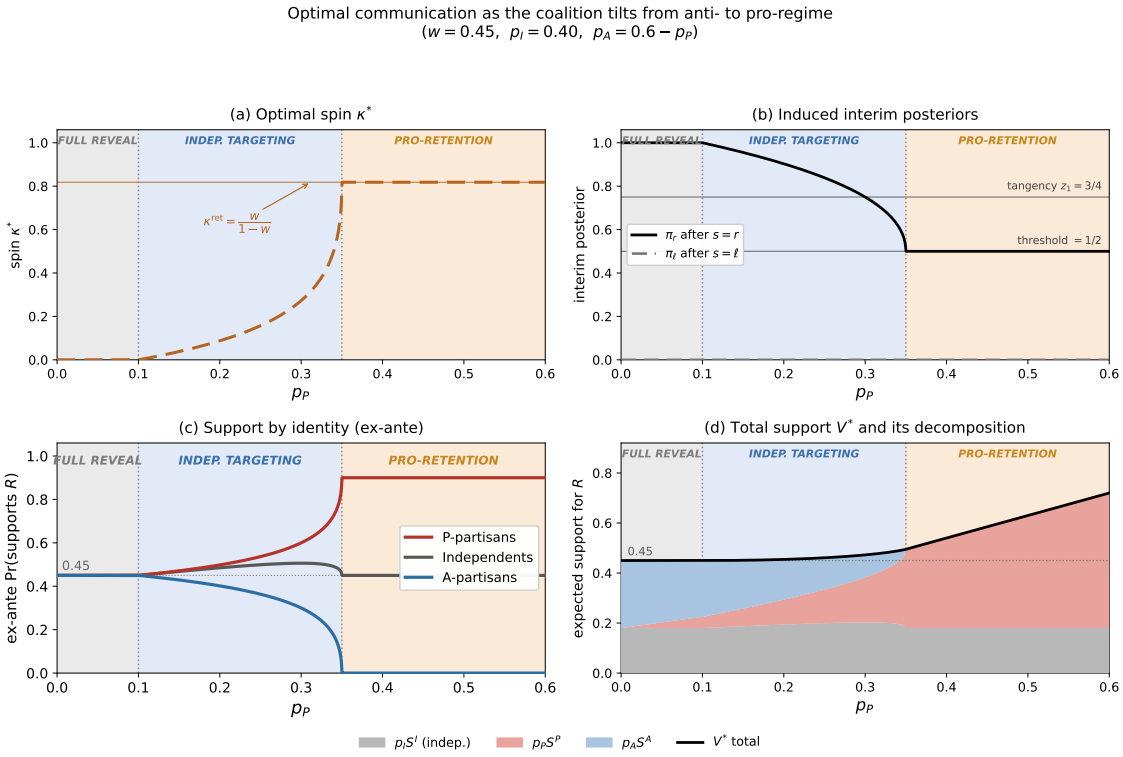


Figure 2 traces the full sweep, showing how the leader's persuasion strategy becomes

more aggressive and abandons full revelation, shifting its target toward a growing share of pro-partisans. As a result, overall approval rises, while becoming increasingly polarized—very high among pro-partisans and very low among anti-partisans. More specifically, we outline the equilibrium communication strategies below.

Pro-leaning ($p_P = 0.45$). With pro-partisans dominating, the leader chooses the *least informative* signal that still maximizes pro-partisan support: $(\kappa^*, \mu^*) = (\kappa_{\text{ret}}, 0)$ with $\kappa_{\text{ret}} \equiv w/(1-w) \approx 0.82$, inducing posteriors $\pi_r = 1/2$ and $\pi_\ell = 0$. Any less informative—any higher κ—and the favorable posterior π_r falls below $1/2$, costing pro-support; any more informative and pro-support also falls, because good news, now stronger than her identity requires, arrives less often: $\Pr[s = r] = w/\pi_r$. Both partisans choose the uninformative experiment N: pro-partisans' tie-break converts $\pi_r = 1/2$ into a vote for R; anti-partisans' tie-break passes them passively to L; independents contribute exactly w in expectation. The partisan media market is silent. Total support is $V^* = 0.585$, against the $w = 0.45$ she would collect under full revelation. Silence, however, would already deliver $p_I\, h(w) + 2w\, p_P = 0.575$: with a base this large, its own Inside media does nearly all the work, and public communication adds little on top.

Balanced ($p_P = p_A = 0.30$). When pro-partisans no longer dominate, the leader trades retention for persuasion. The optimum $(\kappa^*, \mu^*) = (3/11, 0)$ pushes π_r to $3/4$, the concavification tangency of $h(\pi) = 3\pi^2 - 2\pi^3$. Pro-partisan support falls to $S^P = 4w/3 = 0.6$, but independents now contribute $S^I = 9w/8 \approx 0.506$ and the leader extracts $S^A = 2w/3 = 0.3$ from anti-partisans in state R. The new force is endogenous counter-design: facing $\pi_r = 3/4$, anti-partisans' tie-break no longer protects them, so they adopt a GT experiment $(p_{g|R},\, p_{g|L}) = (1/3, 1)$ that pins their combined posterior at $1/2$ exactly. Total support is $V^* = 0.4725$.

Anti-leaning ($p_P = 0.15$). As the coalition tilts further against the leader, the optimal π_r continues past the tangency, solving $p_I\, \pi^2(3 - 4\pi) = p_P - p_A$. At $p_P = 0.15$, $\pi_r \approx 0.955$—that is, $\kappa^* \approx 0.038$—nearly fully revealing; anti-partisans refine GT to $(0.047, 1)$, still pinning the combined posterior at $1/2$. Once $p_P \leq 0.10$, persuasion collapses entirely: no posterior worth persuading independents survives anti-counter-design without losing more pro-partisans than it buys. The unique optimum is full revelation $(\kappa^*, \mu^*) = (0, 0)$, with $V^* = w = 0.45$—the same value the leader obtains at $p_P = 0$, where no pro-partisan exists.

The example exhibits two patterns formalized in the rest of the paper. First, optimal communication moves through three phases rather than smoothly: κ^* is flat at κ_{ret} across

the pro-retention plateau $p_P \geq 0.35$; it then falls continuously—though with a vertical tangent at that boundary, since $\pi_r^2(3 - 4\pi_r)$ is stationary at $\pi_r = \frac{1}{2}$—while π_r climbs from $1/2$ through the tangency $3/4$ to 1; and it is flat at 0 once the leader concedes to truth, $p_P \leq 0.10$. Second, opposition counter-design caps the leader's reach: anti-partisans' GT response pins the combined posterior at the voting threshold, so each additional unit of leader manipulation in the balanced and anti-leaning cases is absorbed by an additional unit of anti-partisan refinement—an identity-driven arms race. The cross-sectional support ordering $S^P > S^I > S^A$ is therefore not assumed but the equilibrium consequence of these endogenous choices.

3 Related Literature

Our analysis lies at the intersection of four literatures: information design and Bayesian persuasion with heterogeneous audiences; belief-based utility and political identity; endogenous media choice and polarization; and the political economy of propaganda. We summarize how the paper relates to, and contributes to, each.

Bayesian persuasion with strategic, heterogeneous audiences. Methodologically, the paper applies the tools of Bayesian persuasion and information design developed in Kamenica and Gentzkow [2011], Bergemann and Morris [2016], Bergemann and Morris [2019], and Kolotilin [2018] to a political-communication environment in which the audience is itself composed of strategic information designers. A growing literature studies persuasion of multiple receivers who differ in preferences or priors [Alonso and Câmara, 2016; Bardhi and Guo, 2018; Chan et al., 2019; Arieli and Babichenko, 2019], and a complementary line allows the sender to be uncertain about receiver types or their information [Kolotilin et al., 2017; Lipnowski and Mathevet, 2018]. Matysková and Montes [2023] study persuasion when a single receiver can acquire her own costly information after observing the sender's signal, and Matějka and Tabellini [2021] endogenize voters' attention in an electoral model where candidates choose platforms anticipating it. Closest to our setting is Gitmez and Molavi [2022], who study Bayesian persuasion by an autocrat facing a heterogeneous audience and show that greater diversity of preferences and beliefs disciplines manipulation. Our framework is distinct on two dimensions. First, partisans are not merely heterogeneous receivers; they are *Bayesian counter-designers* who, having observed the leader's commitment, design a private experiment to neutralize unfavorable updates—and the leader's optimum is therefore the solution to a public concavification problem in which the relevant posteriors are those that survive the partisans' counter-design. Second, the audience contains both a pro-aligned and an anti-aligned

designer, so the public message must trade off retention of the former against persuasion of the latter through the independents. The resulting equilibrium concept, public persuasion under endogenous private counter-design by adversarially aligned audiences, delivers predictions—in particular the four-regime structure of optimal communication and the equilibrium ordering $S^P > S^I > S^A$—that have no analogue in the standard homogeneous-audience benchmark.

Identity, motivated beliefs, and political behavior. The preferences we ascribe to partisans build on the literature on belief-based utility, beginning with Akerlof and Dickens [1982] and developed in Caplin and Leahy [2001], Brunnermeier and Parker [2005], Köszegi [2006], and Bénabou and Tirole [2016]. We adopt the identity-based formulation of Akerlof and Kranton [2000] and follow Herrera and Sethi [2025], who show how citizens whose identity-consistent beliefs enter the utility function endogenously sort into media that protect those beliefs.[1] Our contribution is to embed identity-driven information design into a strategic communication game rather than a purely decision-theoretic problem. Beliefs are a choice variable for the partisan but a stochastic outcome for the leader, and the two are linked through Bayes plausibility. The framework therefore provides a rational microfoundation for the affective polarization documented in Iyengar et al. [2019] and Mason [2018] that does not rely on heterogeneous priors or non-Bayesian updating. This is in contrast to the identity-politics program of Bonomi et al. [2021], Gennaioli and Tabellini [2025], and Gennaioli et al. [2026], where identity distorts beliefs through ingroup stereotypes.[2]

Endogenous media choice, segmentation, and polarization. The empirical premise that citizens consume systematically different media along partisan lines is well established. Gentzkow and Shapiro [2011], Flaxman et al. [2016], and Halberstam and Knight [2016] document substantial ideological segregation in news consumption and online networks. DellaVigna and Kaplan [2007], Martin and Yurukoglu [2017] , and Manacorda et al. [2026] show that exposure to partisan outlets has measurable electoral effects, while Boxell et al. [2017] find that polarization has grown most among demographic groups with the lowest internet use, suggesting the media environment is one driver among several. Theoretical work has emphasized the supply side, in which outlets choose slant in response to audience demand [Mullainathan and Shleifer, 2005; Gentzkow and

[1] Related contributions include Eliaz and Spiegler [2006] on belief-driven information acquisition, and Schwardmann et al. [2022] and Kaanders et al. [2022] on motivated information sampling. Political applications of related ideas include Little [2019], Kahan [2017], and Acharya et al. [2018].

[2] D'Amico and Tabellini [2025] document the partisan behavior our model generates: relative to independents, partisans avoid bad news about their own candidate but actively seek it about the opponent.

Shapiro, 2006; Bernhardt et al., 2008; Chan and Suen, 2008; Perego and Yuksel, 2022]. The demand-side counterpart—how citizens choose among outlets given their identity—is the focus of Herrera and Sethi [2025] and is the foundation of partisan behavior in our model. Our contribution is to interlace demand and supply through a strategic leader. The leader internalizes how citizens of different identities will sort across media in response to her message, and the citizens internalize that message in choosing their information environment. Segmentation, public communication, and identity-consistent beliefs are jointly determined in equilibrium, and the framework speaks directly to the stylized fact that approval is sharply polarized across groups but largely unresponsive to fundamentals [Jones, 2004; Scala and Johnson, 2017; Iyengar et al., 2019].

Propaganda and strategic political communication. A long tradition in political economy studies propaganda and information control by leaders. In the coordination framework of Edmond [2013], propaganda biases citizens' inferences about regime strength to deter collective action; related contributions include Shadmehr and Bernhardt [2015] on censorship and Egorov et al. [2009] on the dictator's tradeoff between censorship and the informational value of a free press. Guriev and Treisman [2020] and Guriev and Treisman [2019] develop a signaling theory of *informational autocracy* in which incumbents manipulate beliefs about their competence in lieu of repression.[3] Most of this literature shares two features that we relax. First, the audience is treated as homogeneous, or as differing only in private signals about a common state; identity-based heterogeneity in *preferences over beliefs* is absent. Second, the audience is passive in its information acquisition: citizens process whatever signals the regime and elite transmit. By contrast, in our framework citizens of different identities actively curate their information environments, and the anti-regime audience can effectively *undo* the leader's persuasion by counter-designing a private experiment that pins their posterior at the voting threshold. A central implication is that propaganda is most effective precisely where access to countervailing inside media is restricted—a prediction consistent with the empirical evidence that propaganda has its strongest effects in information-controlled environments [Adena et al., 2015; Gagliarducci et al., 2020; Yanagizawa-Drott, 2014; Enikolopov et al., 2011; Durante et al., 2019]—and that, in open media markets, partisan retention rather than persuasion of the opposition is the typical equilibrium use of public communication, as in Glaeser et al. [2005] and as documented by Boeri et al. [2026], who find that leaders' rhetoric polarizes in the quarters before an election. The framework therefore provides a unified rationale for both the resilience of approval among supporters of populist leaders in liberal democracies and the

[3]Further contributions include Gehlbach and Sonin [2014], Kolotilin et al. [2022], and the survey by Gehlbach et al. [2016].

heavier reliance on information control in autocratic settings [Guriev and Papaioannou, 2022].

4 The model

Suppose there are two states of the world (R and L), which indicate whether the leader is good or bad. All citizens share a common prior

$$w \equiv \Pr[\omega = R] \in (0, \tfrac{1}{2}], \tag{1}$$

so the prior is stacked against the leader. This is the case in which the leader has something to do: were the prior already on her side, communication would be slack; with the prior against her, persuasion has bite.[4] A countable infinity of citizens each decide whether to support her. A citizen is either an independent, a pro-partisan, or an anti-partisan, with population shares (p_I, p_P, p_A) in the unit simplex $\Delta^2 = \{p \in \mathbb{R}^3_+ : p_I + p_P + p_A = 1\}$. The leader's payoff is total support.

Preferences An independent simply wants to back the better side: she supports the leader if she believes the state is at least as likely to be R as L.[5] Each independent also observes a conditionally i.i.d. private signal about the state, of precision q drawn uniformly from $[\frac{1}{2}, 1]$,

$$\Pr[s^I = r \mid \omega = R,\, q] = q, \qquad \Pr[s^I = r \mid \omega = L,\, q] = 1 - q. \tag{2}$$

Independents thus differ in how well informed they are, from ignorant ($q = \frac{1}{2}$) to perfectly informed ($q = 1$), with no tilt in either direction. What the leader will exploit is not the particular distribution but the heterogeneity itself: dispersion in precision turns a posterior into a *smooth* spread of support, and the shape of that spread is what she optimizes against.

A partisan, by contrast, identifies with a side. She gains utility if she is able to believe, after all the information has been received and processed rationally by her, that the state of the world matches her identity. Write $F_P = (s, \sigma)$ for the pair she ends up holding —

[4] For $w > \frac{1}{2}$ the trade-offs collapse rather than reverse. The uninformative design already secures the base outright, $S^P = 1$, because the favorable message leaves every pro-partisan above her threshold without any help from her own media; and once $w \geq \frac{3}{4}$ the independent benchmark too is attained without informative communication. While we can solve this model for all parameters, the interesting problem is the one a leader with something to prove faces.

[5] Ties are broken in the leader's favor for independents and pro-partisans, and against her for anti-partisans. In the Bayesian-persuasion literature it is standard to break ties in favor of the sender; here each partisan plays a persuasion game within herself, and without this convention the partisan's problem would have no solution.

the leader's public message s together with the realization σ of her own Inside signal. For a pro-partisan, identity preservation means a posterior belief that the leader is good with probability at least one-half,

$$U^P = \mathbf{1}\{\Pr[\omega = R \mid F_P] \geq \tfrac{1}{2}\}, \tag{3}$$

and an anti-partisan derives the symmetric utility from concluding the opposite, $U^A = \mathbf{1}\{\Pr[\omega = L \mid F_A] \geq \frac{1}{2}\}$. A pro-partisan supports the leader exactly when $U^P = 1$, and an anti-partisan exactly when $U^A = 0$.[6] Her problem is therefore to choose an Inside signal structure σ so as to maximize the chance that her own posterior comes out on her side,

$$\max_{\sigma} \;\; \Pr\Big[\, \Pr\big[\omega = R \mid s, \sigma\big] \;\geq\; \tfrac{1}{2} \,\Big], \tag{4}$$

subject to Bayes plausibility — her expected posterior must equal her prior — and taking the leader's design as given.

Equivalently, as in Herrera and Sethi [2025], one can think of each partisan as having two selves—an emotional self (heart) and a rational self (mind). The heart chooses the media the partisan consumes, in order to preserve her political identity; the mind processes all the information it receives rationally and decides whom to support. The heart cannot choose the verdict, only the evidence. This is a rational-choice model in which rationality is the constraint and the objective is the value of the posterior: there is motivated information *acquisition*, but no motivated *reasoning*. Crucially, partisans are not biased updaters—they hold the common prior and update rationally. The only thing identity governs is their choice of information.

The leader's signal The leader speaks through official media—state television, the briefing room, her own megaphone—which we model as a binary public signal $s \in \{r, \ell\}$ observed by every citizen. Her instrument is *spin*. Success is always announced; a failure she may dress up, reporting it as a success with probability κ, to which she commits before the state is realized:

$$\Pr[s = r \mid \omega = R] = 1, \qquad \Pr[s = r \mid \omega = L] = \kappa. \tag{5}$$

Leaders trumpet their successes; the strategic question is what happens to their failures.[7]

[6] Alternatively, utility could attach to an action rather than a belief: a partisan gains from *supporting* her own side, but can bring herself to do so only if she first convinces herself that her side is right. That action-based formulation is isomorphic to ours.

[7] Section 5 enriches the technology by allowing the leader to *underplay* good news as well, reporting a genuine success as a failure with some probability μ.

Because admitted bad news is conclusive, the message ℓ carries posterior $\Pr[\omega = R \mid \ell] = 0$, while the message r carries interim posterior

$$\pi_r \;=\; \frac{w}{w + (1-w)\,\kappa} \;\in\; [w, 1], \tag{6}$$

a strictly decreasing bijection between her spin and the credibility of good news. It is mathematically equivalent to let the leader choose π_r directly: spin makes good news *frequent but weak*. The favorable message arrives with probability

$$\Pr[s = r] \;=\; \frac{w}{\pi_r}, \tag{7}$$

so the design splits the prior into the posteriors $\{0, \pi_r\}$ with the unique Bayes-consistent weights.

Inside media In a rich media landscape, a partisan can choose among outlets with a wide range of political slants. We model her chosen diet as an *Inside* signal structure, which is a finite set of signals correlated with the state. The set of signal structures is only restricted by Bayes plausibility; that the expected posterior equal the prior.[8] Each partisan commits to her Inside structure after observing the leader's design but before any signal realization. This is reminiscent to a standing loyalty to outlets, rather than a reaction to headlines.

Timing First, the leader commits to her spin κ, equivalently to the target π_r. Next, each partisan, having observed the design, chooses her Inside signal structure. The state is then realized, and every citizen receives her signals: the leader's message, plus the Inside realization for partisans and the private signal for independents. Finally, citizens form Bayesian posteriors, support decisions are made, and payoffs are realized. The solution concept is subgame-perfect equilibrium; since independents are non-strategic and partisans move after the leader, the leader simply chooses π_r anticipating best responses that are unique up to payoff-irrelevant relabeling.

The object of analysis is therefore not the posterior the leader's signal generates, but the posterior that survives partisans' response to it. Her public message is a designed information shock, and each partisan group arranges its own Inside media around it to protect its identity, so the belief that drives support is the residual one this optimal counter-design leaves in place. The leader chooses not final beliefs but an input to a

[8] Any distribution of posteriors generated by consuming several outlets can be generated by a single appropriately designed structure, so multi-outlet diets are subsumed.

contest over them, and equilibrium is the fixed point between design and counter-design.

4.1 Solution of the citizen's problem

We use backward induction. Each citizen's problem is decision-theoretic: she takes the leader's design κ as given. Since it is equivalent to choose κ or π_r by (6), we express the following equations in terms of π_r for convenience.

Independents An independent who holds interim posterior π after the public message supports the leader if, after also observing her private signal, her posterior belief is at least one-half. Integrating over signal realizations and precisions, her support probability is the cubic

$$h(\pi) \;=\; 3\pi^2 - 2\pi^3, \tag{8}$$

derived in Lemma 12 (Appendix B). It is increasing throughout, strictly convex on $[0, \frac{1}{2}]$, strictly concave on $[\frac{1}{2}, 1]$, and tangent to the ray from the origin at $\pi = \frac{3}{4}$. That single change of curvature is the source of a *concavification motive*: by splitting the prior into a high and a low posterior rather than pooling, the leader does strictly better on this group.

Since the unfavorable message is conclusive—$h(0) = 0$—only the favorable signal realization contributes, and aggregate independent support is

$$S^I(\pi_r) \;=\; \Pr[s = r] \cdot h(\pi_r) \;=\; w\,\pi_r\,(3 - 2\pi_r). \tag{9}$$

Independent support is single-peaked in the strength of good news, maximized at $\pi_r = \frac{3}{4}$: this is the leader's only reason to be candid short of a hostile electorate.

Partisans The solution to the partisans' problem (4) follows Herrera and Sethi [2025], specialized to an Outside signal whose bad news is conclusive. Three Inside signals are sufficient. A *Good* (G) realization, paired with the public message that is unfavorable to the partisan, is just enough to hold her final posterior at the threshold $\frac{1}{2}$. A *Bad* (B) realization is redeemable: paired with the public message that is *favorable* to her, it too lands at $\frac{1}{2}$. A *Terrible* (T) realization concedes the state, driving her posterior to certainty against her identity. The Martingale property prices these signals—identity-preserving realizations are expensive, concessions are free—and the heart mixes the two with the best utility-to-price ratios. When she can do no better than to remain ignorant, she chooses the uninformative structure (N).

Consider first the *anti-partisan*. The message ℓ confirms her identity outright and needs no answer; the threat is the favorable message r, which, if $\pi_r > \frac{1}{2}$, would leave her unable

to hold an identity-consistent belief. Her Good signal is designed against exactly that threat: paired with r, it pins her final posterior at $\frac{1}{2}$, where the tie breaks toward her own side. Bayes plausibility prices this by forcing the complementary realization to be Terrible—she concedes the state—so her structure is GT: a heavily slanted outlet whose favorable coverage is routine and uninformative, and whose rare criticism is conclusive. She supports the leader with probability $2\pi_r - 1$ when the leader's signal is r, so

$$S^A(\pi_r) \;=\; w \max\Big\{0,\; 2 - \tfrac{1}{\pi_r}\Big\}. \tag{10}$$

Her counter-design caps what manipulation can extract: each increment in the strength of good news is met by an inside GT signal structure that counters it.

The *pro-partisan* faces the mirror problem, with one twist: her unfavorable message is beyond rescue. After ℓ her posterior is zero, and no Inside realization can move it. Her curation is therefore aimed entirely at the realization of the favorable signal (r). If $\pi_r \geq \frac{1}{2}$, that realization already preserves her identity and she optimally consumes no further information, choosing N. If $\pi_r < \frac{1}{2}$—good news too weak to carry her on its own—she chooses a *Bad* realization that, paired with r, tops her up to exactly $\frac{1}{2}$, and concedes (T) otherwise.[9] In either case her support is

$$S^P(\pi_r) \;=\; \min\Big\{2w,\; \tfrac{w}{\pi_r}\Big\}, \tag{11}$$

flat at the cap $2w$ for every $\pi_r \leq \frac{1}{2}$, and declining once good news is strengthened beyond what her identity requires.

Lemma 1 (Inside media) *Fix a design with target $\pi_r \in [w, 1]$. Partisans support the leader with probabilities (11) and (10). For $\pi_r \geq \frac{1}{2}$—the only region the leader ever chooses—the pro-partisan plays N and the anti-partisan plays GT, and*

$$S^P(\pi_r) = \frac{w}{\pi_r}, \qquad S^A(\pi_r) = w\Big(2 - \frac{1}{\pi_r}\Big), \qquad S^P(\pi_r) + S^A(\pi_r) = 2w.$$

The proof is in Appendix D. The cap $S^P \leq 2w$ is Markov's inequality applied to a posterior with mean w: a belief that averages w can reach $\frac{1}{2}$ no more often than a fraction $2w$ of the time.[10]

[9] The labels G, B, T classify a realization by the belief it sustains once combined with the public message, not by its content in isolation: G is the realization that rescues the partisan's *unfavorable* message, B the one that rescues her *favorable* message. When $\pi_r < \frac{1}{2}$ the pro-partisan's B realization is accordingly weak good news rather than bad news. Nothing below turns on the label: by Lemma 1 the leader never chooses $\pi_r < \frac{1}{2}$.

[10] We report the pro-partisan's choice as N because it is the simplest structure delivering her optimum. It is not the only one: on this slice the two-sided menu of Section 5.1 selects a degenerate GB structure, which

4.2 Spin carries, Inside media counters

Before optimizing, it is worth watching Lemma 1 happen. Fix $w = 0.45$ and read four designs off the geometry.

One object does the work. By (6) the favorable message persuades a citizen only if her belief already clears

$$a(\kappa) \;=\; \frac{\kappa}{1+\kappa}, \tag{12}$$

the *escape threshold*: the belief at which her post-r posterior is exactly $\frac{1}{2}$. Applying this at $\pi = w$ translates the design condition that runs through the rest of the paper into belief space,

$$\pi_r \;\geq\; \tfrac{1}{2} \qquad\Longleftrightarrow\qquad w \;\geq\; a(\kappa) \qquad\Longleftrightarrow\qquad \kappa \;\leq\; \kappa_{\text{ret}} \;=\; \frac{w}{1-w}, \tag{13}$$

and the retention design is exactly the one at which the two meet: $a(\kappa_{\text{ret}}) = w$. Spin raises a, and by (7) raises reach with it: frequent but weak, read in belief space.

A citizen who consumes no Inside signal—the uninformative structure N—supports the leader when the favorable message arrives and clears her threshold, so her support as a function of her belief π is

$$S(\pi;\kappa) \;=\; \big[\pi + (1-\pi)\kappa\big] \cdot \mathbf{1}\{\pi \geq a(\kappa)\}, \tag{14}$$

a jump of height $2a(\kappa)$ at the escape threshold followed by a rise to $S(1;\kappa) = 1$. Each panel below plots this curve for one κ, with the dashed line at the prior. Read at w, the curve is a citizen's support *without* the optimal use of Inside media. With it she is not read off the curve at all: she splits her belief, so her value is an envelope of S at w—concave for the pro-partisan, who maximizes support, convex for the anti-partisan, who minimizes it. Both envelopes are available in closed form,

$$(\operatorname{cav} S)(\pi) = \min\big\{2\pi,\ \pi + (1-\pi)\kappa\big\}, \qquad (\operatorname{vex} S)(\pi) = \max\Big\{0,\ \frac{\pi - a(\kappa)}{1 - a(\kappa)}\Big\}, \tag{15}$$

and evaluating them at $\pi = w$ returns exactly (11) and (10). Lemma 1 is that statement: *the two partisan supports are the two envelopes of one curve at one point*, and the vertical gap between them at w is everything Inside media can do.[11]

induces the same distribution of final posteriors, $\{0, \pi_r\}$, and hence the same support. Since $\pi_\ell = 0$, whatever her Inside signal reports alongside ℓ is payoff-irrelevant, and the two structures cannot be told apart from behavior. Appendix E's boundary check at $\mu = 0$ makes the identification explicit.

[11] The anti-partisan's tie-break runs against the leader, so her indicator in (14) is strict. This is why she can sit *at* $a(\kappa)$ rather than just below it, and hence why her envelope is attained rather than approached.

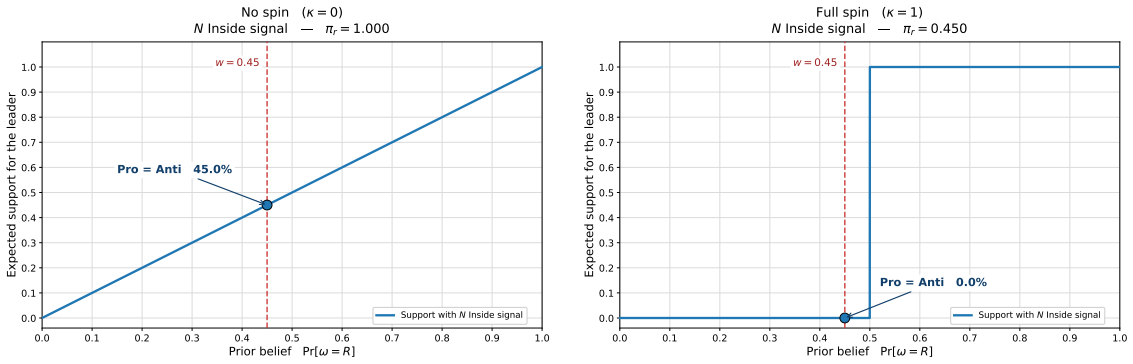
Take the extremes first, with both partisans at N.

No spin ($\kappa = 0$). The signal is fully revealing—r means R, ℓ means L—so a citizen at π supports exactly when the state is R, and $S(\pi) = \pi$. Support at the prior is $w = 0.45$ for every group.

Full spin ($\kappa = 1$). Now r arrives in both states and carries nothing. Beliefs stay put, a citizen supports iff $\pi \geq \frac{1}{2}$, and S is a step at $a(1) = \frac{1}{2}$. The prior falls to the left of it, so support is zero.

The shapes matter more than the numbers. A line is both concave and convex, so it is its own pair of envelopes: under full revelation no split of belief helps either partisan, and both are pinned at w. Full revelation is the one design against which there is nothing to answer. A step is neither, so a split helps whichever partisan wants to cross it: at full spin the envelopes at w are $2w = 0.90$ and 0, the widest partisan gap any design attains. Everything else interpolates between the two.

Figure 3: The two extremes, without the optimal use of Inside media. Panel (a): at zero spin S is the 45° line and every group supports with probability w. Panel (b): at full spin S is a step at $\frac{1}{2}$, and since $w < \frac{1}{2}$ support is zero. A line is its own pair of envelopes; a step is as far from either as a design can get.



High spin ($\kappa = 0.7$), without optimal Inside media. Good news is now frequent, $\Pr[s = r] = 0.835$, and barely credible, $\pi_r = 0.539$. The jump in S has migrated up to $a(0.7) = 0.412$, just under the prior, which therefore clears it with little room to spare: support is 0.835, close to double w, and it comes from both partisan types. Structure N prices only one of the two effects of spin. A citizen who consumes no Inside signal cares whether π_r clears $\frac{1}{2}$, not by how much, so by (13) weakening good news costs the leader

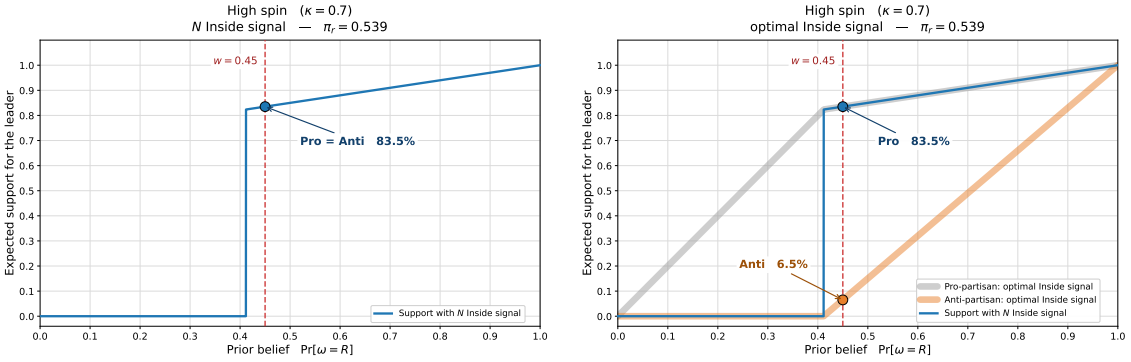
nothing until the escape threshold reaches the prior at $\kappa = \kappa_{\text{ret}} = 0.818$; at $\kappa = 0.7$ she is nearly there.

High spin, with optimal Inside media. The two answers could hardly be more different, and Subsection 4.1 said why. The pro-partisan has nothing to answer—her unfavorable message is beyond rescue and her favorable one already preserves her—so she plays N, the concave envelope meets the curve at w, and her support stays at 0.835. The anti-partisan's GT answer cuts hers from 0.835 to 0.065, and the geometry prices it. Her structure leaves her at belief exactly $a(0.7) = 0.412$ with probability 0.935—the point at which r leaves her exactly indifferent, where her tie-break saves her—and concedes the state with the remaining 0.065, so her support is the weight she must place on that concession,

$$S^A = \frac{w - a(\kappa)}{1 - a(\kappa)} = w - (1 - w)\kappa, \tag{16}$$

which is the convex envelope in (15) evaluated at w, in agreement with (10). Here the concession is cheap: $a(0.7)$ sits 0.038 below the prior, so she moves her belief barely at all to become immune, and the Martingale property charges her almost nothing for it. The two supports sum to $0.90 = 2w$, as Lemma 1 requires. High spin created no support. It moved it.

Figure 4: High spin, $\kappa = 0.7$. Panel (a): without the optimal use of Inside media the design carries both partisan types at 0.835. Panel (b): with it, the pro-partisan is unmoved and plays N, while the anti-partisan's GT answer cuts her to 0.065. The two still sum to $2w$.



Low spin ($\kappa = 0.2$). Good news is strong, $\pi_r = 0.804$, and rare, $\Pr[s = r] = 0.56$. Without the optimal use of Inside media both partisans support with probability 0.56—worse for the leader than high spin, and by a wide margin. With it, the pro-partisan again plays N and again holds 0.56; what changes is the opposition, and it falls only to 0.34. The escape

threshold $a(0.2) = 0.167$ now lies far below the prior, so escaping a genuinely informative r means moving her belief a long way down, and by (16) the Martingale property prices that curation at a 34% chance of conceding outright. Strong good news is expensive to escape. Weak good news is not.

Figure 5: Low spin, $\kappa = 0.2$. Panel (a): without the optimal use of Inside media both partisans support with probability 0.56. Panel (b): with it, the pro-partisan again plays N, while the anti-partisan must move her belief much further to escape r and pays a 34% chance of conceding for it.

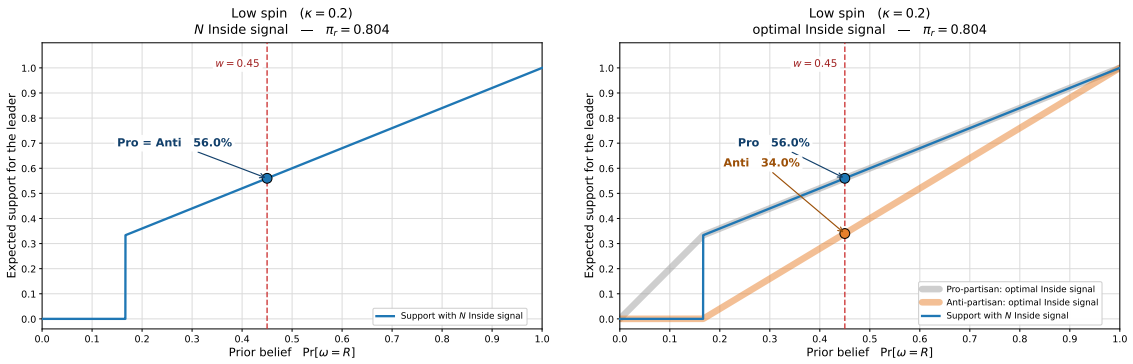


Table 1: The four designs at $w = 0.45$. Without the optimal use of Inside media both partisan types support with the same probability, and it is the reach of good news whenever the prior clears the escape threshold, $w \geq a(\kappa)$. With it, $S^P + S^A = 2w = 0.90$ in every row. The last row is the only one in which $a(\kappa)$ exceeds the prior, and it is the only one in which the two columns differ for the pro-partisan.

				without	with optimal Inside media	
	π_r	$a(\kappa)$	$\Pr[s = r]$	$S^P = S^A$	S^P	S^A
No spin, $\kappa = 0$	1.000	0.000	0.450	0.450	0.450	0.450
Low spin, $\kappa = 0.2$	0.804	0.167	0.560	0.560	0.560	0.340
High spin, $\kappa = 0.7$	0.539	0.412	0.835	0.835	0.835	0.065
Full spin, $\kappa = 1$	0.450	0.500	1.000	0.000	0.900	0.000

The four designs exhibit three patterns. First, the gain from Inside media is one-sided. Wherever $w \geq a(\kappa)$ the pro-partisan plays N and her support is simply the reach of good news, w/π_r; all of the action is on the other side, where by (16) the cost of the GT answer is governed entirely by the distance from the prior down to the escape threshold. Second, spin transfers what it cannot create: $S^P + S^A = 2w$ at every design, so moving from low to high spin buys the base 0.275 and costs the opposition exactly 0.275. Facing partisans alone the leader is not persuading anyone in aggregate; she is choosing which camp to

serve, which Remark 4 makes general. Third, the two extremes are the single-audience optima of Proposition 2. Facing only the base she wants $\pi_r \leq \frac{1}{2}$—any $\kappa \geq \kappa_{\text{ret}}$, full spin included—because the base's own Inside media then manufactures $\{0, \frac{1}{2}\}$ and delivers the cap $2w$: she can stop supplying credibility because the base supplies it for her. Facing only the opposition she wants full revelation, the one curve that is its own convex envelope and hence the one design admitting no answer. What lies between is settled by the independents, who sit on no budget of this kind and whose optimum is interior, at $\pi_r = \frac{3}{4}$.

Partisanship and identity preservation: an aside It is worth being precise about what kind of partisan this is, because the model's central claim depends on it. In the standard formulation a partisan votes for her side regardless of what she comes to believe: the bias sits in her utility over outcomes, in the location of an ideal point, or in a behavioral type, and her vote is therefore a constant. Here the bias sits one step earlier, in the beliefs she *wants to hold*, and her support is whatever her posterior turns out to license. She may fail. Table 1 shows her failing: at $\kappa = 0.7$ the opposition still hands the leader 6.5% of its support, at $\kappa = 0.2$ fully 34%, and under full revelation both camps converge on w and the partisan media market falls silent. Every group's support moves with the state and with the design. These partisans are persuadable.

What identity buys them is not a vote bank but a systematically tilted information diet, and that distinction is what makes the paper's headline result possible: the cross-sectional ordering $S^P > S^I > S^A$ that emerges below is not an assumption about preferences over candidates but an equilibrium consequence of three rational agents choosing different evidence. Polarization here is an outcome of the information market, not an input to it.

4.3 The leader's problem

The leader aims to maximize the ex-ante expected support from all citizens:

$$V \;=\; p_I\, S^I \;+\; p_P\, S^P \;+\; p_A\, S^A \tag{17}$$

where S^j is the probability that a type-j citizen supports her, given by (9), (11) and (10).

Using π_r as the design variable, the prior factors out of the problem entirely: for $\pi_r \geq \frac{1}{2}$,

$$V(\pi_r; w, p) \;=\; w\, g(\pi_r; p), \qquad g(\pi; p) \;=\; p_I\, \pi(3 - 2\pi) \;+\; \frac{p_P - p_A}{\pi} \;+\; 2p_A, \tag{18}$$

while on $[w, \frac{1}{2})$ the pro-partisan cap binds and $g(\pi; p) = p_I\, \pi(3 - 2\pi) + 2p_P$. The marginal

value of stronger good news is therefore

$$g'(\pi) \;=\; \underbrace{p_I\,(3-4\pi)}_{\text{independent effect}} \;-\; \underbrace{\frac{p_P - p_A}{\pi^2}}_{\text{net partisan effect}} \quad . \tag{19}$$

Everything in the baseline follows from the shape of g. Note already that w scales the leader's value but does not appear in g: the prior cannot affect what she says, only how much she collects for saying it.

Benchmark audiences We first ask how the leader would persuade a single citizen type, by setting two of the three shares to zero. Each benchmark isolates one of the three forces in (19).

Proposition 2 (Single audiences) *Fix $w \in (0, \frac{1}{2}]$. Facing a single audience, the leader's optimal spin and resulting support are as follows.*

(i) Independents only *($p_I = 1$). The unique optimum is $\pi_r^* = \frac{3}{4}$, which implies $\kappa^* = \frac{w}{3(1-w)}$, and $V^* = \frac{9}{8}\,w$. The leader splits the prior into the posteriors $\{0, \frac{3}{4}\}$.*

(ii) Pro-partisans only *($p_P = 1$). Any design with $\pi_r \leq \frac{1}{2}$—equivalently any spin $\kappa \geq \kappa_{\text{ret}}$, including full spin $\kappa = 1$—attains the cap $V^* = 2w$. Once her signal is that uninformative, the base's own Inside media manufactures the posteriors $\{0, \frac{1}{2}\}$, so the optimum is not unique.*

(iii) Anti-partisans only *($p_A = 1$). Zero spin $\kappa^* = 0$ (full revelation) is the unique optimum, with $V^* = w$. Any positive spin induces $\pi_r \in (\frac{1}{2}, 1)$ and is met by the GT counter-design, which lets anti-partisans hold their preferred beliefs more often and so strictly reduces support.*

The proof is in Appendix C. The benchmark values order the audiences endogenously,

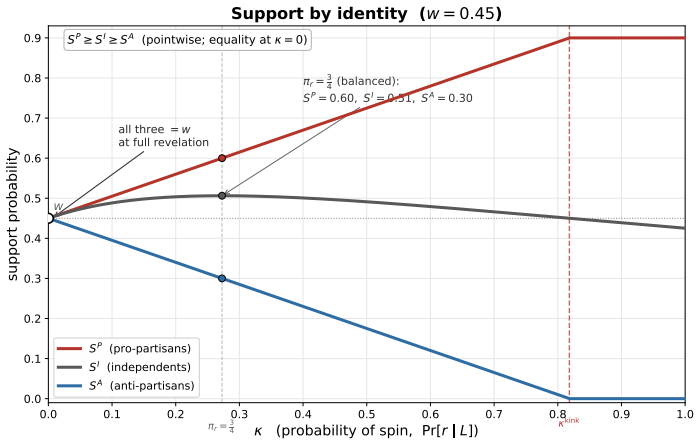
$$\underbrace{2w}_{\text{base}} \;>\; \underbrace{\tfrac{9}{8}w}_{\text{independents}} \;>\; \underbrace{w}_{\text{opposition}} \;,$$

for every $w \in (0, \frac{1}{2}]$: the base offers the most support, the opposition the least, the independents in between. The three optima are also the three corners of the partition that follows.

The full electorate With all three audiences present the leader faces a single one-dimensional trade-off, read off (19): raising the strength of good news buys independents while the target is below the tangency $\frac{3}{4}$ and loses them above it, and it moves support from the base to the opposition throughout. Figure 6 shows the three curves. Partisan

supports are mirror images around w and sum to $2w$ at every design, so the partisan terms enter only through the imbalance $p_P - p_A$; independent support is the only term with an interior peak.

Figure 6: Support by identity, as a function of the spin κ. Pro-partisan support and anti-partisan support are mirror images around w and sum to $2w$ at every design; independent support peaks at the tangency $\pi_r = \frac{3}{4}$.



The three regimes are the three ways this trade-off can resolve. When the two partisan camps are of equal size they cancel, and the leader targets independents exactly as if they were her only audience. Tilt the coalition toward the base and she raises spin; tilt it toward the opposition and she lowers it. At each end the interior solution runs out: once the base leads by enough, she stops at the point where the escape threshold meets the prior and holds the base whole; once the opposition is large enough, counter-design has made every persuasive design unprofitable and only truth survives.

Proposition 3 (Optimal spin) *Fix $w \in (0, \frac{1}{2}]$ and $(p_I, p_P, p_A) \in \Delta^2$. When $p_I > 0$ the optimal design is unique in π_r; when $p_I = 0$ the retention optimum is attained at every $\pi_r \leq \frac{1}{2}$ (Proposition 2(ii)). In either case the simplex partitions into three regimes, and in each the optimal target is implemented by the spin*

$$\kappa^* = \frac{w}{1-w} \cdot \frac{1-\pi_r^*}{\pi_r^*}. \tag{20}$$

(i) Pro-partisan retention, *optimal when $p_P - p_A \geq \frac{p_I}{4}$. The target is $\pi_r^* = \frac{1}{2}$, so $\kappa^* = \kappa_{\text{ret}}$: this is the design at which the escape threshold meets the prior, $a(\kappa^*) = w$. Both partisans play N, and*

$$V^* = w\,(p_I + 2p_P), \qquad (S^I, S^P, S^A) = (w,\, 2w,\, 0).$$

(ii) Independent targeting, *optimal when* $-p_I < p_P - p_A < \frac{p_I}{4}$. *The target* π_r^* *is the unique root in* $(\frac{1}{2}, 1)$ *of*

$$p_I\, \pi_r^2\, (3 - 4\pi_r) \;=\; p_P - p_A, \tag{21}$$

implemented by a $\kappa^* \in (0, \kappa_{\text{ret}})$. *The pro-partisan plays N, the anti-partisan plays GT, and*

$$(S^I, S^P, S^A) = \Big(w\, \pi_r^*(3 - 2\pi_r^*),\;\; \frac{w}{\pi_r^*},\;\; w\Big(2 - \frac{1}{\pi_r^*}\Big)\Big),$$
$$V^* = 2w\big[\, 3p_I\, \pi_r^*(1 - \pi_r^*) + p_A\,\big].$$

(iii) Full revelation, *optimal when* $p_P - p_A \le -p_I$, *equivalently* $p_A \ge \frac{1}{2}$. *The target is* $\pi_r^* = 1$, *that is* $\kappa^* = 0$. *All three groups learn the state, so* $V^* = w$ *and* $(S^I, S^P, S^A) = (w, w, w)$.

The proof is in Appendix D. Since $\pi \mapsto \pi^2(3 - 4\pi)$ is strictly decreasing on $(\frac{1}{2}, 1)$, with values $\frac{1}{4}$, 0 and -1 at $\pi = \frac{1}{2}$, $\frac{3}{4}$ and 1, the target in regime (ii) falls continuously as $p_P - p_A$ rises, sweeps the whole interval $(\frac{1}{2}, 1)$, and passes through the tangency $\frac{3}{4}$ exactly when the two camps are of equal size.

The simplex forms of the two boundaries deserve emphasis. Full revelation becomes optimal precisely when anti-partisans are a *majority*—not merely when they outnumber the base—because the presence of independents keeps some spin profitable. In the retention regime the opposite holds: further spin would cost independent support without adding anything on either partisan side, since the base is already capped at $2w$ and the opposition already at zero, while less spin would gain independents by less than it costs in net partisan support. Both boundaries, $5p_P - 3p_A = 1$ and $p_A = \frac{1}{2}$, are straight lines in (p_P, p_A), and, as the next result shows, they do not move with the prior.

Remark 4 (Partisans only: a zero-sum budget) *With no independents* ($p_I = 0$), *the two partisan supports always add to* $2w$, *whatever spin the leader picks: by Lemma 1, whatever she gains from one camp she loses from the other, so partisan persuasion is a pure transfer from base to opposition. Her problem then collapses to backing the larger camp—pro-retention worth* $2w\, p_P$ *when the base holds the majority, full revelation worth* w *otherwise—and no design lifts combined partisan support above* $2w$. *Getting past that cap takes a second instrument, which is what Section 5 adds.*

4.4 Composition steers, fundamentals scale

The regimes in Proposition 3 are cut out by the composition (p_I, p_P, p_A) alone; the prior w never enters the conditions. The prior is the model's stand-in for fundamentals, and it turns out to set how much support the leader collects, not what she does to collect it.

Proposition 5 (Prior-invariant communication) *For every composition $p \in \Delta^2$ and all $w \in (0, \frac{1}{2}]$:*

(i) The optimal target π_r^*, the regime, and the regime boundaries of Proposition 3 do not depend on w; when $p_I > 0$ they depend on the composition only through the imbalance $(p_P - p_A)/p_I$.

(ii) Every group's support, the total V^*, and the probability that the leader spins, $\Pr[s = r,\ \omega = L] = (1 - w)\,\kappa^*$, are all proportional to w.

(iii) Support ratios and support shares are therefore independent of w. The one object that does move with the prior is the spin itself, $\kappa^* = \frac{w}{1-w} \cdot \frac{1-\pi_r^*}{\pi_r^*}$, which rises with w.

The proof is in Appendix D.

Worse fundamentals scale every group's support down by the same factor and change nothing else. The leader aims at the same posterior, stays in the same regime, and leaves the whole cross-section of polarization intact; she only reaches that posterior by spinning less often. Good news becomes rarer, not weaker. Two leaders who face very different fundamentals but similar electorates should therefore communicate the same way, and differ only in how frequently they spin.

The same point, run in reverse, is a test. If a leader's target—not merely how often she spins—moves with fundamentals, the baseline cannot produce it. Such a response is the mark of underplayed good news, the extension of Section 5, where the invariance breaks, and does so exactly when $w > \frac{1}{4}$.

None of this rests on the particular functional forms. Because admitted bad news is conclusive, the message ℓ is worthless to every audience, so all of the leader's value arrives through the favorable message, which by (7) shows up with probability w/π_r. The prior enters there and nowhere else, as a multiplicative reach factor, while the menu of posterior strengths she can aim at is prior-free.

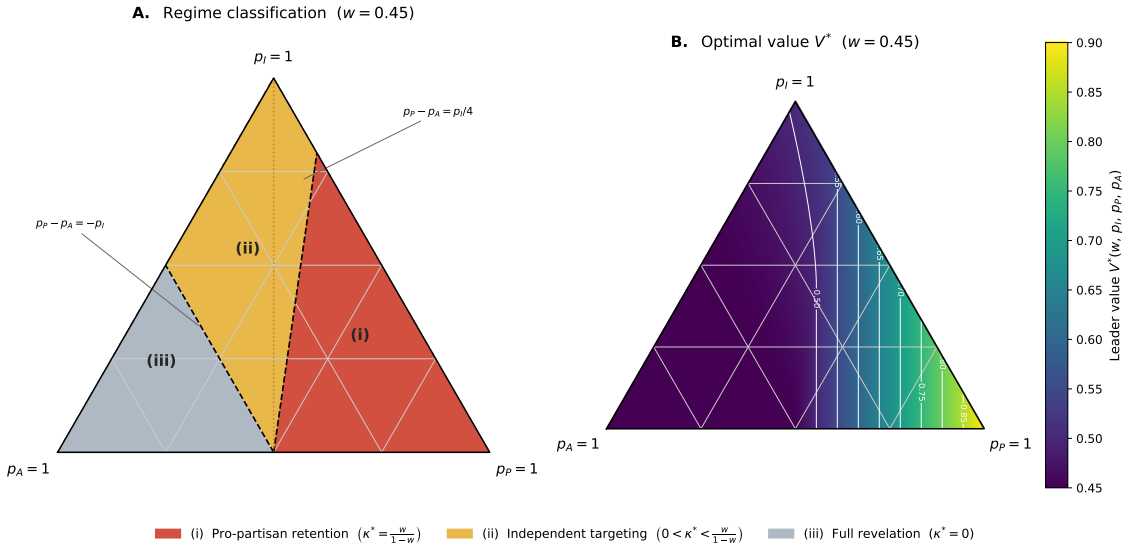
5 Underplayed good news: the two-sided design

The baseline rests on two idealizations: the leader always announces her successes, and she can commit to whatever she announces. This section drops the first; Section 6 drops the second. Neither changes how audiences reason—only what the leader can reach—and each breaks the prior-invariance of Proposition 5, in opposite directions.

The baseline lets the leader dress up a failure but never mute a success. Now let her do both. Alongside her spin κ she chooses an *underplay* rate μ—the probability that she reports a genuine success as a failure—and commits to the pair (κ, μ),

$$\Pr[s = r \mid \omega = R] \;=\; 1 - \mu, \qquad \Pr[s = r \mid \omega = L] \;=\; \kappa, \tag{22}$$

Figure 7: The baseline regime map at $w = 0.45$. Panel (a): the three regimes of Proposition 3 partition the simplex along the two straight lines $5p_P - 3p_A = 1$ and $p_A = \frac{1}{2}$; by Proposition 5 the partition is the same for every prior $w \in (0, \frac{1}{2}]$. Panel (b): the corresponding optimal value V^* over the simplex, which scales with the prior while the partition does not.



with $\kappa + \mu \leq 1$ so that r stays good news: her two lie rates cannot sum past one. The message is uninformative when $\kappa + \mu = 1$ and fully revealing when $\kappa = \mu = 0$; the baseline of Section 4 is the slice $\mu = 0$, where she spins but never underplays. A citizen who sees only r, or only ℓ, now holds the interim posterior

$$\pi_r(\kappa, \mu) = \frac{w(1-\mu)}{w(1-\mu) + (1-w)\kappa}, \qquad \pi_\ell(\kappa, \mu) = \frac{w\mu}{w\mu + (1-w)(1-\kappa)}, \tag{23}$$

and the favorable message arrives with probability $\Pr[r \mid (\kappa, \mu)] = w(1-\mu) + (1-w)\kappa$. Her problem is

$$\max_{(\kappa,\mu)\in[0,1]^2,\, \kappa+\mu\leq 1} V(w, \kappa, \mu;\, p_I, p_P, p_A), \qquad V \;=\; p_I\, S^I \;+\; p_P\, S^P \;+\; p_A\, S^A, \tag{24}$$

with independent support the response function h evaluated at whichever posterior each message induces,

$$S^I(w, \kappa, \mu) \;=\; \Pr[r \mid (\kappa, \mu)]\, h\big(\pi_r(\kappa, \mu)\big) \;+\; \big(1 - \Pr[r \mid (\kappa, \mu)]\big)\, h\big(\pi_\ell(\kappa, \mu)\big). \tag{25}$$

Muting your own successes looks perverse, and for a hostile enough prior it is. For

$w \leq \frac{1}{4}$ the baseline is without loss: every two-sided design is weakly dominated by a spin-only design (Lemma 26), so Propositions 3 and 5 are the whole story and Figure 7 is the complete map. For $w > \frac{1}{4}$ underplay earns its keep, and the rest of this section shows how.

Two escape thresholds The mechanism is easiest to see in the belief-space language of Subsection 4.2. With both messages live, the single escape threshold $a(\kappa)$ of (12) becomes a pair. The favorable message tips a citizen over one-half only if her belief already clears a_r, while surviving the *unfavorable* message requires her belief to clear a_ℓ, where

$$a_r(\kappa, \mu) \;=\; \frac{\kappa}{\kappa + 1 - \mu}, \qquad a_\ell(\kappa, \mu) \;=\; \frac{1 - \kappa}{1 - \kappa + \mu}. \tag{26}$$

On the baseline slice these are $a_r(\kappa, 0) = a(\kappa)$ and $a_\ell(\kappa, 0) = 1$: the unfavorable branch is beyond rescue, which is exactly what made the baseline one-dimensional. Underplay pulls a_ℓ down from 1 into range. And the feasibility constraint has a clean translation: $\kappa + \mu \leq 1$ holds if and only if

$$a_r \;\leq\; \tfrac{1}{2} \;\leq\; a_\ell, \tag{27}$$

with all three equal at the uninformative designs. The two lie rates cannot sum past one because the two escape thresholds must bracket the voting threshold.

Now recall why the order of moves was irrelevant in the baseline (Section 4, Timing): with one live branch, a structure aimed at that branch is also the best reaction to it. With two live branches a partisan's single committed structure must serve both at once. She cannot aim one realization at the message r and a different one at the message ℓ, because the realization arrives before she knows which message it will be paired with. That single friction is the whole extension. The leader exploits it in two directions. *Toward her base:* an ambiguous unfavorable message can be laundered. The pro-partisan's Good realization, paired with ℓ, lands her exactly at $\frac{1}{2}$—possible only when $\pi_\ell > 0$—and the probability that her Inside media preserves her works out to the exact mirror of the baseline reach formula,

$$S^P_{GT} \;=\; \frac{w}{a_\ell(\kappa, \mu)}, \tag{28}$$

just as her baseline support was w/π_r: in the baseline the leader served her base through the strength of good news, here through the ambiguity of bad news. *Toward her opposition:* good news can stay strong while this happens, because π_r and a_ℓ are controlled by two different instruments. The anti-partisan's counter-design must now pay for its protection against a strong r with realizations that are themselves graded against a live ℓ, and the leader tunes the two branches so that her shield and its price cannot both be efficient. The

zero-sum budget $S^P + S^A = 2w$ of Remark 4 gives way exactly when the unfavorable posterior can be parked above $\frac{1}{4}$, which Bayes plausibility ($\pi_\ell < w$) allows only when $w > \frac{1}{4}$. Below we watch it break in numbers.

5.1 The partisans' problem with two live branches

With both messages live, the one-branch splitting of Lemma 1 becomes the full taxonomy of Herrera and Sethi [2025]. Recall the three useful realizations: a *good* one (G), which paired with the partisan's unfavorable message just holds her posterior at $\frac{1}{2}$; a *bad* one (B), redeemable because paired with her favorable message it too lands at $\frac{1}{2}$; and a *terrible* one (T), which concedes the state. Bayes plausibility prices them—the expected posterior must equal the prior, so favorable realizations have to be paired with unfavorable ones. A good realization is valuable but dear, a terrible one is free, and a bad one sits between. Three structures span every case: do nothing (N), pair good with terrible (GT), or pair good with bad (GB).[12]

Which one a partisan picks is a threshold rule in the design. Take the pro-partisan and look at her interim posterior after the public message. If it is already at least $\frac{1}{2}$ after r and at most $\frac{1}{2}$ after ℓ, she has nothing to gain and plays N. Otherwise she weighs the two active structures, choosing GT when the index

$$\Delta_P(\kappa, \mu) \;\equiv\; \kappa + 3\mu - 2\kappa\mu - 1 \tag{29}$$

is nonnegative and GB when it is negative. The anti-partisan's rule is the mirror image, run by $\Delta_A(\kappa, \mu) \equiv 3\kappa + \mu - 2\kappa\mu - 1$—the same index with κ and μ swapped, reflecting the symmetry of the two types. On the baseline slice $\mu = 0$ the unfavorable posterior collapses to $\pi_\ell = 0$, which kills the redeemable realization B as a source of value: since $\Delta_P(\kappa, 0) = \kappa - 1 < 0$, the index selects GB for the pro-partisan, but that structure is degenerate there—whatever her Inside signal reports alongside the conclusive ℓ is payoff-irrelevant, so it induces the same final posteriors $\{0, \pi_r\}$, and the same support w/π_r, as the uninformative N that Lemma 1 reports. The anti-partisan's GT and GB likewise coincide on the slice, both giving $S^A = w \max\{0, 2 - 1/\pi_r\}$, recovering Lemma 1 exactly. Appendix E collects the state-by-state support probabilities and verifies both boundary checks.

[12] N is the special case of GT in which the good realization occurs with probability one. As noted in Subsection 4.1, the menu is exhaustive on the region the leader ever chooses, $\pi_r \geq \frac{1}{2}$; the pairing BT can be optimal for the pro-partisan only when $\pi_r < \frac{1}{2}$, which the domination argument of Appendix G (Step 2) rules out through the Bayes-plausibility cap $S^P \leq 2w$.

5.2 The full electorate

With underplay available, the leader's optimum takes one of two forms. On the family $\mathcal{F}_1$ she leaves underplay off, $\mu = 0$, and concedes the unfavorable state—this is the baseline. On $\mathcal{F}_2$ she sets $\mu > 0$ to drive the anti-partisan into GB. Writing V_1^* and V_2^* for the best she can do on each,

$$V_1^*(w, p_I, p_P, p_A) \equiv \max_{\kappa \in [0,1]} V(w, \kappa, 0; \cdot), \tag{30}$$

$$V_2^*(w, p_I, p_P, p_A) \equiv \max_{(\kappa,\mu) \in \overline{\mathcal{F}_2}} V(w, \kappa, \mu; \cdot), \tag{31}$$

both attained by the continuity and compactness recorded in Lemma 11. Lemmas 21, 22 and 23—in Appendix H—show her optimum is simply the better of the two,

$$V^*(w, p_I, p_P, p_A) = \max\{V_1^*(w, p_I, p_P, p_A),\ V_2^*(w, p_I, p_P, p_A)\}. \tag{32}$$

Since V_1^* is exactly the baseline value of Proposition 3, underplay adds one new regime, and only where V_2^* wins.

Proposition 6 (Optimal communication with the full electorate) *Fix $w \in (0, \frac{1}{2}]$ and $(p_I, p_P, p_A) \in \Delta^2$. The leader's optimum lies in $\mathcal{F}_1$ if $V_1^* \geq V_2^*$ and in $\mathcal{F}_2$ if $V_2^* > V_1^*$.*

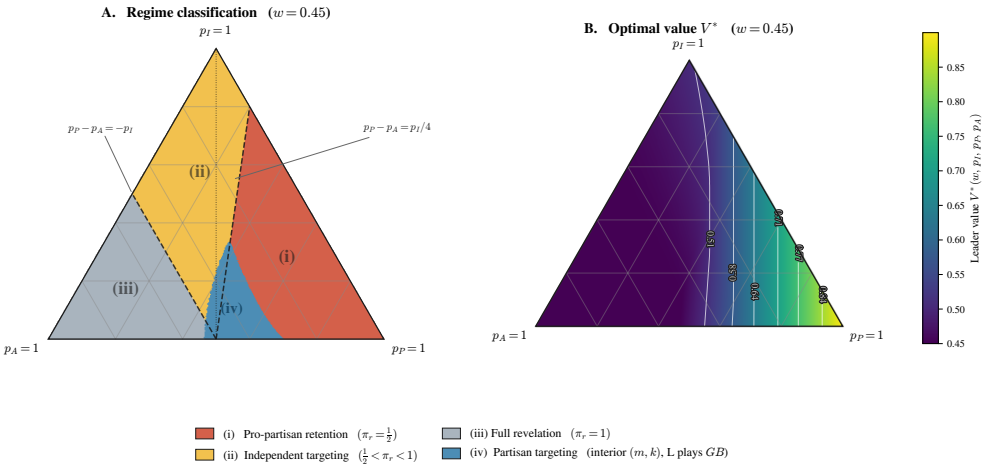
- On $\mathcal{F}_1$, *the optimum is exactly Proposition 3: retention when $p_P - p_A \geq p_I/4$, independent targeting when $-p_I < p_P - p_A < p_I/4$, and full revelation when $p_P - p_A \leq -p_I$. Underplay is unused, $\mu^* = 0$, and prior-invariance holds.*

- On $\mathcal{F}_2$, *the optimum is a fourth regime,* partisan targeting*: it has $\pi_r(\kappa^*, \mu^*) > \frac{1}{2}$, $\Delta_P(\kappa^*, \mu^*) > 0$, and $\Delta_A(\kappa^*, \mu^*) < 0$, so the pro-partisan plays GT and the anti-partisan plays GB. The optimal design is characterized exactly by the first-order system of Lemma 23(i). That system has no closed-form solution in the population shares, but its* inverse *does: every critical design corresponds to shares given in closed form by (83), and the optimum is the value-maximizing design on that stationary locus. This regime is nonempty if and only if $w > \frac{1}{4}$: Lemma 26 gives the "only if" and Proposition 7(ii) the "if".*

The proof is in Appendix G.

Figure 8 shows the four regimes at $w = 0.45$. The two dashed lines are the baseline cuts $p_P - p_A = p_I/4$ and $p_P - p_A = -p_I$, which do not move with the prior. The solid navy curve is the partisan-targeting boundary $V_2^* = V_1^*$: it is the one boundary that does move with the prior, absent at $w \leq \frac{1}{4}$ and expanding as w rises.

Figure 8: Optimal communication regimes at $w = 0.45$ over the population simplex. Panel A displays the four regimes of Proposition 6. Panel B indicates the ex-ante expected support V^* for the leader.



Each regime is the footprint of whichever benchmark force wins the leader's calculus. In retention, the base wins: she stops at the threshold to keep it whole, and the partisan media market falls silent. In independent targeting, the concavification gain wins at the favorable message: she pays a retention cost for sharper independent persuasion, and the anti-partisan answers with GT to neutralize the move at the threshold. In full revelation, the opposition wins outright: counter-design makes any persuasive design unprofitable, and only truth survives. Partisan targeting is the one regime where both partisan structures are active at once: she uses the favorable message to extract independent persuasion and the *unfavorable* message to force the anti-partisan into the costly GB hedge—worthwhile only when the prior is high enough to make the unfavorable message informative. The four-regime partition is not ad hoc; it is these three forces meeting across population mixes.

5.3 Persuading partisans only

The cleanest look at what the second instrument buys comes from turning the independents off. Set $p_I = 0$, so $p_A = 1 - p_P$. As p_P runs from 1 to 0 the electorate travels between the pure-pro and pure-anti benchmarks of Proposition 2: this edge of the simplex isolates the identity trade-off—holding loyalists against poaching opponents—with no persuadable middle in between.

On the baseline the answer is the bang-bang of Remark 4: the leader can never beat $\max\{2w\, p_P,\, w\}$, because every unit of manipulation transfers support one-for-one

from base to opposition. The escape runs through the unfavorable message. Leaving ℓ only partly revealing while keeping r strong restores pro-partisan support—by (28) their Inside media launders an ambiguous unfavorable message, but a conclusive one leaves it nothing to work with—while the strong favorable message still pulls support from a counter-designing anti-partisan.

Proposition 7 (Persuading partisans only) *Fix $p_I = 0$ and define*

$$\overline{p}_P(w) \;\equiv\; \frac{1}{1+8w(1-2w)} \;=\; \frac{1}{2-(4w-1)^2} \;\in\; \left[\tfrac{1}{2},\,1\right] \quad \text{for } w \in [\tfrac{1}{4},\tfrac{1}{2}]. \tag{33}$$

(i) Low priors. *If $w \le \frac{1}{4}$, then $V^*(p_P) = \max\{2w\,p_P,\, w\}$: retention is optimal if $p_P \ge \frac{1}{2}$, full revelation if $p_P \le \frac{1}{2}$.*

(ii) High priors. *If $w \in (\frac{1}{4},\frac{1}{2}]$, there is a $\underline{p}_P(w) \in (0,\frac{1}{2})$ with full revelation optimal iff $p_P \le \underline{p}_P$, partisan targeting* strictly *optimal on $(\underline{p}_P, \overline{p}_P)$, and retention optimal iff $p_P \ge \overline{p}_P$. On the targeting band $V^*(p_P) > \max\{2w\,p_P,\, w\}$ strictly. The thresholds satisfy $\overline{p}_P(\frac{1}{4}) = \frac{1}{2}$, $\overline{p}_P(\frac{1}{2}) = 1$, and $\underline{p}_P(\frac{1}{2}) = (4+3\sqrt{3})/22 \approx 0.418$.*

(iii) The targeting design. *For $w < \frac{1}{2}$ and $p_P \in (\underline{p}_P, \overline{p}_P)$ the optimum is interior to the partisan-targeting cell, at the critical point of Lemma 23(i), whose system collapses at $p_I = 0$, in the coordinates (42), to*

$$b^* = \frac{a^*}{1-2(a^*)^2}, \qquad \frac{p_P}{1-p_P} = \frac{(1+a)^2\,b\,\left(2ab^2+abt-a-b^2t+b+2t\right)}{(1+t)\,(a+b)^3}\,\Bigg|_{(a,b)=(a^*,b^*)}. \tag{34}$$

As $p_P \uparrow \overline{p}_P$ the design degenerates to the uninformative corner; at $w = \frac{1}{2}$ the maximizer migrates to the locus $\Delta_A = 0$ of Lemma 23(ii)–(iii).

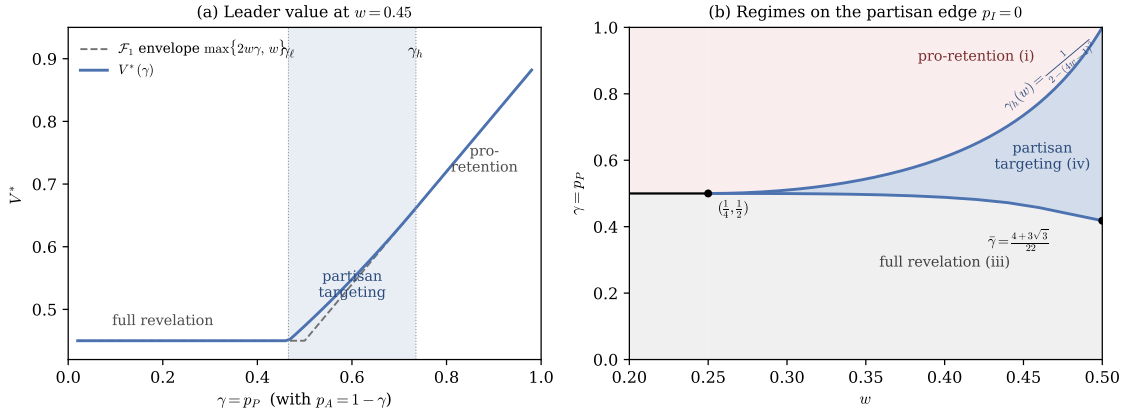
The proof is in Appendix I.

A worked point makes the mechanism concrete. At $w = 0.45$ and $p_P = 0.55$, solving (34) gives $(\kappa^*, \mu^*) \approx (0.14,\, 0.75)$, inducing posteriors $(\pi_\ell, \pi_r) \approx (0.42,\, 0.59)$. Both lies are at work, and the bigger one is underplay: the leader mutes three-quarters of her genuine successes, leaving the unfavorable message nearly uninformative (π_ℓ just below the prior) while good news stays moderately strong. The supports are $S^P \approx 0.845$ and $S^A \approx 0.100$, so combined partisan support is

$$S^P + S^A \;\approx\; 0.945 \;>\; 0.90 \;=\; 2w\,:$$

the zero-sum budget of Remark 4, which held at *every* spin-only design, is broken. The leader's value is $V^* \approx 0.510$ against the baseline envelope 0.495. In the belief-space

Figure 9: Persuading partisans only. Panel (a): the leader's value $V^*(p_P)$ at $w = 0.45$ coincides with the baseline envelope $\max\{2w\, p_P,\, w\}$ (dashed) outside the targeting band $(\underline{p}_P, \overline{p}_P) \approx (0.466,\, 0.735)$, and rises strictly above it inside. Panel (b): the regime map on the partisan edge $p_I = 0$. The targeting wedge opens at $(w, p_P) = (\frac{1}{4}, \frac{1}{2})$; its upper boundary is $\overline{p}_P(w) = 1/(2-(4w-1)^2)$ of (33), and its lower boundary reaches $\underline{p}_P(\frac{1}{2}) = (4+3\sqrt{3})/22$ at $w = \frac{1}{2}$.



accounting of (28), the design parks the unfavorable escape threshold at $a_\ell \approx 0.53$, so the base is preserved with probability $w/a_\ell \approx 0.845$—her Inside media lifts her from the prior to a_ℓ cheaply, because the nearly uninformative ℓ leaves that distance short.

Three things stand out. First, the leader's favorable message is always set the same way. Because pro-partisan support under GT depends only on the unfavorable posterior—this is (28), in which π_r does not appear—she picks the favorable posterior purely to maximize extraction from the anti-partisan given π_ℓ: the first equation of (34), which involves neither p_P nor w. The coalition weight and the prior choose only where along this ridge she sits, which is why her value is tangent to the retention line at $\overline{p}_P$ (Figure 9(a)).

Second, $\overline{p}_P$ measures how far targeting reaches into pro-dominant territory, and it grows with the prior. Just above $w = \frac{1}{4}$ the wedge is a sliver around the balanced coalition; at $w = \frac{1}{2}$ retention is never strictly optimal for any interior p_P— with an even prior, even a heavily loyalist coalition invites the leader to work both messages. At $w = 0.45$, $\overline{p}_P = 25/34 \approx 0.735$.

Third, targeting trades a little polarization for reach. Against the retention outcome $(S^P, S^A) = (0.90,\, 0)$, the worked point gives up 0.055 of loyalist support—the base's Inside media can no longer fully launder a now-informative unfavorable message—in exchange for 0.100 pulled from opponents whose counter-design is imperfect against a design that works both messages. The ordering $S^P > S^A$ survives, but the gap narrows from 0.90 to 0.745. Manipulation aimed *across* the aisle is, at the margin, depolarizing.

Remark 8 (Where prior-invariance breaks) *Inside the partisan-targeting wedge—and only there—the leader's strategy depends on the prior. The regime boundary itself moves with w (Figure 8), the optimal posteriors (π_ℓ^*, π_r^*) solve a system in which $t = 1 - 2w$ appears explicitly (34), and the support structure is no longer a fixed multiple of w. The contrast with Proposition 5 gives the model a testable signature:* prior-dependence of communication strategy is the fingerprint of underplayed good news. *A leader who only spins her failures communicates in a way invariant to fundamentals; a leader who also mutes her successes does not.*

6 Extension: weak commitment and ex-post spin

The baseline also assumes the leader can commit fully to the experiment she announces. Public communication is rarely that institutionalized: even after announcing a strategy, a leader may sometimes bury an unfavorable realization and send a favorable message instead. Following Lipnowski et al. [2022], a single parameter $\rho \in [0, 1)$ governs this ex-post spin. She follows her announced structure with probability $1 - \rho$, and with probability ρ she is free to send her preferred message—always r.

Formally, she chooses a *planned* experiment $(\widehat{\kappa}, \widehat{\mu})$ with $\widehat{\kappa} + \widehat{\mu} \leq 1$. After the planned message is drawn, commitment fails with probability ρ and the message is overwritten to r; partisans commit their Inside media knowing ρ and $(\widehat{\kappa}, \widehat{\mu})$ but before the failure shock, which is independent of the state. The *observed* experiment—the one every posterior, support function and best response actually depends on—is therefore

$$\kappa = \rho + (1 - \rho)\widehat{\kappa}, \qquad \mu = (1 - \rho)\widehat{\mu}. \tag{35}$$

Lemma 9 (Reduction to a shrunken menu) *Fix $w \in (0, \frac{1}{2}]$, $(p_I, p_P, p_A) \in \Delta^2$ and $\rho \in [0, 1)$. Under weak commitment the equilibrium solves the full-commitment problem on the restricted feasible set*

$$D_\rho = \left\{(\kappa, \mu) \in [0,1]^2 : \ \kappa + \mu \leq 1, \ \kappa \geq \rho, \ \mu \leq 1 - \rho\right\}, \qquad V_\rho^* = \max_{(\kappa,\mu)\in D_\rho} V(w, \kappa, \mu; \, p), \tag{36}$$

with every support function, cell inequality, and first-order condition of Proposition 6 unchanged. The value V_ρ^* is weakly decreasing in ρ, and any full-commitment optimum with $\kappa^* \geq \rho$ and $\mu^* \leq 1 - \rho$ survives unchanged, implemented by the plan $\widehat{\kappa}^* = (\kappa^* - \rho)/(1 - \rho)$, $\widehat{\mu}^* = \mu^*/(1 - \rho)$.

Proof. The failure shock is independent of the state, so the message the citizens see is generated by the observed experiment (35), and their posteriors, Inside structures and support probabilities are exactly those of Section 5 evaluated at (κ, μ). The map (35) is a

bijection from $\{\widehat{\kappa}+\widehat{\mu}\leq 1\}$ onto D_ρ for every $\rho<1$ (and $\kappa+\mu=\rho+(1-\rho)(\widehat{\kappa}+\widehat{\mu})\leq 1$, so $D_\rho\subseteq\mathcal{D}$), which gives (36). Monotonicity follows from $D_{\rho'}\subseteq D_\rho$ for $\rho'\geq\rho$, and the implementation formulas invert the bijection. ■

Weak commitment therefore introduces no new force: partisans still curate identity-preserving media after seeing the leader's style, independents still respond to the induced posteriors. What shrinks is the leader's reach—which promises she can credibly make. On the spin family the floor $\kappa\geq\rho$ is a cap on credibility,

$$\pi_r \;\leq\; \bar{\pi}_\rho \;\equiv\; \frac{w}{w+(1-w)\rho}, \tag{37}$$

and the next result traces what that cap does to the three benchmark forces of Proposition 2.

Proposition 10 (Erosion of the benchmarks) *Fix $w\in(0,\frac{1}{2}]$ and $\rho\in[0,1)$.*

(i) Pro-partisans only. *$V^*_\rho=2w$ for every ρ: the cap is attained by any feasible design with $\pi_r\leq\frac{1}{2}$, for instance $(\kappa,\mu)=(\max\{\rho,\kappa_{\text{ret}}\},0)$. The base benchmark is untouched by institutional weakness.*

(ii) Anti-partisans only. *$V^*_\rho=w-(1-w)\rho$, uniquely attained at $(\kappa,\mu)=(\rho,0)$—the most transparent feasible design—and strictly decreasing in ρ, reaching zero at $\rho=\kappa_{\text{ret}}$. Full revelation is no longer available: the favorable message pools true successes with the ρ-fraction of buried failures, anti-partisans discount it accordingly, and the leader cannot even secure the revelation floor w.*

(iii) Independents only. *$V^*_\rho=\frac{9}{8}w$ for $\rho\leq\frac{w}{3(1-w)}$, attained at the tangency design $\pi^*_r=\frac{3}{4}$ as under full commitment; for $\rho>\frac{w}{3(1-w)}$ the tangency is out of reach and*

$$V^*_\rho \;=\; w\,\bar{\pi}_\rho\,(3-2\bar{\pi}_\rho),$$

attained at the cap and strictly decreasing in ρ.

(iv) The baseline on $\mathcal{F}_1$. *For a general electorate the spin-family value is Proposition 3 solved on the truncated menu $\pi_r\leq\bar{\pi}_\rho$: the single-peakedness of Lemma 22 clips the optimal target to $\pi^*_{1,\rho}=\min\{\pi^0_1,\bar{\pi}_\rho\}$, so realized and planned spin are*

$$\kappa^*_{1,\rho}=\max\{\kappa^0_1,\rho\},\qquad \widehat{\kappa}^*_{1,\rho}=\max\Big\{\tfrac{\kappa^0_1-\rho}{1-\rho},\,0\Big\}, \tag{38}$$

where (π^0_1,κ^0_1) are the full-commitment target and spin on $\mathcal{F}_1$.

Proof. (i) $(\max\{\rho,\kappa_{\text{ret}}\},0)\in D_\rho$, and it induces $\pi_r\leq\frac{1}{2}$, at which the base's Inside media manufactures the posteriors $\{0,\frac{1}{2}\}$ and delivers the Markov cap $2w$ (Lemma 14(ii)); no

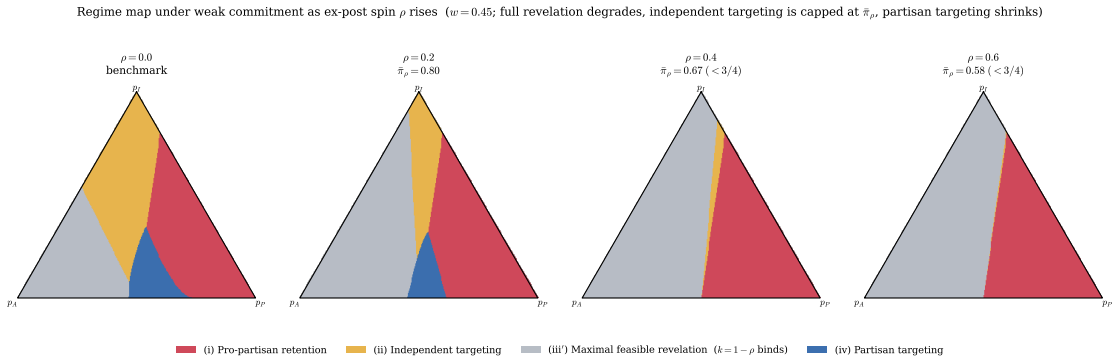
design beats the cap. (ii) The anti-partisan's best response holds the leader to $S^A \leq w(2 - 1/\pi_r)$ at every design (Lemma 14(iii)), which is increasing in π_r; and on D_ρ, $\kappa \geq \rho$ with $\mu \geq 0$ implies $\pi_r \leq \bar{\pi}_\rho$ by (23). Hence $V^*_\rho \leq w(2 - 1/\bar{\pi}_\rho) = w - (1 - w)\rho$, attained at $(\rho, 0)$, where the bound binds and the anti-partisan's GT counter-design is exact. Uniqueness follows as in the full-commitment case: both inequalities bind only at the corner. The value hits zero when $\bar{\pi}_\rho = \frac{1}{2}$, that is $\rho = \kappa_{\text{ret}}$. (iii) The leader's value is $\mathbb{E}[h(\Pi_0)]$ over Bayes-plausible interim posteriors supported in $[0, \bar{\pi}_\rho]$. For $\bar{\pi}_\rho \geq \frac{3}{4}$ the unconstrained concavification $\{0, \frac{3}{4}\}$ is feasible and optimal (Lemma 14(i)); the condition $\bar{\pi}_\rho \geq \frac{3}{4}$ is $\rho \leq \frac{w}{3(1-w)}$. Beyond it, the concave closure of h on the truncated interval is the chord from 0 to $\bar{\pi}_\rho$, since the tangency lies outside; the optimal split is $\{0, \bar{\pi}_\rho\}$, worth $w\, h(\bar{\pi}_\rho)/\bar{\pi}_\rho = w\bar{\pi}_\rho(3 - 2\bar{\pi}_\rho)$. (iv) On $\mathcal{F}_1$ the observed $\mu = 0$ leaves $\mu \leq 1 - \rho$ slack, while $\kappa \geq \rho$ is equivalent to $\pi_r \leq \bar{\pi}_\rho$ by (37); since the $\mathcal{F}_1$ value is single-peaked with peak π^0_1 (Lemma 22), the constraint clips the maximizer to $\min\{\pi^0_1, \bar{\pi}_\rho\}$, and the spin rates follow from inverting (6), with $\kappa(\bar{\pi}_\rho) = \rho$. ■

The proposition has two lessons. The first is distributional: *institutional weakness is free for the base and ruinous against the opposition*. The benchmark ordering $2w > \frac{9}{8}w > w$ of Proposition 2 becomes one fixed term and two eroding ones—the base's $2w$ never moves, because serving the base never required credibility in the first place; the independents' term survives until ρ exceeds $\frac{w}{3(1-w)}$ and erodes thereafter; and the opposition's term erodes from the first unit of ρ, vanishing entirely at $\rho = \kappa_{\text{ret}}$. A leader whose institutions cannot bind her loses nothing at her rallies and everything across the aisle.

The second is a wedge between planned and realized spin. Because audiences expect unfavorable messages to be suppressed, the leader rebuilds credibility by promising a *more* informative planned signal: by (38), planned spin falls as ρ rises. The compensation has a limit. Once the forced spin exceeds what she would have chosen anyway ($\rho > \kappa^0_1$), she can no longer offset it: planned spin hits zero, yet audiences still see favorable messages with probability ρ in state L, so realized spin is floored at ρ however transparent she tries to be. This is the identity-model version of the general lesson—institutions that let a sender occasionally cheat force honest senders to over-invest in informativeness.

For the full electorate, Figure 10 traces the computed regime maps as ρ rises. The cap (37) works through the partition exactly as the benchmarks suggest. Retention is unaffected as long as its design remains feasible, $\rho \leq \kappa_{\text{ret}}$. Full revelation degrades into a maximal-feasible-revelation band on which the floor $\kappa = \rho$ binds: the leader would reveal if she could, and does the most transparent thing her institutions permit. Independent targeting loses access to the concavification tangency once $\rho > \frac{w}{3(1-w)}$, and in the computed maps the partisan-targeting wedge shrinks with ρ and closes at that same threshold. The cap on the strength of good news thus does double duty: it is the anti-partisan counter-

Figure 10: Regime map under weak commitment as ex-post spin ρ rises ($w = 0.45$). As ρ grows the feasible favorable posterior is capped at $\bar{\pi}_\rho = w/(w + (1-w)\rho)$: full revelation degrades into a maximal-feasible-revelation regime (iii′, the band on which the floor $\kappa = \rho$ binds), independent targeting (ii) loses access to the concavification tangency once $\rho > w/(3(1-w)) \approx 0.27$, and in the computed maps partisan targeting (iv) shrinks and closes at that same threshold. Retention (i) is unaffected while $\bar{\pi}_\rho \geq \frac{1}{2}$, that is $\rho \leq \kappa_{\text{ret}}$.



design's work done for free. As ρ rises the opposition has less to neutralize, and the environment collapses toward retention and degraded revelation. Institutional weakness and Inside-media counter-design are substitutes, not complements: the more freely a leader can spin ex post, the less her opponents need to refine ex ante.

Weak commitment is thus the second departure that reintroduces prior-dependence: the cap $\bar{\pi}_\rho = w/(w + (1-w)\rho)$ moves with w, so two leaders facing the same electorate but different fundamentals no longer communicate alike once their institutions are weak. Underplayed good news and imperfect commitment break the invariance of Proposition 5 in opposite directions—the first by widening the leader's reach, the second by narrowing it.

7 Conclusion

This paper offers a simple perspective on a striking feature of contemporary politics: support for political leaders is both highly polarized and remarkably stable within groups, even in the face of substantial changes in fundamentals. We argue that this pattern can be understood as the outcome of an interaction between identity, media choice, and strategic communication.

Political identity shares important features with other forms of group attachment, such as support for a sports team. Individuals derive value from holding beliefs that align with their identity, and are naturally drawn to information environments that reinforce those

beliefs. What distinguishes the current environment is the ease with which such alignment can be sustained. The expansion of media choice—particularly the growth of partisan and segmented information sources—allows individuals to curate their information in ways that preserve identity-consistent views with minimal effort.

Leaders operate within this fragmented information landscape and anticipate how a common public message will be processed across different audiences. While communication is broadcast uniformly, its effects are heterogeneous because citizens filter it through distinct media environments shaped by identity. The leader therefore chooses messages with these differential responses in mind, trading off how the same signal will be amplified by supporters and counteracted by opponents. In settings where supporters are embedded in reinforcing media ecosystems, maintaining their support requires little informational content, while attempts to persuade other groups are constrained by the expectation that they will access countervailing information. Communication and media choice thus form a mutually reinforcing system.

This perspective helps rationalize why approval ratings exhibit both persistence and sharp cross-group differences across a wide range of political contexts. The same pattern is visible in contemporary democracies, such as the United States under Donald Trump and Brazil under Jair Bolsonaro, as well as in more centralized systems under leaders such as Recep Tayyip Erdoğan and Viktor Orbán. In each case, supporters and opponents consume systematically different information, and the same public messages generate sharply divergent interpretations.

More broadly, the analysis suggests that polarization is not simply a reflection of disagreement over policies or fundamentals, but an equilibrium outcome shaped by identity and the structure of the media environment. As access to identity-consistent information expands, the conditions sustaining polarized and resilient support become more prevalent, even in formally competitive democracies. Understanding how these forces interact is therefore central to understanding modern political communication.

In this sense, polarization is not a failure of information, but a consequence of how it is endogenously selected, filtered, and used.

References

Avidit Acharya, Matthew Blackwell, and Maya Sen. Explaining preferences from behavior: A cognitive dissonance approach. *The Journal of Politics*, 80(2):400–411, 2018.

Maja Adena, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhu-

ravskaya. Radio and the rise of the Nazis in prewar Germany. *Quarterly Journal of Economics*, 130(4):1885–1939, 2015.

George A Akerlof and William T Dickens. The economic consequences of cognitive dissonance. *The American economic review*, 72(3):307–319, 1982.

George A Akerlof and Rachel E Kranton. Economics and identity. *The quarterly journal of economics*, 115(3):715–753, 2000.

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

Itai Arieli and Yakov Babichenko. Private Bayesian persuasion. *Journal of Economic Theory*, 182:185–217, 2019.

Arjada Bardhi and Yingni Guo. Modes of persuasion toward unanimous consent. *Theoretical Economics*, 13(3):1111–1149, 2018.

Roland Bénabou and Jean Tirole. Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64, 2016.

Dirk Bergemann and Stephen Morris. Information design, Bayesian persuasion, and Bayes correlated equilibrium. *American Economic Review*, 106(5):586–591, 2016.

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

Dan Bernhardt, Stefan Krasa, and Mattias Polborn. Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5-6):1092–1104, 2008.

Tito Boeri, Nina Nikiforova, and Guido Tabellini. Do elections moderate or polarize political rhetoric? Discussion Paper 21317, CEPR, 2026.

Giampaolo Bonomi, Nicola Gennaioli, and Guido Tabellini. Identity, beliefs, and political conflict. *The Quarterly Journal of Economics*, 136(4):2371–2411, 2021.

Levi Boxell, Matthew Gentzkow, and Jesse M. Shapiro. Greater internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*, 114(40):10612–10617, 2017.

Markus K Brunnermeier and Jonathan A Parker. Optimal expectations. *American Economic Review*, 95(4):1092–1118, 2005.

Andrew Caplin and John Leahy. Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1):55–79, 2001.

Jimmy Chan and Wing Suen. A spatial theory of news consumption and electoral competition. *The Review of Economic Studies*, 75(3):699–728, 2008.

Jimmy Chan, Seher Gupta, Fei Li, and Yun Wang. Pivotal persuasion. *Journal of Economic Theory*, 180:178–202, 2019.

Leonardo D'Amico and Guido Tabellini. Disengaging from reality: Online behavior and unpleasant political news. Working paper, Bocconi University, 2025.

Stefano DellaVigna and Ethan Kaplan. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.

Ruben Durante, Paolo Pinotti, and Andrea Tesei. The political legacy of entertainment TV. *American Economic Review*, 109(7):2497–2530, 2019.

Chris Edmond. Information manipulation, coordination, and regime change. *Review of Economic studies*, 80(4):1422–1458, 2013.

Georgy Egorov, Sergei Guriev, and Konstantin Sonin. Why resource-poor dictators allow freer media: A theory and evidence from panel data. *American Political Science Review*, 103(4):645–668, 2009.

Kfir Eliaz and Ran Spiegler. Can anticipatory feelings explain anomalous choices of information sources? *Games and Economic Behavior*, 56(1):87–104, 2006.

Ruben Enikolopov, Maria Petrova, and Ekaterina Zhuravskaya. Media and political persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–3285, 2011.

Seth Flaxman, Sharad Goel, and Justin M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016.

Stefano Gagliarducci, Massimiliano Gaetano Onorato, Francesco Sobbrio, and Guido Tabellini. War of the waves: Radio and resistance during world war II. *American Economic Journal: Applied Economics*, 12(4):1–38, 2020.

Scott Gehlbach and Konstantin Sonin. Government control of the media. *Journal of public Economics*, 118:163–171, 2014.

Scott Gehlbach, Konstantin Sonin, and Milan W. Svolik. Formal models of nondemocratic politics. *Annual Review of Political Science*, 19:565–584, 2016.

Nicola Gennaioli and Guido Tabellini. Presidential address: Identity politics. *Econometrica*, 93(6):1937–1967, 2025.

Nicola Gennaioli, Frederik Schwerter, and Guido Tabellini. "us vs them": Salient conflict and belief polarization. Discussion Paper 21324, CEPR, 2026.

Matthew Gentzkow and Jesse M Shapiro. Media bias and reputation. *Journal of political Economy*, 114(2):280–316, 2006.

Matthew Gentzkow and Jesse M. Shapiro. Ideological segregation online and offline. *Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

A Arda Gitmez and Pooya Molavi. Informational autocrats, diverse societies. *arXiv preprint arXiv:2203.12698*, 2022.

Edward L. Glaeser, Giacomo A. M. Ponzetto, and Jesse M. Shapiro. Strategic extremism: Why Republicans and Democrats divide on religious values. *The Quarterly Journal of Economics*, 120(4):1283–1330, 2005.

Sergei Guriev and Elias Papaioannou. The political economy of populism. *Journal of Economic Literature*, 60(3):753–832, 2022.

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

Sergei Guriev and Daniel Treisman. A theory of informational autocracy. *Journal of Public Economics*, 186:104158, 2020.

Yosh Halberstam and Brian Knight. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143: 73–88, 2016.

Helios Herrera and Ravideep Sethi. Identity-based elections. *Journal of Political Economy Microeconomics*, 2025. doi: 10.1086/740209. Forthcoming.

Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22:129–146, 2019.

David A Jones. Why americans don't trust the media: A preliminary analysis. *Harvard international journal of press/politics*, 9(2):60–75, 2004.

Paula Kaanders, Pradyumna Sepulveda, Tomas Folke, Pietro Ortoleva, and Benedetto De Martino. Humans actively sample evidence to support prior beliefs. *Elife*, 11:e71768, 2022.

DM Kahan. Misconceptions, misinformation, and the logic of identity-protective cognition. *Yale Law & Economics Research Paper*, (164), 2017.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

Anton Kolotilin. Optimal information disclosure: A linear programming approach. *Theoretical Economics*, 13(2):607–635, 2018.

Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. Persuasion of a privately informed receiver. *Econometrica*, 85(6):1949–1964, 2017.

Anton Kolotilin, Timofiy Mylovanov, and Andriy Zapechelnyuk. Censorship as optimal persuasion. *Theoretical Economics*, 17(2):561–585, 2022.

Botond Köszegi. Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4):673–707, 2006.

Elliot Lipnowski and Laurent Mathevet. Disclosure to a psychological audience. *American Economic Journal: Microeconomics*, 10(4):67–93, 2018.

Elliot Lipnowski, Doron Ravid, and Denis Shishkin. Persuasion via weak institutions. *Journal of Political Economy*, 130(10):2705–2730, 2022. doi: 10.1086/720462.

Andrew T. Little. The distortion of related beliefs. *American Journal of Political Science*, 63 (3):675–689, 2019.

Marco Manacorda, Guido Tabellini, and Andrea Tesei. Mobile internet and the rise of communitarian politics. *The Review of Economic Studies*, 2026. Forthcoming.

Gregory J Martin and Ali Yurukoglu. Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99, 2017.

Lilliana Mason. *Uncivil agreement: How politics became our identity*. University of Chicago Press, 2018.

Filip Matějka and Guido Tabellini. Electoral competition with rationally inattentive voters. *Journal of the European Economic Association*, 19(3):1899–1935, 2021.

Ludmila Matysková and Alfonso Montes. Bayesian persuasion with costly information acquisition. *Journal of Economic Theory*, 211:105678, 2023.

Sendhil Mullainathan and Andrei Shleifer. The market for news. *American economic review*, 95(4):1031–1053, 2005.

Jacopo Perego and Sevgi Yuksel. Media competition and social disagreement. *Econometrica*, 90(1):223–265, 2022.

Dante J Scala and Kenneth M Johnson. Political polarization along the rural-urban continuum? the geography of the presidential vote, 2000–2016. *The ANNALS of the American Academy of Political and Social Science*, 672(1):162–184, 2017.

Peter Schwardmann, Egon Tripodi, and Joël J Van der Weele. Self-persuasion: Evidence from field experiments at international debating competitions. *American Economic Review*, 112(4):1118–1146, 2022.

Mehdi Shadmehr and Dan Bernhardt. State censorship. *American Economic Journal: Microeconomics*, 7(2):280–307, 2015.

David Yanagizawa-Drott. Propaganda and conflict: Evidence from the Rwandan genocide. *Quarterly Journal of Economics*, 129(4):1947–1994, 2014.

The appendices are ordered so that every cross-reference runs backwards. Appendix A fixes coordinates and records the partition of the design space that all later arguments use. Appendices B–D concern the baseline: the independent response function h, the single-audience benchmarks, and the citizens' and leader's problems on the spin family. Appendix E records the partisans' Inside structures for a general two-sided design. Appendix F isolates the four inequalities that carry the main proof, which Appendix G then assembles into the cell-by-cell domination argument behind Proposition 6. Appendix H reads that argument as a reduction to two candidate families and characterizes the optimum within each; Appendix I solves the partisan edge $p_I = 0$.

A Notation, coordinates, and standing conventions

Three coordinate systems appear below. Each is convenient for a different step, and the arguments move between them freely, so we record all three and the maps among them here, once.

Designs, posteriors, and Bayes plausibility. The leader chooses a spin κ and an underplay μ in the compact feasible set

$$\mathcal{D} \;:=\; \big\{(\kappa,\mu) \in [0,1]^2 \;:\; \kappa+\mu \le 1\big\},$$

so that $\Pr[s = r \mid \omega = R] = 1-\mu$ and $\Pr[s = r \mid \omega = L] = \kappa$ as in (22). The favorable message arrives with probability $\alpha := w(1-\mu) + (1-w)\kappa$ and induces interim posteriors

$$\pi_r = \frac{w(1-\mu)}{w(1-\mu)+(1-w)\kappa}, \qquad \pi_\ell = \frac{w\mu}{w\mu+(1-w)(1-\kappa)}, \tag{39}$$

which satisfy $0 \le \pi_\ell \le w \le \pi_r \le 1$. Bayes plausibility is the martingale identity

$$\alpha\,\pi_r + (1-\alpha)\,\pi_\ell \;=\; w. \tag{40}$$

On the interior $\{\kappa+\mu<1\}$ the map $(\kappa,\mu) \mapsto (\pi_\ell,\pi_r)$ is a bijection onto $\{0 \le \pi_\ell < w < \pi_r \le 1\}$, inverted by

$$\alpha = \frac{w-\pi_\ell}{\pi_r-\pi_\ell}, \qquad \mu = \frac{\pi_\ell(\pi_r-w)}{w(\pi_r-\pi_\ell)}, \qquad \kappa = \frac{(w-\pi_\ell)(1-\pi_r)}{(1-w)(\pi_r-\pi_\ell)}. \tag{41}$$

Every citizen's final posterior Π — after the public message and any Inside realization — is a martingale from the common prior, so $\mathbb{E}[\Pi] = w$, $\mathbb{E}[\Pi \mid r] = \pi_r$ and $\mathbb{E}[\Pi \mid \ell] = \pi_\ell$. This

is the only property of the Inside stage used in the domination arguments of Appendix G.

Symmetric coordinates. Several expressions that are rational and asymmetric in (π_ℓ, π_r) become polynomial and symmetric after centring at one-half. Write

$$t \equiv 1-2w, \qquad a \equiv 1-2\pi_\ell, \qquad b \equiv 2\pi_r - 1, \tag{42}$$

so that $w = \frac{1-t}{2}$, $\pi_\ell = \frac{1-a}{2}$, $\pi_r = \frac{1+b}{2}$. The restriction $w \le \frac{1}{2}$ gives $t \in [0,1)$; $\pi_\ell \le w \le \pi_r$ gives $a \ge t$ and $b \ge -t$; and the upper half $\pi_r > \frac{1}{2}$ is exactly $b > 0$. The favorable-message probability is $\alpha = (a-t)/(a+b)$, and the design is recovered by

$$\mu = \frac{(1-a)(b+t)}{(1-t)(a+b)}, \qquad \kappa = \frac{(1-b)(a-t)}{(1+t)(a+b)}. \tag{43}$$

Dictionary. The objects that recur below, in all three coordinate systems:

	(κ, μ)	(π_ℓ, π_r)	(a, b)
Full revelation	$(0,0)$	$(0,1)$	$(1,1)$
Uninformative	$\kappa + \mu = 1$	$\pi_\ell = \pi_r = w$	$(t, -t)$
Spin-only family $\mathcal{F}_1$	$\mu = 0$	$\pi_\ell = 0$	$a = 1$
Voting threshold	—	$\pi_r = \frac{1}{2}$	$b = 0$
Retention design	$\kappa = \kappa_{\text{ret}}$	$\pi_r = \frac{1}{2}$	$(1,0)$
Concavification tangency	$\kappa = \frac{w}{3(1-w)}$	$\pi_r = \frac{3}{4}$	$b = \frac{1}{2}$
Prior threshold $w = \frac{1}{4}$	—	—	$t = \frac{1}{2}$

Two entries of this table do the work in Lemma 26. Because $\pi_\ell \le w$ always, we have $a \ge t$; hence

$$w \le \tfrac{1}{4} \iff t \ge \tfrac{1}{2} \implies a \ge \tfrac{1}{2} \iff \pi_\ell \le \tfrac{1}{4}. \tag{44}$$

The prior threshold $w > \frac{1}{4}$ is therefore not a free parameter restriction: it is exactly the condition under which Bayes plausibility permits the unfavorable posterior to be parked above $\frac{1}{4}$, which is what the two-sided design needs in order to exceed the partisan budget $2w$.

Inside structures and the cell partition. By Section 5.1 each partisan plays one of N, GT, or GB, selected by the threshold indices

$$\Delta_P(\kappa, \mu) = \kappa + 3\mu - 2\kappa\mu - 1, \qquad \Delta_A(\kappa, \mu) = 3\kappa + \mu - 2\kappa\mu - 1, \tag{45}$$

matching (29): the pro-partisan plays GT when $\Delta_P \geq 0$ and GB when $\Delta_P < 0$; the anti-partisan plays GT when $\Delta_A \geq 0$ and GB when $\Delta_A < 0$. On the upper half $\{\pi_r > \frac{1}{2}\}$ this partitions the design space into four open cells

$$\mathcal{C}^{XY}, \qquad X, Y \in \{GT, GB\},$$

where X is the pro-partisan's structure and Y the anti-partisan's. *One of these cells is the object of the whole two-sided analysis*, and we give it a single name throughout:

$$\mathcal{F}_2 \;\equiv\; \mathcal{C}^{GT,GB} \;=\; \big\{(\kappa,\mu) \in \mathcal{D} :\; \pi_r > \tfrac{1}{2},\; \Delta_P > 0,\; \Delta_A < 0\big\}, \tag{46}$$

the partisan-targeting cell, on which the pro plays GT and the anti plays GB. Suprema over $\mathcal{F}_2$ are taken over its closure $\overline{\mathcal{F}_2}$. Alongside it sits the spin-only family

$$\mathcal{F}_1 \;\equiv\; \big\{(\kappa, 0) : \kappa \in [0,1]\big\}, \qquad \text{on which } \pi_\ell = 0. \tag{47}$$

In the coordinates (42) the two cell restrictions are polynomial: $\Delta_A < 0 \iff \Gamma_A > 0$ and $\Delta_P > 0 \iff \Gamma_P < 0$, where

$$\begin{aligned} \Gamma_A(a,b;t) &= (a+1-t)\,b^2 + (a+t)\,b + t - a(1-t), \\ \Gamma_P(a,b;t) &= a^2 b + a^2 t + a^2 + ab - at - bt - b - t. \end{aligned} \tag{48}$$

Indeed, substituting (43) into (45),

$$\Delta_A = \frac{2(a-t)}{(1-t^2)(a+b)^2}\,(-\Gamma_A), \qquad \Delta_P = \frac{2(b+t)}{(1-t^2)(a+b)^2}\,(-\Gamma_P), \tag{49}$$

and both prefactors are positive on the upper half. Hence the image of $\mathcal{F}_2$ in (a,b) coordinates is

$$\mathcal{F}_2^{ab} = \big\{(a,b) :\; t < a < 1,\; 0 < b < 1,\; \Gamma_A(a,b;t) > 0,\; \Gamma_P(a,b;t) < 0\big\}. \tag{50}$$

The boundary $\Delta_A = 0$, equivalently $\kappa = (1-\mu)/(3-2\mu)$, equivalently $\Gamma_A = 0$, is the *L-indifference locus*; the tie-breaking rule selects GB for the anti-partisan on it.

Continuity and existence. The support functions are given cell by cell, in posterior coordinates, by

$$S^P_{GT} = \frac{w + \pi_\ell - 2w\pi_\ell}{1 - \pi_\ell}, \qquad S^P_{GB} = 1 - \frac{(\pi_r + \pi_\ell - 2\pi_r\pi_\ell)(\pi_r - w)(1 - 2\pi_\ell)}{(\pi_r - \pi_\ell)^2}, \tag{51}$$

$$S^A_{GT} = w\Big(2 - \tfrac{1}{\pi_r}\Big), \qquad S^A_{GB} = \frac{(\pi_r + \pi_\ell - 2\pi_r\pi_\ell)(w - \pi_\ell)(2\pi_r - 1)}{(\pi_r - \pi_\ell)^2}. \tag{52}$$

Lemma 11 (Continuity and existence) *V is continuous on $\mathcal{D}$, and the leader's problem attains a maximum. Moreover, on the upper half $\{\pi_r > \frac{1}{2}\}$,*

$$V \;=\; p_I S^I + p_P \max\{S^P_{GT}, S^P_{GB}\} + p_A \min\{S^A_{GT}, S^A_{GB}\}, \tag{53}$$

and V is smooth on the interior of each cell.

Proof. Expanding (51)–(52) in (κ, μ) and collecting terms gives the two cell-boundary identities

$$S^P_{GT} - S^P_{GB} = \frac{\kappa\,\Delta_P\,[(1-w)(1-\kappa) - w\mu]}{(1-\kappa)(1-\kappa-\mu)}, \qquad S^A_{GB} - S^A_{GT} = \frac{\mu\,\Delta_A\,[w(1-\mu) - (1-w)\kappa]}{(1-\mu)(1-\kappa-\mu)}, \tag{54}$$

each vanishing on its own cell boundary. Since the bracketed factors are nonnegative on the upper half, the sign of the first is the sign of Δ_P and the sign of the second is the sign of Δ_A; the partisan best responses therefore select the maximum and the minimum in (53), and the smooth cell formulas glue continuously across interfaces. Continuity across $\pi_r = \frac{1}{2}$ into the lower half follows from Step 2 of Appendix G. A maximum of a continuous function on the compact set $\mathcal{D}$ is attained, so throughout the appendices we write max rather than sup. ■

Polynomial identities. Several displays below are polynomial identities in (κ, μ, w) or (a, b, t) that are proved by expanding both sides and collecting terms; where the expansion is long we say so and give the intermediate form rather than writing "direct calculation." All such identities have been verified symbolically, and the verification script is included with the replication package.

B The independent response function

An independent starts from the public posterior π, then draws her own private signal: an accuracy q pulled uniformly from $[\frac{1}{2}, 1]$ and a reading at that accuracy. She backs the leader when her final belief weakly favors R. This appendix computes how often that happens. To avoid a clash with the leader's normalized objective $g(\cdot\,; p)$ of (18), we write ρ_R, ρ_L for the state-conditional support probabilities.

Lemma 12 (Independent response) *For an independent with interim public posterior π, the state-conditional probabilities of supporting the leader are*

$$\rho_R(\pi) \;=\; 2\pi - \pi^2, \qquad \rho_L(\pi) \;=\; \pi^2, \tag{55}$$

and her ex-ante support probability is

$$h(\pi) \;=\; \pi\,\rho_R(\pi) + (1-\pi)\,\rho_L(\pi) \;=\; 3\pi^2 - 2\pi^3. \tag{56}$$

Proof. A private reading r at accuracy q leaves posterior $\pi q/[\pi q + (1-\pi)(1-q)]$, which clears $\frac{1}{2}$ exactly when $q \geq 1-\pi$; a reading ℓ clears it when $q \leq \pi$. Averaging over q with density 2 on $[\frac{1}{2}, 1]$,

$$\rho_R(\pi) \;=\; 2\int_{1/2}^{1} \big[\, q\,\mathbf{1}\{q \geq 1-\pi\} + (1-q)\,\mathbf{1}\{q \leq \pi\}\,\big]\, dq \;=\; 2\pi - \pi^2,$$

since for $\pi \geq \frac{1}{2}$ the first term runs over all of $[\frac{1}{2}, 1]$ and the second over $[\frac{1}{2}, \pi]$, while for $\pi \leq \frac{1}{2}$ only the first survives, on $[1-\pi, 1]$. The symmetric computation gives $\rho_L(\pi) = \pi^2$, and h combines the two states. ∎

The shape of h is what the leader trades on. It rises throughout, is convex below $\frac{1}{2}$ and concave above ($h''(\pi) = 6 - 12\pi$), and is tangent to the ray from the origin at $\pi = \frac{3}{4}$, where $h(\pi)/\pi = h'(\pi)$ reads $3\pi - 2\pi^2 = 6\pi - 6\pi^2$. The tangent line is $y = \frac{9}{8}\pi$, and

$$h(\pi) - \tfrac{9}{8}\pi \;=\; -2\pi\big(\pi - \tfrac{3}{4}\big)^2, \tag{57}$$

a simple zero at 0 and a double zero at $\frac{3}{4}$. Hence the contact set of the concave closure is exactly $\{0, \frac{3}{4}\}$ and

$$(\operatorname{cav} h)(\pi) = \begin{cases} \frac{9}{8}\,\pi, & \pi \in [0, \frac{3}{4}], \\ h(\pi), & \pi \in [\frac{3}{4}, 1]. \end{cases} \tag{58}$$

That single bend is the whole concavification motive: splitting the prior $w \leq \frac{1}{2} < \frac{3}{4}$ into the posteriors $\{0, \frac{3}{4}\}$ wrings out the most independent support Bayes plausibility allows.

Remark 13 (What the uniform assumption is doing) *Only four properties of h are used anywhere below: $h(0) = 0$; h is increasing; h has a single change of curvature, convex then concave; and the resulting tangency point lies strictly inside $(w, 1)$. Any distribution of private precisions delivering these gives the same three-regime structure with $\frac{3}{4}$ replaced by the corresponding tangency; the uniform is chosen because it makes that point, and hence every threshold in the paper, a rational number.*

C Single-audience benchmarks

Proposition 2 states the benchmarks on the spin family. We prove the stronger statement directly — the same values and optimizers when the leader may also underplay — and read Proposition 2 off as the $\mu = 0$ corollary. This also establishes that the enrichment of Section 5 can only matter for *mixed* electorates.

Lemma 14 (Single-audience benchmarks, two-sided designs) *Fix $w \in (0, \frac{1}{2}]$ and let the leader choose any $(\kappa, \mu) \in \mathcal{D}$. Facing a single audience, her optimum is:*

(i) Independents only *($p_I = 1$): $(\kappa^*, \mu^*) = \left(\frac{w}{3(1-w)}, 0\right)$, with $\pi_r^* = \frac{3}{4}$, $\pi_\ell^* = 0$, and $V^* = \frac{9}{8}w$; the optimum is unique, so underplay is strictly suboptimal.*

(ii) Pro-partisans only *($p_P = 1$): every design with $\pi_r(\kappa, \mu) \le \frac{1}{2}$ is optimal, with final posteriors $\{0, \frac{1}{2}\}$ and $V^* = 2w$; in particular every uninformative design is optimal, and the optimum is not unique.*

(iii) Anti-partisans only *($p_A = 1$): full revelation $(\kappa^*, \mu^*) = (0, 0)$ is the unique optimum, with $V^* = w$.*

For every $w \in (0, \frac{1}{2}]$ the benchmark values are ordered $V_P^* = 2w > V_I^* = \frac{9}{8}w > V_A^* = w$.

Proof. Each part switches off two audiences so that a single force operates in isolation, and in every case the argument has the same shape: exhibit a design, then show that a Bayes-plausibility cap makes it unbeatable. Throughout, index a design by the interim posteriors $\pi_\ell \le w \le \pi_r$ it induces, with α and the martingale identity (40) as in Appendix A. Because r is good news, π_r is the only interim posterior that can exceed $\frac{1}{2}$.

(i) Independents only. An independent at interim posterior π supports the leader with ex-ante probability $h(\pi)$ (Lemma 12), the private precision already integrated out. The leader's payoff is $\mathbb{E}[h(\Pi_0)] = \alpha\, h(\pi_r) + (1-\alpha)\, h(\pi_\ell)$ over the interim public posterior Π_0, and she chooses its Bayes-plausible distribution (mean w) to maximize it; by the concavification theorem [Kamenica and Gentzkow, 2011] the value is $(\operatorname{cav} h)(w)$. By (58) and $w \le \frac{1}{2} < \frac{3}{4}$, this equals $\frac{9}{8}w$. By (57) the maximizing distribution is supported on $\{0, \frac{3}{4}\}$, with weights $\left(1 - \frac{4w}{3}, \frac{4w}{3}\right)$, feasible since $\frac{4w}{3} \le \frac{2}{3}$. Inverting (41), $\pi_\ell^* = 0 \Rightarrow \mu^* = 0$ and $\pi_r^* = \frac{3}{4} \Rightarrow \kappa^* = \frac{w}{3(1-w)}$. The double zero at $\frac{3}{4}$ makes the contact set exactly $\{0, \frac{3}{4}\}$, so the optimal posteriors, and hence (κ^*, μ^*), are unique.

(ii) Pro-partisans only. Here $V = S^P = \Pr(\Pi \ge \frac{1}{2})$ after the pro-partisan best-responds. Markov's inequality applied to $\Pi \ge 0$ with $\mathbb{E}[\Pi] = w$ caps *every* design:

$$S^P = \Pr\left(\Pi \ge \tfrac{1}{2}\right) \;\le\; \frac{\mathbb{E}[\Pi]}{1/2} \;=\; 2w.$$

The bound is attained by any uninformative design: with $\pi_r = \pi_\ell = w$ the pro-partisan plays the Inside structure with $\Pr(G \mid R) = 1$ and $\Pr(G \mid L) = \frac{w}{1-w}$, so G fires with probability $2w$ and induces posterior $\frac{1}{2}$ while its complement induces 0. To pin down the optimal set, suppose first $\pi_r > \frac{1}{2}$. Bounding the favorable realization by its mass α, the unfavorable one by Markov ($\Pr(\Pi \geq \frac{1}{2} \mid \ell) \leq 2\pi_\ell$), and using $(1-\alpha)\pi_\ell = w - \alpha\pi_r$,

$$S^P \;\leq\; \alpha + 2(1-\alpha)\pi_\ell \;=\; 2w - \alpha\,(2\pi_r - 1) \;<\; 2w,$$

so every design with $\pi_r > \frac{1}{2}$ is strictly suboptimal. Conversely, if $\pi_r \leq \frac{1}{2}$ then every interim posterior is at most $\frac{1}{2}$, and applying the $\{0, \frac{1}{2}\}$-split conditional on each message yields $S^P = \mathbb{E}[2\Pi_0] = 2w$. The optimal set is therefore exactly $\{(\kappa, \mu) : \pi_r(\kappa, \mu) \leq \frac{1}{2}\}$.

(iii) Anti-partisans only. Here $V = S^A = \Pr(\Pi > \frac{1}{2})$, and the anti-partisan chooses Inside media to *minimize* this probability, her tie at $\frac{1}{2}$ breaking toward L. After ℓ the interim posterior is at most $\frac{1}{2}$ and she retains it, contributing nothing. After r with $\pi_r > \frac{1}{2}$, the support-minimizing mean-π_r spread places mass on $\{\frac{1}{2}, 1\}$ with weights $(2 - 2\pi_r,\, 2\pi_r - 1)$, so $\Pr(\Pi > \frac{1}{2} \mid r) = 2\pi_r - 1$: this is the GT structure. Hence $S^A = \alpha(2\pi_r - 1)$ if $\pi_r > \frac{1}{2}$ and $S^A = 0$ otherwise. For $\pi_r > \frac{1}{2}$, the martingale with $\pi_\ell \geq 0$ gives $\alpha\pi_r \leq w$, so

$$S^A = \alpha(2\pi_r - 1) \;\leq\; \frac{w}{\pi_r}(2\pi_r - 1) = w\Big(2 - \tfrac{1}{\pi_r}\Big) \;\leq\; w,$$

the first inequality binding iff $\pi_\ell = 0$ (that is, $\mu = 0$) and the second iff $\pi_r = 1$. Both bind only at $(\kappa^*, \mu^*) = (0, 0)$: under full revelation the state-R interim posterior is degenerate and cannot be moved, so the anti-partisan supports in state R and not in state L, giving $S^A = \Pr(\omega = R) = w$. Partial spin strictly backfires.

Ordering. $2w > \frac{9}{8}w > w$ for every $w \in (0, \frac{1}{2}]$. ■

Corollary 15 (Proposition 2) *Restricting to $\mathcal{F}_1$ leaves the statement of Lemma 14 unchanged: the independent optimum $\kappa^* = \frac{w}{3(1-w)}$ and the full-revelation optimum $\kappa^* = 0$ both lie on $\mathcal{F}_1$, and the pro-partisan optimal set intersects $\mathcal{F}_1$ in $\{\kappa \geq \kappa_{\text{ret}}\}$. This is Proposition 2.*

The three optima are the population-simplex vertices of Proposition 3: pure-pro lies in retention, pure-independent in independent targeting (at $\pi_r = \frac{3}{4}$, since $p_P = p_A = 0$), and pure-anti in full revelation.

D The baseline: best responses, optimal spin, and invariance

Throughout this appendix the design is a spin level $\kappa \in [0, 1]$, equivalently the target $\pi_r = w/[w + (1-w)\kappa]$ of (6). Bad news is conclusive, so $\pi_\ell = 0$, the favorable message

arrives with probability $\alpha = w/\pi_r$, and Bayes plausibility (40) collapses to $\alpha\pi_r = w$.

Proof of Lemma 1. After ℓ the posterior is 0 whatever the Inside realization, since ℓ never occurs in state R: the unfavorable branch adds nothing and ties nothing down, so the problem is the interim one on the favorable branch. There the heart picks a Bayes-plausible spread of posteriors with mean π_r to maximize the chance of an identity-consistent belief, and the value is the concavification of the relevant indicator [Kamenica and Gentzkow, 2011].

For the pro-partisan the indicator is $\mathbf{1}\{\pi \geq \frac{1}{2}\}$, concavified to $\min\{1, 2\pi\}$. When $\pi_r < \frac{1}{2}$ she puts mass $2\pi_r$ on $\frac{1}{2}$ — where her tie-break keeps her identity — and the rest on 0; when $\pi_r \geq \frac{1}{2}$ she needs nothing further. Times $\alpha = w/\pi_r$, this is $S^P = w\min\{2, 1/\pi_r\}$: the cap $2w$ for $\pi_r \leq \frac{1}{2}$, and w/π_r above.

For the anti-partisan the indicator is $\mathbf{1}\{\pi \leq \frac{1}{2}\}$, concavified to 1 on $[0, \frac{1}{2}]$ and $2(1-\pi)$ on $[\frac{1}{2}, 1]$, achieved by the split $\{\frac{1}{2}, 1\}$. She backs the leader only at posterior 1, which arrives on the favorable branch with probability $2\pi_r - 1$ when $\pi_r > \frac{1}{2}$ and never otherwise — the GT structure. So $S^A = w\max\{0, 2 - 1/\pi_r\}$, and adding the two on $[\frac{1}{2}, 1]$ gives $S^P + S^A = 2w$.

For $\pi_r \geq \frac{1}{2}$ the pro-partisan's optimum is not unique as a signal structure, only as a distribution of final posteriors: the uninformative structure N and the degenerate GB structure of Appendix E both deliver $\{0, \pi_r\}$ and hence $S^P = w/\pi_r$. We report it as N, the simplest representative; Appendix E records the degenerate GB form, which is what the index Δ_P of (45) selects on this slice. ■

Proof of Proposition 3. By (18) the leader maximizes $g(\cdot\,; p)$ on $[w, 1]$. On $[w, \frac{1}{2})$, $g(\pi) = p_I\,\pi(3 - 2\pi) + 2p_P$ is nondecreasing (strictly if $p_I > 0$, since $3 - 4\pi > 0$), so the maximum over $[w, \frac{1}{2}]$ sits at $\frac{1}{2}$ and we work on $[\frac{1}{2}, 1]$. There g is smooth with derivative (19); multiplying by $\pi^2 > 0$,

$$\pi^2 g'(\pi) \;=\; p_I\,\pi^2(3 - 4\pi) - (p_P - p_A),$$

and $\pi^2(3 - 4\pi)$ has derivative $6\pi(1 - 2\pi) < 0$ on $(\frac{1}{2}, 1)$, so g' crosses zero at most once, positive to negative: g is single-peaked with a unique maximizer. The three regimes come from the endpoints of $\pi^2(3 - 4\pi)$, which is $\frac{1}{4}$ at $\pi = \frac{1}{2}$ and -1 at $\pi = 1$. So $g'(\frac{1}{2}) \leq 0$ iff $p_P - p_A \geq p_I/4$ (retention, $\pi_r^* = \frac{1}{2}$); $g'(1) \geq 0$ iff $p_P - p_A \leq -p_I$ (full revelation, $\pi_r^* = 1$); and between them the unique interior root of

$$p_I\,\pi_r^2(3 - 4\pi_r) \;=\; p_P - p_A \tag{59}$$

(independent targeting), which falls as $p_P - p_A$ rises and hits $\frac{3}{4}$ when $p_P = p_A$. When

$p_I = 0$ the derivative is $(p_A - p_P)/\pi^2$, of one sign, and the same split holds with ties going to the corner.

The values are substitution. At $\pi = \frac{1}{2}$, $g = p_I + 2p_P$, so $V^* = w(p_I + 2p_P)$ and $(S^I, S^P, S^A) = (w, 2w, 0)$ by Lemma 1. At $\pi = 1$, $g = 1$, so $V^* = w$ and every group supports with probability w. At an interior optimum, substituting $p_P - p_A = p_I\,\pi^2(3 - 4\pi)$ gives $g = 6p_I\,\pi(1-\pi) + 2p_A$, so $V^* = 2w[\,3p_I\,\pi_r^*(1-\pi_r^*) + p_A\,]$.

The simplex forms use $p_I = 1 - p_P - p_A$: $p_P - p_A \geq \frac{1-p_P-p_A}{4} \iff 5p_P - 3p_A \geq 1$, and $p_P - p_A \leq -(1 - p_P - p_A) \iff p_A \geq \frac{1}{2}$. ■

Proof of Proposition 5. (i) By (18), $V(\pi; w, p) = w\,g(\pi; p)$ with $w > 0$, so the maximizer over $[\frac{1}{2}, 1]$ depends only on p; restricting to $[\frac{1}{2}, 1]$ is legitimate for every $w \leq \frac{1}{2}$ since the menu $[w, 1]$ always contains it. The boundaries $p_P - p_A \geq p_I/4$ and $p_P - p_A \leq -p_I$ and the interior condition (59) involve only p; for $p_I > 0$, dividing by p_I leaves them depending on composition only through $(p_P - p_A)/p_I$.

(ii) By (9), (11), and (10), each support is w times a function of π_r^* alone, and $V^* = w\,g(\pi_r^*; p)$; part (i) makes π_r^* free of w, so all four scale with w. The unconditional spin probability is $(1-w)\kappa^* = w(1-\pi_r^*)/\pi_r^*$ by (6), again w times a w-free quantity.

(iii) Immediate from (ii): any ratio of two quantities proportional to w is w-free. The conditional spin rate is $\kappa^* = \frac{w}{1-w} \cdot \frac{1-\pi_r^*}{\pi_r^*}$, and $w \mapsto w/(1-w)$ rises on $(0, \frac{1}{2}]$, so spin rises with the prior even though the target does not move. ■

E Partisan Inside media under two-sided designs

This appendix works out each partisan's Inside media for a general two-sided design (κ, μ), the input to Section 5. Given the state, the Inside signal is drawn independently of the public message, so a structure is just its two state-conditional laws.

Three facts, and two abbreviations. A terrible realization T settles the partisan's unfavorable state outright, so it never fires in her favorable one: $\Pr(T \mid R) = 0$ for the pro-partisan and $\Pr(T \mid L) = 0$ for the anti. A good realization G and a bad realization B each land her final posterior at exactly $\frac{1}{2}$ — G paired with her unfavorable message, B with her favorable one. The uninformative structure N fires G for sure. Every expression below is built from three quantities:

$$B := (1-w)(1-\kappa) - w\mu, \qquad B' := w(1-\mu) - (1-w)\kappa, \qquad C := 1 - \kappa - \mu. \tag{60}$$

These are exactly the feasibility margins:

$$B \geq 0 \iff \pi_\ell \leq \tfrac{1}{2}, \qquad B' \geq 0 \iff \pi_r \geq \tfrac{1}{2}, \qquad B - B' = 1 - 2w. \tag{61}$$

So the GT structure is available to the pro-partisan exactly when $\pi_\ell \leq \frac{1}{2}$ and to the anti-partisan exactly when $\pi_r \geq \frac{1}{2}$, and GB is available to either exactly when $\pi_\ell \leq \frac{1}{2} \leq \pi_r$.

Derivation. Take the pro-partisan; the anti-partisan is the mirror image under $(\kappa \leftrightarrow \mu,\ R \leftrightarrow L)$. Under GT, $\Pr(T \mid R) = 0$ forces $\Pr(G \mid R) = 1$, and holding the belief after (ℓ, G) at $\frac{1}{2}$ means $w\mu = (1-w)(1-\kappa)\Pr(G \mid L)$, so $\Pr(G \mid L) = w\mu/[(1-w)(1-\kappa)]$. Under GB, the two pins — the beliefs after (ℓ, G) and after (r, B) both at $\frac{1}{2}$ — are two equations in $\Pr(G \mid R)$ and $\Pr(G \mid L)$, whose solution is the GB block of Table 2.

Table 2: Inside signal laws for a general design (κ, μ), in the notation (60). The two types are mirror images under $(\kappa \leftrightarrow \mu,\ R \leftrightarrow L)$.

		$\omega = R$	$\omega = L$
Pro-partisan	GT: $\Pr(G \mid \cdot)$	1	$\dfrac{w\mu}{(1-w)(1-\kappa)}$
	$\Pr(T \mid \cdot)$	0	$\dfrac{B}{(1-w)(1-\kappa)}$
	GB: $\Pr(G \mid \cdot)$	$\dfrac{(1-\kappa)B'}{wC}$	$\dfrac{\mu B'}{(1-w)C}$
	$\Pr(B \mid \cdot)$	$\dfrac{\kappa B}{wC}$	$\dfrac{(1-\mu)B}{(1-w)C}$
Anti-partisan	GT: $\Pr(G \mid \cdot)$	$\dfrac{(1-w)\kappa}{w(1-\mu)}$	1
	$\Pr(T \mid \cdot)$	$\dfrac{B'}{w(1-\mu)}$	0
	GB: $\Pr(G \mid \cdot)$	$\dfrac{\kappa B}{wC}$	$\dfrac{(1-\mu)B}{(1-w)C}$
	$\Pr(B \mid \cdot)$	$\dfrac{(1-\kappa)B'}{wC}$	$\dfrac{\mu B'}{(1-w)C}$

Support. The pro-partisan backs the leader when her final posterior is at least $\frac{1}{2}$. Under GT this holds after G but not after T, so she supports iff $\sigma = G$. Under GB it fails only after (ℓ, B), so she supports unless that pair lands. The anti-partisan's numbers are the mirror; the one asymmetry is her tie-break, which goes against the leader, so she needs a strict majority to support. This gives Table 3.

Table 3: Probability of supporting the leader, by type, structure, and state.

		$\omega = R$	$\omega = L$
Pro-partisan	GT	1	$\dfrac{w\mu}{(1-w)(1-\kappa)}$
	GB	$1 - \dfrac{\mu\kappa B}{wC} = (1-\mu) + \dfrac{(1-\kappa)\mu B'}{wC}$	$\kappa + \dfrac{(1-\kappa)\mu B'}{(1-w)C}$
Anti-partisan	GT	$\dfrac{B'}{w(1-\mu)}$	0
	GB	$\dfrac{(1-\mu)(1-\kappa)B'}{wC}$	$\dfrac{\kappa\mu B'}{(1-w)C}$

Consistency. The laws sum to one in each state through the two identities

$$(1-\kappa)B' + \kappa B = wC, \qquad \mu B' + (1-\mu)B = (1-w)C, \tag{62}$$

each obtained by expanding the left side and collecting the terms in $\kappa\mu$. The two equalities in the pro-GB entry of Table 3 are the first of these, and the alternative form of the pro-GB state-L entry, $1 - (1-\kappa)(1-\mu)B/[(1-w)C]$, is the second. The ex-ante supports are $S^j = w\Pr(\text{supports} \mid R) + (1-w)\Pr(\text{supports} \mid L)$, and substituting Table 3 into that formula and rewriting in posterior coordinates returns (51)–(52).

Boundary check 1: the spin-only slice $\mu = 0$. Here $C = 1-\kappa$, $B = (1-w)(1-\kappa)$ and $B' = w - (1-w)\kappa$, and the table degenerates as follows.

- *Pro-partisan.* GT becomes fully revealing on the unfavorable side ($\Pr(G \mid L) = 0$), worth only $S^P = w$. GB becomes $\Pr(G \mid R) = B'/w$, $\Pr(B \mid R) = \kappa(1-w)/w$, $\Pr(G \mid L) = 0$, $\Pr(B \mid L) = 1$, feasible exactly when $B' \geq 0$, i.e. $\kappa \leq \kappa_{\text{ret}}$, i.e. $\pi_r \geq \frac{1}{2}$. Its support probabilities are 1 in state R and κ in state L, so

$$S^P = w + (1-w)\kappa = \frac{w}{\pi_r}.$$

  Since $\Delta_P(\kappa, 0) = \kappa - 1 < 0$ for $\kappa < 1$, the index of (45) selects GB, not GT — and the value it delivers is exactly that of the uninformative structure N, because with $\pi_\ell = 0$ the realization following ℓ is payoff-irrelevant. This is the sense in which Lemma 1 reports N: on this slice N and GB induce the same distribution of final posteriors, $\{0, \pi_r\}$.

- *Anti-partisan.* $\Pr(G \mid R) = (1-w)\kappa/w$ and $\Pr(G \mid L) = 1$ under *both* GT and GB:

the two structures coincide, with support B'/w in state R and 0 in state L, so

$$S^A \;=\; w-(1-w)\kappa \;=\; w\Big(2-\frac{1}{\pi_r}\Big).$$

Both expressions are those of Lemma 1, and they sum to $2w$.

Boundary check 2: full revelation $\kappa=\mu=0$. Now $C=1$, $B=1-w$, $B'=w$. Every entry collapses to a perfectly informative Inside signal ($\Pr(G\mid R)=1$, $\Pr(B\mid L)=1$ for the pro; the mirror for the anti), support probabilities $(1,0)$ for both types, and $S^P=S^A=w$. Together with $S^I=h(1)\cdot w=w$ this is $(S^I,S^P,S^A)=(w,w,w)$, as in Proposition 3(iii).

F Auxiliary inequalities

Four inequalities carry the domination argument of Appendix G. We state and prove them here, out of the proof, each with the economics it encodes. Throughout, work in posterior coordinates with $0\le\pi_\ell\le w\le\pi_r\le 1$ and $w\le\frac{1}{2}$.

Lemma 16 (Slice convexity on $\mathcal{C}^{GT,GT}$) *On $\mathcal{C}^{GT,GT}$, $V^{GT,GT}:=p_IS^I+p_PS^P_{GT}+p_AS^A_{GT}$ is convex in π_ℓ along each π_r-slice, strictly unless $p_I=p_P=0$. Explicitly,*

$$\frac{\partial^2 V^{GT,GT}}{\partial\pi_\ell^2}=4p_I(\pi_r-w)+\frac{2p_P(1-w)}{(1-\pi_\ell)^3}\;\ge\;0.$$

What it says. When both partisans slant hard, the leader never wants an interior amount of ambiguity in her bad news: she is pushed to one extreme or the other, so this cell can only contribute through its boundary.

Proof. S^A_{GT} does not depend on π_ℓ, and differentiating $S^I=\alpha h(\pi_r)+(1-\alpha)h(\pi_\ell)$ with $\alpha=(w-\pi_\ell)/(\pi_r-\pi_\ell)$ and $S^P_{GT}=(w+\pi_\ell-2w\pi_\ell)/(1-\pi_\ell)$ twice in π_ℓ gives the display. Both terms are nonnegative because $\pi_r\ge w$ and $\pi_\ell<1$. ∎

Lemma 17 (Two-target trade-off) *On $\{\kappa+\mu<1\}$ with $w\le\frac{1}{2}$, writing $R:=(1-\kappa)(1-\mu)+\kappa\mu$ and $C:=1-\kappa-\mu$,*

$$S^P_{GB}+S^A_{GB}\;=\;1-\frac{(1-2w)R}{C}\;\le\;2w. \tag{63}$$

What it says. The zero-sum partisan budget of Remark 4 does not go away when both partisans hedge: the leader still cannot hold both camps at once past $2w$. What two-sided

design buys is not a bigger partisan pie but a different split of it, paid for out of the independents.

Proof. With B, B' as in (60), expanding (51) and (52) in (κ, μ) gives $S^P_{GB} = 1 - RB/C$ and $S^A_{GB} = RB'/C$, so $S^P_{GB} + S^A_{GB} = 1 - R(B - B')/C$. Since $B - B' = 1 - 2w$ by (61), the identity follows. For the bound, $R - C = 2\kappa\mu \geq 0$, so $R \geq C > 0$; combined with $1 - 2w \geq 0$ this gives $(1 - 2w)R/C \geq 1 - 2w$ and hence the right-hand inequality. ■

Lemma 18 (Independent–anti envelope) *On the upper half $\{\pi_\ell < \frac{1}{2} < \pi_r\}$ with $w \leq \frac{1}{2}$ and $\Delta_A < 0$, writing $A := S^A_{GB}$,*

$$S^I - w \;\leq\; \Phi_w(A), \qquad \Phi_w(A) \;:=\; \frac{A\, w\, (w - A)}{(2w - A)^2}. \tag{64}$$

What it says. The concavification gain the leader can still extract from independents is bounded by a function of how much anti-partisan support she is already extracting. The two prizes are not independent: pushing the favorable message harder to win independents is exactly what lets the anti-partisan's counter-design pay off, so the envelope prices one in terms of the other.

Proof. By Bayes plausibility, $S^I - w = \alpha\pi_r(1 - \pi_r)(2\pi_r - 1) - (1 - \alpha)\pi_\ell(1 - \pi_\ell)(1 - 2\pi_\ell)$. On the upper half the second term enters with a nonpositive sign, so

$$S^I - w \;\leq\; \alpha\, \pi_r(1 - \pi_r)(2\pi_r - 1). \tag{65}$$

Reparametrize by $a = 1 - 2\pi_\ell$ and $b = 2\pi_r - 1$ as in (42), with $a, b \in [0, 1]$; substitution gives $\pi_r - \pi_\ell = (a + b)/2$, $\pi_r + \pi_\ell - 2\pi_r\pi_\ell = (1 + ab)/2$ and $\pi_r(1 - \pi_r) = (1 - b^2)/4$. Combining with the closed form for S^A_{GB} in (52),

$$\frac{\alpha\, \pi_r(1 - \pi_r)(2\pi_r - 1)}{A} \;=\; \frac{(1 - b^2)(a + b)}{4(1 + ab)} \;\leq\; \frac{1 - b^2}{4}, \tag{66}$$

the inequality because $1 + ab - (a + b) = (1 - a)(1 - b) \geq 0$. Next, $\Delta_A < 0$ implies $A \leq S^A_{GT} = w(2 - 1/\pi_r) = 2wb/(1 + b)$, so $A/w \leq 2b/(1 + b)$. The function $f(x) := (1 - x)/(2 - x)^2$ has $f'(x) = -x/(2 - x)^3 < 0$ on $(0, 1)$, so

$$\frac{w(w - A)}{(2w - A)^2} = f(A/w) \;\geq\; f\left(\frac{2b}{1 + b}\right) = \frac{1 - b^2}{4}. \tag{67}$$

Chaining (65), (66) and (67) gives $S^I - w \leq A \cdot \frac{1 - b^2}{4} \leq \Phi_w(A)$. ■

Lemma 19 (Anti-matched dominance) *Let (π_ℓ, π_r) lie in the upper half with $\Delta_A < 0$, set*

$A := S^A_{GB} \in [0, w]$, *and let*

$$\hat{\pi} \;:=\; \frac{w}{2w - A} \;\in\; [\tfrac{1}{2}, 1]$$

be the spin-only design that matches it on the anti-partisan. Then

$$S^A_{\mathcal{F}_1}(\hat{\pi}) = A, \qquad S^P_{\mathcal{F}_1}(\hat{\pi}) = 2w - A, \qquad S^I_{\mathcal{F}_1}(\hat{\pi}) - w = \Phi_w(A). \tag{68}$$

Consequently, if in addition $S^P + A \le 2w$ and $S^I - w \le \Phi_w(A)$ at (π_ℓ, π_r), then $V(\pi_\ell, \pi_r) \le V_{\mathcal{F}_1}(\hat{\pi}) \le V_1^*$.

What it says. This is the bridge. Any two-sided design is compared not to an arbitrary spin-only design but to the one that delivers the anti-partisan exactly the same support; on that design the other two audiences are then automatically at least as well served. The three inequalities of Lemmas 17–18 are built to hit this target simultaneously.

Proof. That $A \in [0, w]$: positivity from $\pi_r > \frac{1}{2}$ and $\pi_\ell < w$, and $A \le S^A_{GT} \le w$ from $\Delta_A < 0$ together with $\pi_r \le 1$. Hence $\hat{\pi} \in [\frac{1}{2}, 1]$. From $\hat{\pi} = w/(2w - A)$ we get $1/\hat{\pi} = (2w - A)/w$ and $2\hat{\pi} - 1 = A/(2w - A)$, so

$$S^A_{\mathcal{F}_1}(\hat{\pi}) = w\Big(2 - \tfrac{1}{\hat{\pi}}\Big) = A, \qquad S^P_{\mathcal{F}_1}(\hat{\pi}) = \frac{w}{\hat{\pi}} = 2w - A,$$

$$S^I_{\mathcal{F}_1}(\hat{\pi}) - w = w\hat{\pi}(3 - 2\hat{\pi}) - w = w(2\hat{\pi} - 1)(1 - \hat{\pi}) = w \cdot \frac{A}{2w - A} \cdot \frac{w - A}{2w - A} = \Phi_w(A).$$

The consequence is immediate: $p_I S^I + p_P S^P + p_A A \le p_I[w + \Phi_w(A)] + p_P(2w - A) + p_A A = V_{\mathcal{F}_1}(\hat{\pi})$. ∎

G Proof of Proposition 6

Fix $w \in (0, \frac{1}{2}]$ and $(p_I, p_P, p_A) \in \Delta^2$. Using the notation of Appendix A, put

$$V_1^* := \max_{(\kappa,\mu)\in\mathcal{F}_1} V(w; \kappa, \mu), \qquad V_2^* := \max_{(\kappa,\mu)\in\overline{\mathcal{F}_2}} V^{GT,GB}(w; \kappa, \mu),$$

where $V^{GT,GB} := p_I S^I + p_P S^P_{GT} + p_A S^A_{GB}$ is the smooth function with which V coincides on $\overline{\mathcal{F}_2}$. Both maxima are attained by Lemma 11. In this notation Proposition 6 states that

$$\max_{(\kappa,\mu)\in\mathcal{D}} V(w; \kappa, \mu) \;=\; \max\{V_1^*,\, V_2^*\}, \tag{69}$$

with V_1^* attained on $\mathcal{F}_1$ at the kink $\pi_r = \frac{1}{2}$, at an interior $\pi_r^* \in (\frac{1}{2}, 1)$, or at full revelation $\pi_r = 1$ according to the partisan-imbalance trichotomy, and V_2^* attained on the compact

set $\overline{\mathcal{F}_2}$ at an interior critical point of $V^{GT,GB}$ or at a boundary point satisfying the corresponding KKT condition; shared-boundary values are identified by continuity. We prove this restatement. The whole content is that *every* design outside $\mathcal{F}_1 \cup \overline{\mathcal{F}_2}$ is dominated by a design in $\mathcal{F}_1$; Table 4 summarizes which step handles which region.

Proof. The argument proceeds in six steps, in posterior coordinates (π_ℓ, π_r) with $0 \le \pi_\ell \le w \le \pi_r \le 1$, related to (κ, μ) by (41).

Step 1 (Cells and continuity). By Lemma 11, V is continuous on $\mathcal{D}$, is given on the upper half by (53), and is smooth on the interior of each cell $\mathcal{C}^{XY}$; the cell formulas glue through the identities (54). Continuity across $\pi_r = \frac{1}{2}$ into the lower half follows from Step 2. Since $\mathcal{D}$ is compact, a maximizer exists.

Step 2 (Lower-half reduction). Suppose $\pi_r \le \frac{1}{2}$. The anti-partisan, by playing the constant action $a^A \equiv 0$, secures $S^A_* = 0$. For independents, $h(\pi) \le \pi$ on $[0, \frac{1}{2}]$, so by Bayes plausibility

$$S^I = \alpha h(\pi_r) + (1-\alpha) h(\pi_\ell) \;\le\; \alpha \pi_r + (1-\alpha)\pi_\ell = w.$$

For the pro-partisan, her final posterior is a martingale from w, so Markov gives $\Pr[\Pi \ge \frac{1}{2}] \le 2w$ and $S^P_* \le 2w$. Hence $V \le p_I w + 2w\, p_P$ on the lower half, which is the value of $V_{\mathcal{F}_1}$ at the kink $\pi_r = \frac{1}{2}$. The lower half is therefore dominated by $\mathcal{F}_1$, and we restrict henceforth to $\pi_r > \frac{1}{2}$. In particular the pairing BT, which can be optimal for the pro-partisan only when $\pi_r < \frac{1}{2}$, never arises at an optimum.

Step 3 (Trichotomy on $\mathcal{F}_1$). On $\mathcal{F}_1$, $\pi_\ell = 0$ and $\pi_r \in [w, 1]$ as κ varies over $[0, 1]$. For $\pi_r \in [\frac{1}{2}, 1]$, $S^I = w\pi_r(3 - 2\pi_r)$, $S^P = w/\pi_r$ and $S^A = w(2 - 1/\pi_r)$, so

$$V_{\mathcal{F}_1}(\pi_r) = w\big[p_I \pi_r (3 - 2\pi_r) + (p_P - p_A)/\pi_r + 2p_A\big], \qquad V'_{\mathcal{F}_1}(\pi_r) = \frac{w}{\pi_r^2}\,\Psi(\pi_r),$$

with $\Psi(\pi_r) := p_I \pi_r^2 (3 - 4\pi_r) - (p_P - p_A)$ and $\Psi'(\pi_r) = 6 p_I \pi_r (1 - 2\pi_r)$.

If $p_I > 0$, then $\Psi' < 0$ on $(\frac{1}{2}, 1)$, so Ψ is strictly decreasing, with $\Psi(\frac{1}{2}) = p_I/4 - (p_P - p_A)$ and $\Psi(1) = -p_I - (p_P - p_A)$. The three cases follow from the sign of Ψ on $[\frac{1}{2}, 1]$, and substituting the first-order condition (59) into $V_{\mathcal{F}_1}$ gives the interior value $V_1^* = 2w[3p_I \pi_r^*(1 - \pi_r^*) + p_A]$.

If $p_I = 0$, then $V_{\mathcal{F}_1}(\pi_r) = w[(p_P - p_A)/\pi_r + 2p_A]$ is monotone in π_r on $[\frac{1}{2}, 1]$ according to the sign of $p_P - p_A$: decreasing if $p_P > p_A$, with maximum at the kink and value $2wp_P$; increasing if $p_P < p_A$, with maximum at full revelation and value $w(p_P + p_A) = w$; and constant equal to w when $p_P = p_A = \frac{1}{2}$, the announced knife-edge.

Step 4 (The cell $\mathcal{C}^{GT,GT}$). By Lemma 16, $V^{GT,GT}$ is convex in π_ℓ along each π_r-slice, so its supremum on each slice is attained at a slice endpoint. Each endpoint lies on $\mathcal{F}_1$ ($\pi_\ell = 0$),

on $\partial\overline{\mathcal{F}_2}$ ($\Delta_A = 0$), on the boundary $\Delta_P = 0$ separating $\mathcal{C}^{GT,GT}$ from $\mathcal{C}^{GB,GT}$, or on the lower half ($\pi_r = \frac{1}{2}$). When $p_I = p_P = 0$ the convexity is weak, but then $V^{GT,GT} = p_A S^A_{GT}$ depends only on π_r, is constant in π_ℓ, and the same boundary attribution holds. By continuity (Step 1) together with Steps 2, 3 and 5, each such boundary value is bounded by $\max\{V_1^*, V_2^*\}$.

Step 5 (The cell $\mathcal{C}^{GB,GT}$). Compare any $(\pi_\ell, \pi_r) \in \mathcal{C}^{GB,GT}$ to the $\mathcal{F}_1$ point at the same favorable posterior, $(0, \pi_r)$. We show $V^{GB,GT}(\pi_\ell, \pi_r) \le V_{\mathcal{F}_1}(\pi_r)$ in four claims.

(a) The cell is empty unless $w > \frac{1}{4}$. On it $\Delta_A \ge 0$, which requires $\pi_r \le 3w/(1+2w)$; combined with $\pi_r > \frac{1}{2}$ this forces $w > \frac{1}{4}$. Assume $w > \frac{1}{4}$ for the rest of the step, so that $\pi_r \in \left(\frac{1}{2},\, 3w/(1+2w)\right] \subset (\frac{1}{2}, \frac{3}{4}]$.

(b) The anti-partisan term matches exactly. $S^A_{GT}(\pi_\ell, \pi_r) = w(2 - 1/\pi_r) = S^A_{\mathcal{F}_1}(\pi_r)$, since S^A_{GT} does not depend on π_ℓ.

(c) The independent term is weakly worse than on $\mathcal{F}_1$. Expanding both sides and collecting terms,

$$S^I_{\mathcal{F}_1}(\pi_r) - S^I(\pi_\ell, \pi_r) \;=\; \pi_\ell(\pi_r - w)(3 - 2\pi_r - 2\pi_\ell). \tag{70}$$

On this cell $\pi_r \le \frac{3}{4}$ and $\pi_\ell < w \le \frac{1}{2}$, so $3 - 2\pi_r - 2\pi_\ell \ge 3 - \frac{3}{2} - 1 = \frac{1}{2} > 0$ and the gap is nonnegative.

(d) The pro-partisan term is weakly worse than on $\mathcal{F}_1$. Likewise,

$$\begin{aligned} S^P_{\mathcal{F}_1}(\pi_r) - S^P_{GB}(\pi_\ell, \pi_r) \;&=\; \frac{\pi_\ell(\pi_r - w)\, Q(\pi_\ell, \pi_r)}{\pi_r(\pi_r - \pi_\ell)^2}, \\ Q(\pi_\ell, \pi_r) \;&:=\; \pi_\ell(4\pi_r^2 - 2\pi_r - 1) + \pi_r(3 - 4\pi_r), \end{aligned} \tag{71}$$

so it suffices that $Q \ge 0$. The cell condition $\Delta_A \ge 0$ bounds $\pi_\ell \le \pi_\ell^A := \pi_r[3w - (1 + 2w)\pi_r]/(w + \pi_r - 2\pi_r^2)$, and the zero of $Q(\cdot, \pi_r)$ is $\pi_\ell^Q := \pi_r(3 - 4\pi_r)/(1 + 2\pi_r - 4\pi_r^2)$. Placing the two over a common denominator and simplifying the numerator,

$$\pi_\ell^Q - \pi_\ell^A = \frac{8\left(\frac{1}{2} - w\right)\pi_r^2(1 - \pi_r)^2}{(1 + 2\pi_r - 4\pi_r^2)(w + \pi_r - 2\pi_r^2)} \;\ge\; 0, \tag{72}$$

both denominator factors being positive on the range of (a): the first because $1 + 2\pi_r - 4\pi_r^2 > 0$ for $\pi_r < (1 + \sqrt{5})/4 \approx 0.809$, the second because $\pi \mapsto w + \pi - 2\pi^2$ is concave with positive values at both endpoints of that range. Since the coefficient $4\pi_r^2 - 2\pi_r - 1$ of π_ℓ in Q is negative there, Q is decreasing in π_ℓ, and $\pi_\ell \le \pi_\ell^A \le \pi_\ell^Q$ gives $Q \ge 0$.

Combining (b), (c) and (d),

$$V^{GB,GT}(\pi_\ell, \pi_r) \;\le\; V_{\mathcal{F}_1}(\pi_r) \;\le\; V_1^*. \tag{73}$$

Step 6 (The cell $\mathcal{C}^{GB,GB}$). This is the crux, and it is where the anti-matched comparison of Lemma 19 is used. Set $A := S^A_{GB}$ and $\hat{\pi} := w/(2w - A)$. Lemma 17 gives $S^P_{GB} + A \leq 2w$ and Lemma 18 gives $S^I - w \leq \Phi_w(A)$; both hypotheses of Lemma 19 therefore hold, so

$$V^{GB,GB}(\pi_\ell, \pi_r) \;\leq\; V_{\mathcal{F}_1}(\hat{\pi}) \;\leq\; V_1^*. \tag{74}$$

Synthesis. Steps 2–6 give the cell-by-cell domination structure recorded in Table 4. The only active region not dominated by $\mathcal{F}_1$ is $\overline{\mathcal{F}_2}$, on which V coincides with the smooth $V^{GT,GB}$. Hence

$$\max_{\mathcal{D}} V \;=\; \max\Big\{\max_{\mathcal{F}_1} V,\; \max_{\overline{\mathcal{F}_2}} V^{GT,GB}\Big\} \;=\; \max\{V_1^*, V_2^*\}.$$

The trichotomy of Step 3 characterizes V_1^*; V_2^* is attained on the compact set $\overline{\mathcal{F}_2}$, with first-order or boundary conditions characterizing the maximizer as stated. ■

Table 4: Cell-by-cell domination. Every region other than $\mathcal{F}_1$ and $\mathcal{F}_2$ is dominated by an explicit design in $\mathcal{F}_1$. The last column records that no step in the domination argument uses $p_I > 0$, which is what licenses the reduction at $p_I = 0$ in Appendix I.

Region	Dominated by	Mechanism	Step	Needs $p_I > 0$?
$\pi_r \leq \frac{1}{2}$	$\mathcal{F}_1$ at $\pi_r = \frac{1}{2}$	$S^I \leq w$, $S^P \leq 2w$, $S^A = 0$	2	no
$\mathcal{C}^{GT,GT}$	slice endpoints	convexity in π_ℓ (Lem. 16)	4	no
$\mathcal{C}^{GB,GT}$	$\mathcal{F}_1$ at same π_r	S^A matches; S^I, S^P gaps ≥ 0	5	no
$\mathcal{C}^{GB,GB}$	$\mathcal{F}_1$ at $\hat{\pi} = \frac{w}{2w - S^A_{GB}}$	Lem. 17+18 into 19	6	no
$\mathcal{F}_1$	— (trichotomy)	single-peakedness of $V_{\mathcal{F}_1}$	3	no
$\mathcal{F}_2 = \mathcal{C}^{GT,GB}$	— (survives)	characterized in App. H	—	—

Remark 20 (Why the anti-matched map is the right one) *The map $(\pi_\ell, \pi_r) \mapsto \hat{\pi} = w/(2w - S^A_{GB})$ of Lemma 19 matches the anti-partisan value* exactly. *The two-target trade-off* (63) *then delivers $S^P_{GB} \leq S^P_{\mathcal{F}_1}(\hat{\pi})$ and the envelope* (64) *delivers $S^I \leq S^I_{\mathcal{F}_1}(\hat{\pi})$, with simultaneous equality on $\partial\mathcal{C}^{GB,GB} \cap \overline{\mathcal{F}_1}$. Matching on any other audience would not close: it is the anti-partisan whose counter-design ties the other two together.*

H The leader's problem on the two families

Appendix G shows that the optimum survives in only two places. This appendix reads that conclusion as a reduction and then characterizes the optimum within each piece. Both

lemmas below are corollaries of the domination argument just given; the new content is the characterization of V_2^* and the threshold $w > \frac{1}{4}$.

Lemma 21 (Two candidate families) *Fix $w \in (0, \frac{1}{2}]$ and $(p_I, p_P, p_A) \in \Delta^2$. There is an optimum of the leader's problem (17) lying in $\mathcal{F}_1 \cup \overline{\mathcal{F}_2}$, and*

$$V^*(w, p_I, p_P, p_A) \;=\; \max\Big\{\max_{\mathcal{F}_1} V,\; \max_{\overline{\mathcal{F}_2}} V^{GT,GB}\Big\}. \tag{75}$$

Along $\mathcal{F}_1$ the unfavorable realization is fully revealing ($\pi_\ell = 0$) and both partisan supports are piecewise linear in κ with a kink at κ_{ret}. On $\mathcal{F}_2$ the pro-partisan plays GT and the anti-partisan plays GB.

Proof. Immediate from Table 4: Steps 2, 4, 5 and 6 of the proof of Proposition 6 dominate every design outside $\mathcal{F}_1 \cup \overline{\mathcal{F}_2}$ by a design in $\mathcal{F}_1$, and both maxima are attained by Lemma 11. The description of $\mathcal{F}_1$ is Lemma 1; the description of $\mathcal{F}_2$ is its definition (46). ■

The two sets correspond to qualitatively distinct communication strategies. Family $\mathcal{F}_1$ *concedes* the unfavorable state: when $\omega = L$ the message ℓ fires with probability $1 - \kappa$ and rules the state out, so the leader draws no further support from that realization. The single parameter κ controls both the weight on the favorable realization and where π_r lands, and three salient designs sit inside $\mathcal{F}_1$: full revelation $\kappa = 0$, the independent-optimal design pinning π_r at the tangency $\frac{3}{4}$, and the pro-retention design $\kappa = \kappa_{\text{ret}}$ pinning π_r at the voting threshold $\frac{1}{2}$. Region $\mathcal{F}_2$, by contrast, uses both realizations actively: the anti-partisan is driven out of her cheap GT slant into the costlier GB hedge, which contributes positively to leader support in state R, while the pro-partisan pays a retention cost. The leader trades pro-retention for anti-partisan poaching.

Lemma 22 (Subregions within $\mathcal{F}_1$) *Fix $w \in (0, \frac{1}{2}]$ and $(p_I, p_P, p_A) \in \Delta^2$. The maximum of $V(w, \kappa, 0;\, p)$ over $\kappa \in [0, 1]$ is attained at a κ^* inducing a favorable posterior π_r^* in exactly one of three regimes:*

(i) Pro-partisan retention *($\pi_r^* = \frac{1}{2}$). If $p_P - p_A \geq p_I/4$, then $\kappa^* = \kappa_{\text{ret}}$, both partisans play N, and $V_1^* = w(p_I + 2p_P)$.*

(ii) Independent targeting *($\pi_r^* \in (\frac{1}{2}, 1)$). If $-p_I < p_P - p_A < p_I/4$, then π_r^* is the unique solution in $(\frac{1}{2}, 1)$ of (59), $\kappa^* \in (0, \kappa_{\text{ret}})$ satisfies $\pi_r(\kappa^*, 0) = \pi_r^*$, the pro-partisan plays N and the anti-partisan plays GT.*

(iii) Full revelation *($\pi_r^* = 1$). If $p_P - p_A \leq -p_I$, then $\kappa^* = 0$ and $V_1^* = w$.*

In regime (ii), π_r^* is continuous and strictly decreasing in $p_P - p_A$, sweeps $(\frac{1}{2}, 1)$ monotonically as $p_P - p_A$ moves from $p_I/4$ down to $-p_I$, and crosses the tangency $\frac{3}{4}$ exactly when $p_P = p_A$. The maximized value is continuous in (p_I, p_P, p_A) across the three regimes.

Proof. This is Step 3 of the proof of Proposition 6; on $\mathcal{F}_1$ the statement coincides with Proposition 3, proved in Appendix D. Monotonicity of π_r^* follows from (59) because $\pi \mapsto \pi^2(3-4\pi)$ is strictly decreasing on $(\frac{1}{2}, 1)$, with values $\frac{1}{4}$ at $\frac{1}{2}$, 0 at $\frac{3}{4}$ and -1 at 1. ■

Three forces at the first-order condition. For $\pi_r > \frac{1}{2}$ the $\mu = 0$ specializations are $S^P = w/\pi_r$, $S^A = w(2 - 1/\pi_r)$ and $S^I = w\pi_r(3 - 2\pi_r)$, so

$$\frac{1}{w} \cdot \frac{dV_{\mathcal{F}_1}}{d\pi_r} = \underbrace{p_I(3-4\pi_r)}_{\text{independent effect}} + \underbrace{\frac{p_A - p_P}{\pi_r^2}}_{\text{net partisan effect}}. \tag{76}$$

The independent effect is positive below the tangency $\frac{3}{4}$, zero at it, negative above. The net partisan effect decomposes into a marginal retention loss p_P/π_r^2 and a marginal poaching gain p_A/π_r^2; the two are identical in scale because a unit shift in π_r costs exactly one unit of pro-support and buys exactly one unit of anti-support. The three regimes are the three sign configurations of (76): in (i) the retention loss swamps even the maximal independent gain, attained at $\pi_r = \frac{1}{2}$, so the leader stops at the threshold; in (ii) the two balance at an interior π_r^*, with $p_P = p_A$ collapsing to the independent-only tangency; in (iii) the poaching gain dominates even at $\pi_r = 1$ and the leader reveals.

Geometry on the simplex. In the (p_I, p_P)-plane with $p_A = 1 - p_I - p_P$ the two boundaries are

$$p_P = \tfrac{1}{2} - \tfrac{3}{8}p_I \quad \text{(retention/interior)}, \qquad p_P = \tfrac{1}{2} - p_I \quad \text{(interior/full revelation)},$$

both meeting at the symmetric-partisans point $(p_I, p_P) = (0, \frac{1}{2})$ and fanning into the simplex as p_I increases; region (ii) is a triangle with apex there. The pure-pro vertex $(0, 1, 0)$ sits in (i), the pure-anti vertex $(0, 0, 1)$ in (iii), and the pure-independent vertex $(1, 0, 0)$ in (ii) at $\pi_r^* = \frac{3}{4}$ — the concavification benchmark of Lemma 14(i).

The objective on $\mathcal{F}_2$. By construction the pro plays GT and the anti plays GB there, so

$$\begin{aligned} V^{GT,GB}(\kappa, \mu; w, p) = p_I\, S^I(w, \kappa, \mu) + p_P\, \frac{w(1-\kappa+\mu)}{1-\kappa} \\ + p_A\, \frac{[(1-\kappa)(1-\mu) + \kappa\mu]\, B'}{1-\kappa-\mu}, \end{aligned} \tag{77}$$

with S^I given by (25) and B' by (60). In the symmetric coordinates (42) this becomes

$$V^{GT,GB} \;=\; p_I\,\mathcal{I}(a,b;t) + p_P\,\mathcal{P}(a;t) + p_A\,\mathcal{A}(a,b;t), \tag{78}$$

with $\mathcal{I} = S^I$, $\mathcal{P} = S^P_{GT}$, $\mathcal{A} = S^A_{GB}$ reducing to

$$\mathcal{I}(a,b;t) = \tfrac{1}{4}\big(a^2b + a^2t - ab^2 - abt + b^2t - 3t + 2\big), \tag{79}$$

$$\mathcal{P}(a;t) = \frac{1-at}{1+a}, \qquad \mathcal{A}(a,b;t) = \frac{b(a-t)(1+ab)}{(a+b)^2}. \tag{80}$$

Note that $\mathcal{P}$ does not depend on b: the pro-partisan playing GT is exposed only to the unfavorable posterior. The domain is $\mathcal{F}_2^{ab}$ of (50).

Lemma 23 (Maximum on $\mathcal{F}_2$) *Fix $w \in (0, \frac{1}{2}]$ and $(p_I, p_P, p_A) \in \Delta^2$, and write V_2^* for the maximum of $V^{GT,GB}$ over $\overline{\mathcal{F}_2}$.*

(i) Exact interior characterization. *In the coordinates (42), any interior maximizer $(a^*, b^*) \in \mathcal{F}_2^{ab}$ satisfies $\partial_a V^{GT,GB} = \partial_b V^{GT,GB} = 0$, namely*

$$0 = \frac{p_I}{4}(b+t)(2a-b) - \frac{p_P(1+t)}{(1+a)^2} + \frac{p_A\,b\,\big(2ab^2 + abt - a - b^2t + b + 2t\big)}{(a+b)^3}, \tag{81}$$

$$0 = \frac{p_I}{4}(a-2b) + \frac{p_A\big(2a^2b + a - b\big)}{(a+b)^3}, \tag{82}$$

where (82) is $\partial_b V^{GT,GB} = 0$ after division by the strictly positive factor $a - t = 2(w - \pi_\ell)$. For $p_I > 0$ the system is solved explicitly in inverse form: every interior critical point corresponds to the shares

$$p_I = \frac{1}{1 + R_A + R_P}, \qquad p_A = \frac{R_A}{1 + R_A + R_P}, \qquad p_P = \frac{R_P}{1 + R_A + R_P}, \tag{83}$$

where

$$R_A(a,b) = \frac{(2b-a)(a+b)^3}{4\big(2a^2b + a - b\big)},$$

$$R_P(a,b;t) = \frac{(1+a)^2}{1+t}\left[\frac{(b+t)(2a-b)}{4} + R_A(a,b)\,\frac{b\big(2ab^2 + abt - a - b^2t + b + 2t\big)}{(a+b)^3}\right]. \tag{84}$$

The interior stationary locus is the image of $\{(a,b) \in \mathcal{F}_2^{ab} : R_A, R_P > 0\}$ under (83).

(ii) Locus restriction at $w = \frac{1}{2}$. *At $w = \frac{1}{2}$ the restriction of $V^{GT,GB}$ to the L-indifference locus*

$\Delta_A = 0$, *equivalently* $\kappa = (1-\mu)/(3-2\mu)$, *is*

$$\widetilde{V}(\mu; p) = p_I\, \widetilde{S}^I(\mu) + p_P\, \frac{1+\mu-\mu^2}{2-\mu} + p_A\, \frac{1-\mu}{3-2\mu}, \qquad \mu \in \left[1-\tfrac{1}{\sqrt{2}},\, 1\right], \tag{85}$$

with $\widetilde{S}^I(\mu) \equiv S^I\big(\frac{1}{2}, (1-\mu)/(3-2\mu), \mu\big)$; *the maximum over the locus solves* $\widetilde{V}'(\mu^*) = 0$, *with* $\kappa^* = (1-\mu^*)/(3-2\mu^*)$.

(iii) Partisans only at $w = \frac{1}{2}$. *When in addition* $p_I = 0$, V_2^* *is attained on the locus and* $\widetilde{V}'(\mu) = 0$ *reduces to*

$$(2-\mu)^2 \;=\; p_P\big(4\mu^4 - 28\mu^3 + 70\mu^2 - 76\mu + 31\big), \tag{86}$$

which has a unique solution $\mu^*(p_P)$ *in* $((3-\sqrt{3})/2, 1)$ *for every* $p_P \in (0,1)$. *Then* $V_2^* > \frac{1}{2}$ *if and only if* $p_P > \underline{p}_P(\frac{1}{2}) \equiv (4+3\sqrt{3})/22 \approx 0.418$.

Proof. (i) On the open cell $V^{GT,GB}$ is smooth, so an interior maximizer is stationary. Substituting the recovery map (43) into S^I, S^P_{GT} and S^A_{GB} and simplifying yields (79)–(80) as polynomial identities in (a, b, t). Differentiating (78) in a gives (81) directly. For the b-derivative, $\partial_b \mathcal{P} = 0$ while

$$\partial_b \mathcal{I} = \tfrac{1}{4}(a-t)(a-2b), \qquad \partial_b \mathcal{A} = \frac{(a-t)\big(2a^2b + a - b\big)}{(a+b)^3},$$

so $\partial_b V^{GT,GB} = (a-t)\big[\frac{p_I}{4}(a-2b) + p_A(2a^2b+a-b)/(a+b)^3\big]$; since $a > t$ on the cell this vanishes iff the bracket does, which is (82). For $p_I > 0$, (82) is linear in p_A/p_I and gives $p_A/p_I = R_A(a,b)$; substituting into (81), itself linear in p_P/p_I, gives $p_P/p_I = R_P(a,b;t)$, and normalizing yields (83). The cell equivalences are (49), whose prefactors are positive on $\mathcal{F}_2^{ab}$.

(ii) At $w = \frac{1}{2}$, substituting the locus $\kappa = (1-\mu)/(3-2\mu)$ into $V^{GT,GB}$ collapses the partisan terms to $p_P\frac{1+\mu-\mu^2}{2-\mu} + p_A\frac{1-\mu}{3-2\mu}$ and the independent term to $p_I\widetilde{S}^I(\mu)$, giving (85).

(iii) With $p_I = 0$ the independent term vanishes and, holding μ fixed, $V^{GT,GB}$ is strictly increasing in κ throughout the GB region, so the unrestricted maximum is attained at the highest feasible κ — the locus. Then $\widetilde{V}(\mu) = p_P\frac{1+\mu-\mu^2}{2-\mu} + (1-p_P)\frac{1-\mu}{3-2\mu}$; setting $\widetilde{V}'(\mu) = 0$ and clearing denominators gives (86). The quartic in the bracket and $(2-\mu)^2$ cross once on the admissible range, so the root $\mu^*(p_P)$ is unique; substituting it and comparing $\widetilde{V}(\mu^*)$ to the truth value $\frac{1}{2}$ gives the threshold. ■

Remark 24 (Interior versus boundary optima) *Part (iii) is the only case in which the optimum provably lies on the L-indifference locus. With independents present* ($p_I > 0$), *or for* $w < \frac{1}{2}$,

the maximizer is instead an interior point with $\Delta_A(\kappa^*, \mu^*) < 0$, characterized exactly by the first-order system of part (i). The locus $\Delta_A = 0$ is therefore a convenient anchor, not the solution.

Remark 25 (No closed form) *Away from the knife-edge of Lemma 23(iii) the maximizer of $V^{GT,GB}$ is located exactly, though not in closed form. It is one of two things. Either it is an interior critical point, pinned by the first-order system of part (i); or it lies on the boundary $\Gamma_A = 0$, where b is pinned to the positive root*

$$b_A(a;t) = \frac{-(a+t) + \sqrt{(a+t)^2 - 4(a+1-t)\big[t - a(1-t)\big]}}{2(a+1-t)}, \tag{87}$$

and the problem reduces to maximizing $V^{GT,GB}\big(a, b_A(a;t)\big)$ in a alone, subject to $t < a < 1$, $0 < b_A < 1$ and $\Gamma_P \leq 0$. That one-dimensional problem is the analogue of parts (ii)–(iii); equivalently, the tangency system

$$V_a\,\Gamma_{A,b} - V_b\,\Gamma_{A,a} = 0, \qquad \Gamma_A = 0,$$

holds, with $\Gamma_{A,a} = b^2 + b + t - 1$ and $\Gamma_{A,b} = 2(a+1-t)b + a + t$.

Eliminating b between the two first-order conditions yields a univariate polynomial of degree 25 in a for generic (w, p_P, p_A). The forward map from shares to (a^*, b^*) therefore has no useful closed form, and the operative exact description is the rational inverse parameterization (83) together with the selection rule that the maximizer is the value-maximizing real root in $\overline{\mathcal{F}_2^{ab}}$. The partisan-targeting boundary in Figure 8 is computed from that rule on a grid; the locus $\Delta_A = 0$ is a close numerical proxy over the range plotted, and the replication package reports the distance between the two.

Lemma 26 (Partisan targeting requires $w > \frac{1}{4}$) *Fix $w \in (0, \frac{1}{4}]$ and $(p_I, p_P, p_A) \in \Delta^2$. Then $V_2^* \leq V_1^*$: the partisan-targeting regime is empty and the optimum lies in $\mathcal{F}_1$.*

Proof. On $\mathcal{F}_2$ the pro plays GT and the anti plays GB, so $V = p_I S^I + p_P S^P_{GT} + p_A S^A_{GB}$. Write $A := S^A_{GB} \in [0, w]$ and $\hat{\pi} := w/(2w - A) \in [\frac{1}{2}, 1]$ as in Lemma 19. We dominate each audience term by its $\mathcal{F}_1$ value at $\hat{\pi}$.

Anti-partisan: equal to A by construction, (68).

Independents: the envelope inequality (64) uses only Bayes plausibility and $\Delta_A < 0$, both of which hold on $\mathcal{F}_2$, while $S^I_{\mathcal{F}_1}(\hat{\pi}) - w = \Phi_w(A)$ by (68); hence $S^I \leq S^I_{\mathcal{F}_1}(\hat{\pi})$.

Pro-partisan: since $S^P_{\mathcal{F}_1}(\hat{\pi}) = 2w - A$, it suffices that $S^P_{GT} + S^A_{GB} \leq 2w$. Note that this is *not* Lemma 17, which bounds $S^P_{GB} + S^A_{GB}$; the pro plays GT here, and the corresponding

bound is the one that fails for $w > \frac{1}{4}$. From $S^P_{GT} - 2w = (\pi_\ell - w)/(1 - \pi_\ell)$,

$$S^P_{GT} + S^A_{GB} - 2w = (w - \pi_\ell)\,\Theta, \qquad \Theta := \frac{(\pi_r + \pi_\ell - 2\pi_r\pi_\ell)(2\pi_r - 1)}{(\pi_r - \pi_\ell)^2} - \frac{1}{1 - \pi_\ell},$$

and since $w - \pi_\ell \geq 0$ the claim is $\Theta \leq 0$. In the coordinates (42), using $\pi_r - \pi_\ell = \frac{a+b}{2}$, $\pi_r + \pi_\ell - 2\pi_r\pi_\ell = \frac{1+ab}{2}$ and $1 - \pi_\ell = \frac{1+a}{2}$, this is

$$F(a,b) := (a+b)^2 - b(1+a)(1+ab) = (1 - a - a^2)b^2 - (1-a)b + a^2 \;\geq\; 0. \tag{88}$$

By the dictionary (44), $w \leq \frac{1}{4}$ forces $a \geq 1 - 2w \geq \frac{1}{2}$. For $a \in [\frac{1}{2}, 1]$ and $b \in [0,1]$: $F(a,0) = a^2 \geq 0$, $F(a,1) = 0$, and the vertex $b^* = \frac{1-a}{2(1-a-a^2)}$ satisfies $b^* \geq 1 \iff (1-a) \geq 2(1-a-a^2) \iff (2a-1)(a+1) \geq 0 \iff a \geq \frac{1}{2}$. So on $[0,1]$ the quadratic is either monotone decreasing (if $1 - a - a^2 > 0$, vertex to the right) or concave with an endpoint minimum (if $1 - a - a^2 \leq 0$), and in both cases $\min_{[0,1]} F = \min\{F(a,0), F(a,1)\} = 0$. Hence $\Theta \leq 0$ and $S^P_{GT} \leq S^P_{\mathcal{F}_1}(\hat{\pi})$.

Combining the three, $V \leq V_{\mathcal{F}_1}(\hat{\pi}) \leq V_1^*$ pointwise on $\mathcal{F}_2$ and hence on its closure. ∎

Remark 27 (Sharpness of the threshold) *The binding step is $a \geq \frac{1}{2}$, i.e. $\pi_\ell \leq \frac{1}{4}$, which $\pi_\ell \leq w$ enforces exactly when $w \leq \frac{1}{4}$ by (44). For $a < \frac{1}{2}$ the discriminant of (88), namely $1 - 2a - 3a^2 + 4a^3 + 4a^4$, is positive (it vanishes at $a = \frac{1}{2}$), so $F < 0$ on an interval of b and $S^P_{GT} + S^A_{GB} > 2w$ becomes attainable. Thus for every $w > \frac{1}{4}$ the leader can reach designs with $\pi_\ell \in (\frac{1}{4}, w]$ on which combined partisan reach exceeds the single-audience cap $2w$ — the opening that makes partisan targeting nonempty precisely for $w > \frac{1}{4}$.*

I Persuading partisans only

Throughout this appendix $p_I = 0$, and we work in the coordinates (42). On $\overline{\mathcal{F}_2}$ the pro-partisan plays GT and the anti-partisan plays GB, with supports $\mathcal{P}(a;t) = (1 - at)/(1 + a)$ and $\mathcal{A}(a,b;t) = b(a-t)(1+ab)/(a+b)^2$ as in (80); $\mathcal{F}_2^{ab}$ is (50).

Proof of Proposition 7.

Step 0 (Reduction). By Table 4, no step of the domination argument uses $p_I > 0$, so Steps 2, 4 and 5 of the proof of Proposition 6 dominate every design outside $\overline{\mathcal{F}_2}$ by a design in $\mathcal{F}_1$ at $p_I = 0$ as well. By Step 3, the $\mathcal{F}_1$ value at $p_I = 0$ is $\max\{2w\,p_P,\, w\}$: monotone in π_r according to the sign of $p_P - \frac{1}{2}$, with the retention value $2w\,p_P$ at the kink and the revelation value w at $\pi_r = 1$. Hence

$$V^*(p_P) \;=\; \max\big\{2w\,p_P,\; w,\; V_2^*(p_P)\big\}, \qquad V_2^*(p_P) \;\equiv\; \max_{\overline{\mathcal{F}_2}}\big[p_P\,\mathcal{P} + (1 - p_P)\,\mathcal{A}\big]. \tag{89}$$

Part (i) is then immediate from Lemma 26: for $w \le \frac{1}{4}$, $V_2^* \le V_1^* = \max\{2w\, p_P, w\}$. Assume $w > \frac{1}{4}$, i.e. $t < \frac{1}{2}$, for the remainder.

Step 1 (The gain identity). From $\mathcal{P} = (1 - at)/(1 + a)$ and $2w = 1 - t$,

$$2w - \mathcal{P} \;=\; \frac{(1-t)(1+a) - (1-at)}{1+a} \;=\; \frac{a-t}{1+a},$$

so on $\overline{\mathcal{F}_2}$,

$$p_P\, \mathcal{P} + (1 - p_P)\, \mathcal{A} - 2w\, p_P \;=\; (a-t)\left[(1-p_P)\,\frac{b(1+ab)}{(a+b)^2} \;-\; \frac{p_P}{1+a}\right]. \tag{90}$$

Since $a \ge t$ on the cell, a design in $\overline{\mathcal{F}_2}$ strictly beats pro-retention if and only if $a > t$ and $G(a,b) > p_P/(1-p_P)$, where

$$G(a,b) \;\equiv\; \frac{b\,(1+ab)\,(1+a)}{(a+b)^2}. \tag{91}$$

Step 2 (The gain envelope and the upper threshold). For fixed a,

$$\frac{\partial G}{\partial b} \;=\; \frac{(1+a)\big(a - b(1-2a^2)\big)}{(a+b)^3},$$

so along each a-slice G is single-peaked with peak at the *ridge* $b = a/(1-2a^2)$ when $a \le \frac{1}{2}$ — exactly when that peak lies in $[0,1]$ — and is increasing on $b \in [0,1]$ otherwise. On the ridge, $1 + ab = (1-a^2)/(1-2a^2)$ and $a + b = 2a(1-a^2)/(1-2a^2)$, so the peak value collapses to

$$G\Big(a, \tfrac{a}{1-2a^2}\Big) \;=\; \frac{1}{4a(1-a)}, \tag{92}$$

while at the endpoint $b = 1$, $G(a,1) = 1$. For $a \in (t, \frac{1}{2}]$ we have $a < 1 - t$, hence $a(1-a) > t(1-t)$ and $1/(4a(1-a)) < 1/(4t(1-t))$; and $1 \le 1/(4t(1-t))$ since $t < \frac{1}{2}$. Therefore

$$\sup_{a \ge t,\, b \in [0,1]} G(a,b) \;=\; \frac{1}{4t(1-t)}, \tag{93}$$

approached only as $(a,b) \to \big(t,\, t/(1-2t^2)\big)$ and not attained for $a > t$. The approaching ridge points lie in the open cell: at $a = t$,

$$\Gamma_A(t,b;t) = (b+t)^2 > 0, \qquad \Gamma_P(t,b;t) = -(1-t^2)(b+t) < 0,$$

so by continuity nearby ridge points satisfy the strict cell inequalities, and $t/(1-2t^2) \in$

$(0,1]$ exactly when $t \le \frac{1}{2}$, i.e. $w \ge \frac{1}{4}$. Combining (90)–(93) with the definition (33), which is equivalent to $\overline{p}_P/(1-\overline{p}_P) = 1/(4t(1-t))$:

- if $p_P < \overline{p}_P(w)$, then $p_P/(1-p_P) < \sup G$, and cell points near the corner give $V_2^*(p_P) > 2w\,p_P$;
- if $p_P \ge \overline{p}_P(w)$, then for every cell point with $a > t$ we have $G(a,b) < 1/(4t(1-t)) \le p_P/(1-p_P)$, so the bracket in (90) is strictly negative, while $a = t$ gives equality: $V_2^*(p_P) \le 2w\,p_P$, strictly at every non-degenerate design.

Evaluating at $p_P = \frac{1}{2}$, where $2w\,p_P = w$ and $p_P/(1-p_P) = 1$: $V_2^*(\frac{1}{2}) > w$ if and only if $4t(1-t) < 1$, i.e. if and only if $w > \frac{1}{4}$. This is an independent proof of the threshold of Lemma 26 on the $p_I = 0$ edge, and of its sharpness.

Step 3 (Convexity and the lower threshold). For a fixed design the partisan best responses do not depend on p_P, so $p_P \mapsto V(p_P;\kappa,\mu)$ is affine and $V_2^*(\cdot)$, a pointwise maximum of affine functions over the compact set $\overline{\mathcal{F}_2}$, is convex and continuous. At $p_P = 0$,

$$V_2^*(0) \;=\; \max_{\overline{\mathcal{F}_2}} \mathcal{A} \;<\; w\,:$$

by Lemma 14(iii), $S^A = w$ is attained only at full revelation, which lies outside $\overline{\mathcal{F}_2}$ since $\Gamma_P(1,1;t) = 2(1-t) > 0$ violates $\Gamma_P \le 0$; the strict bound follows by compactness. Since $V_2^*(\frac{1}{2}) > w$ by Step 2, and since for a convex function the difference quotients from $p_P = 0$ are nondecreasing, the set $\{p_P \in [0,\frac{1}{2}] : V_2^*(p_P) > w\}$ is an interval $(\underline{p}_P, \frac{1}{2}]$ with $\underline{p}_P \in (0,\frac{1}{2})$. The three-regime classification of part (ii) now assembles from (89): for $p_P \le \underline{p}_P$, $V_2^* \le w = \max\{2w\,p_P, w\}$ and full revelation is optimal; for $p_P \in (\underline{p}_P, \frac{1}{2}]$, $V_2^* > w \ge 2w\,p_P$; for $p_P \in (\frac{1}{2}, \overline{p}_P)$, $V_2^* > 2w\,p_P \ge w$ by Step 2; and for $p_P \ge \overline{p}_P$, $V_2^* \le 2w\,p_P$ and pro-retention is optimal. At $w = \frac{1}{2}$, $\overline{p}_P = 1$ and $\underline{p}_P = (4+3\sqrt{3})/22$ by Lemma 23(iii).

Step 4 (The targeting design). On the interior of $\mathcal{F}_2$ the value is smooth, and any interior maximizer satisfies the first-order system (81)–(82) of Lemma 23(i). At $p_I = 0$, condition (82) reduces to $2a^2b + a - b = 0$, i.e. the ridge $b = a/(1-2a^2)$ of Step 2 — as it must, since $\mathcal{P}$ does not depend on b, so $\partial_b V = (1-p_P)\partial_b \mathcal{A}$ and the optimal favorable posterior maximizes anti-partisan extraction given π_ℓ. Substituting into (81) and solving for the population ratio yields the second equation of (34). For $w < \frac{1}{2}$ the maximizer is interior (Remark 24), so (34) characterizes it; as $p_P \uparrow \overline{p}_P$ the gain (90) forces $a^* \downarrow t$, the degenerate corner at which $\mu \to 1$, $\kappa \to 0$ and $V \to 2w\,p_P$, which is the tangency of $V^*(\cdot)$ with the retention line visible in Figure 9(a). At $w = \frac{1}{2}$ the maximizer lies on the locus $\Delta_A = 0$ and is characterized by Lemma 23(ii)–(iii). ∎