Why and How: Public-Sphere Persuasion and Institutional Emergence

Laure Athias

January 15, 2026

Abstract

This paper develops a tractable theory of institutional emergence through persuasion in the public sphere. Citizens evaluate an institutional norm along two latent content dimensions: *merit* – whether the norm is desirable – and *feasibility* – whether it can be implemented. Endorsement is bundle-based: persuasive impact requires a matched pair of reasons that jointly establish *why* and *how*. Information arrives through two channels – state agents and peers – that may have different comparative advantage in supplying feasibility versus merit reasons (state agents supply feasibility more cheaply due to role-based spillovers from legal–rational bureaucratic expertise into public discourse), with exposure scaling the precision of each channel's arguments in a Bayesian learning environment. In the benchmark, bundles are assembled within channels, implying diminishing returns and substitutability across channels. I then study an extension in which citizens can assemble bundles across channels; when channels specialize by comparative advantage, cross-channel fit generates local complementarity at low reach and can steepen aggregate updating. To capture institutional emergence, I embed persuasion in a simple feedback dynamic: aggregate endorsement raises subsequent exposure in the public sphere. When cross-channel fit is strong, the resulting endorsement update map can become S-shaped, so small shocks to reach or content can trigger tipping-point transitions between low- and high-endorsement regimes; when bundling remains primarily within channels, endorsement evolves more smoothly. This yields a sharp distinction between disciplined legitimacy formation – when feasibility completes a contested merit case – and instrumental amplification, in which implementation talk raises endorsement for whichever merit discourse prevails. A welfare comparison highlights an internalization wedge in precision provision and motivates policy guardrails that preserve informational gains while limiting amplification risk.

1 Introduction

How do institutions emerge? We know far more about how institutions persist than about how they originate. Political philosophy (e.g., Castoriadis (1987)) emphasizes that institutions combine a functional role with a symbolic (meaning-making) dimension, yet we lack tractable economic models linking social meaning to institutionalization.

This paper develops a tractable theory of institutional emergence through persuasion in the public sphere. The central premise is that institutional endorsement is *bundle-based*: for a norm to become legitimate, citizens must encounter public reasons that jointly establish both its *merit* – why it is desirable – and its *feasibility* – how it can be implemented. I model these as two latent content dimensions, m and h, and embed persuasion in a Gaussian learning environment in which exposure scales the effective precision of arguments delivered by distinct channels.

The framework delivers three main insights. First, bundling creates a tight link between the *architecture of discourse* and aggregate endorsement. In the benchmark in which complete (m, h) bundles are assembled *within* channels, endorsement is increasing in exposure but features diminishing returns. Because channel-specific bundle arrival rates add, the marginal effect of one channel is reduced by the presence of the other: channels are substitutes at the endorsement stage. This provides a sharp baseline for empirical work and for interpretation: negative interactions in exposure need not reflect audience mechanics (overlap, saturation) per se, but can arise from how persuasive opportunities aggregate when bundles are channel-contained.

Second, allowing citizens to assemble bundles *across* channels fundamentally changes the exposure-to-endorsement mapping when channels specialize. When one channel disproportionately supplies feasibility talk while the other disproportionately supplies merit talk, cross-channel fit creates *local complementarity* at low reach: increasing exposure in either channel raises the probability that citizens encounter a matched pair. Comparative advantage on the production side generates specialization pressure in equilibrium (e.g., state agents supply h more cheaply due to role-based spillovers from legal–rational bureaucratic expertise into public discourse). This yields a clean conceptual distinction between two modes of legitimacy formation. When merit discourse is contestable and bundles are disciplined, feasibility arguments complete a persuasive case and facilitate legitimacy. Under strong specialization and weak contestation, feasibility arguments can instead become a *content-agnostic amplifier*: they raise endorsement for whichever merit narrative happens to dominate. In this sense, the model isolates a composition risk that is orthogonal to the efficiency gains from specialization.

Third, once exposure is embedded in a feedback loop – public endorsement increases subsequent reach through public-sphere propagation – cross-channel fit can steepen aggregate updating and generate nonlinear dynamics. When fit is strong, the endorsement update map becomes S-shaped: small changes in reach or content mix have little effect at very low endorsement, but can have large effects around intermediate levels, creating the possibility of tipping-point regime shifts and multiplicity. The dynamic result clarifies when persuasion produces gradual evolution versus discontinuous transitions: it is not persuasion *per se* that yields cascades, but persuasion combined with (i) endogenous reach and (ii) cross-channel complementarity created by specialization.

The welfare analysis highlights two distinct policy problems. The first is a standard internalization wedge: when producers do not fully internalize the common benefit of legitimacy, precision is under-provided. The second is specific to bundle-based persuasion under specialization: the endorsement system can become excessively sensitive when feasibility talk scales up endorsement without disciplining the underlying merit case. This motivates a set of *guardrails* that map directly into model primitives: (i) limiting extreme h-only specialization in the state-agent channel by embedding substantive professional norms alongside legal-rational expertise; (ii) strengthening the quality and contestability of merit discourse (media quality, education, and broader deliberative capacity);

(iii) investing in public-sphere epistemic infrastructure that creates common reference points; (iv) relaxing time and attention constraints that raise the opportunity cost of producing and processing public reasons; and (v) institutional reforms that increase internalization where public endorsement is politically consequential. A key implication is that policies which expand implementation capacity alone (higher π_S^h or higher E^S) can be double-edged: they may increase legitimacy formation when merit discourse is disciplined, but increase amplification risk when it is not.

Overall, the model highlights a classic tension in accounts of Enlightenment rationality. Horkheimer and Adorno (1944) warn that instrumental reason, detached from substantive ends, can facilitate domination; by contrast, Habermas (1989) emphasizes the democratic promise of a public sphere in which reasons are contestable. In my framework, since endorsement is bundle-based by construction, the key question is whether feasibility completes a contested merit case or converts a one-sided one: with strong contestation, feasibility supports disciplined legitimacy; with skewed merit and strong specialization, feasibility amplifies prevailing narratives.

2 A Reason-Based Model of Norm Endorsement

This section formalizes how public arguments induce citizens' endorsement of a social norm. Two content dimensions matter for persuasion: *merit* (is the norm desirable?) and *implementability* (is its institutionalization feasible?). Bringing such arguments into informal talk is not cheap talk: arguments are citable and contestable, and require information acquisition. Citizens incur information costs because a norm's legitimacy is increasing in the extent of public endorsement; higher legitimacy raises compliance and implementability, yielding a common benefit. Crucially, while the public sphere treats citizens as equals, state agents face lower marginal costs of producing *implementation* precision, reflecting Weberian legal–rational expertise embedded in bureaucratic roles (Weber (1922)).

2.1 Model Setup

A continuum of citizens i in a public sphere c at date t considers whether to endorse a norm as a rule for society. Let $y_{i,c,t} \in \{0,1\}$ denote the observed public stance, and

$$s_{i,c,t} \;\equiv\; \mathbb{E}[y_{i,c,t} \,|\, \mathcal{F}_{i,c,t}] \in [0,1]$$

be the endorsement propensity given citizen i's information set $\mathcal{F}_{i,c,t}$, generated by the public arguments to which she is exposed. Public arguments are reasons offered to an open audience and thus publicly accessible and contestable (Habermas (1989)). I define the *public sphere* in cell (c,t) as the set of venues and interactions in which reasons about common affairs are produced and exchanged. Concretely, it includes (i) mass and local media; (ii) open civic forums (associations, clubs, assemblies); and (iii) informal interpersonal communication. In the model, citizens observe only the subset of arguments they are exposed to; I formalize this below via the information set $\mathcal{F}_{i,c,t}$. I abstract from censorship: any agent can contribute to and access this discourse, though exposure varies across citizens.

Two latent dimensions matter for persuasion: a *merit* dimension m (is the norm optimal or desirable?) and an *implementation/how* dimension h (is its institutionalization feasible/can we make it work?). A necessary condition for endorsement is that citizens can jointly infer that the norm is desirable and implementable; persuasion therefore requires both dimensions. They draw on different knowledge bases (value-arguments vs. legal-rational expertise). Priors are Gaussian and independent:

$$m \sim \mathcal{N}(\mu_m, \sigma_m^2), \qquad h \sim \mathcal{N}(\mu_h, \sigma_h^2).$$

Two producer channels supply public reasons: $j = S$ (*state agents*: public employees spilling over from their roles) and $j = P$ (*peers*: the other citizens). Each channel's communication has (i) *reach* $E^j_{i,c,t} \geq 0$, the realized exposure intensity in (i, c, t), and (ii) *content precision* on the two dimensions $k \in \{m, h\}$.

For each channel $j \in \{S, P\}$ (state agents, peers) and dimension $k \in \{m, h\}$ (merit, how), producing precision $\pi^k_j \geq 0$ requires effort $e^k_j \geq 0$ and entails a convex cost $C^k_j(e^k_j)$ with $C^{k\prime}_j(e) > 0$ and $C^{k\prime\prime}_j(e) \geq 0$. Precision is generated by an increasing, concave technology $\pi^k_j = \upsilon^k_j(e^k_j)$ with $\upsilon^{k\prime}_j(e) > 0$ and $\upsilon^{k\prime\prime}_j(e) \leq 0$.

Assumption 1 (Role-based cost asymmetry on implementation precision) *There exists* $\delta > 0$ *such that for all* $x \geq 0$*,*

$$C^{h\prime}_S(x) \;\leq\; (1-\delta)\, C^{h\prime}_P(x).$$

Moreover, the two channels are symmetric on the merit dimension:

$$C^{m\prime}_S(x) = C^{m\prime}_P(x) \quad \forall x \geq 0, \quad \textit{and} \quad \upsilon^m_S = \upsilon^m_P.$$

Producers invest in precision because legitimacy is increasing in aggregate endorsement and higher legitimacy raises compliance and implementability, generating a common benefit $V(S_{c,t})$ with $V'(S) > 0$. *Baseline normalization.* I take $V(S) = B\,S$ with $B > 0$; all results below go through for any V with $V' > 0$ (and, if desired, $V'' \leq 0$). I model each channel as a non-atomistic producer: its choice of precision affects the information received by a positive mass of citizens (through reach), and hence equilibrium endorsement.

Internalization. Channels may internalize only a fraction of the common benefit generated by legitimacy. Let $\beta_j \in (0, 1]$ denote channel j's internalization weight. In the producer problem below, the common benefit BS is scaled by β_j.

For each channel j, individual exposure within (c, t) decomposes as

$$E^j_{i,c,t} = E^j_{c,t} + \eta^j_{i,c,t}, \qquad \mathbb{E}[\eta^j_{i,c,t} \mid c, t] = 0,$$

where $E^j_{c,t}$ is aggregate reach in public sphere (c, t) for channel j and $\eta^j_{i,c,t}$ is idiosyncratic reach. Assume $\eta^j_{i,c,t}$ has support such that $E^j_{i,c,t} \geq 0$ a.s. Agents condition on their realized exposure.

Assumption 2 (Within-period predetermined/atomistic exposures) *Within* (c, t)*, the aggregate exposures* $E^j_{c,t}$ *(for* $j \in \{S, P\}$*) are realized by many micro-broadcasts and information-environment constraints (e.g., communication technology, media-market reach). Any single citizen or micro-sender is negligible and aggregates* $E^j_{c,t}$ *are treated as predetermined within the period.*

Exposures $E^j_{c,t}$ are treated as predetermined within the period, but each channel's precision choice is non-atomistic: changing π^k_j shifts the informativeness of the signals received by the (positive) mass of citizens exposed through reach, and thus affects equilibrium endorsement.

Assumption 3 (Exposure–signal aggregation) *For each channel* $j \in \{S, P\}$ *and dimension* $k \in \{m, h\}$*, citizen* i *observes an aggregated signal*

$$\tilde{x}^k_{j,i,c,t} = \begin{cases} m + \tilde{\varepsilon}^m_{j,i,c,t}, & k = m, \\ h + \tilde{\varepsilon}^h_{j,i,c,t}, & k = h, \end{cases}$$

where, if $\pi_j^k E_{i,c,t}^j > 0$*,*

$$\tilde{\varepsilon}_{j,i,c,t}^k \sim \mathcal{N}\Big(0,\ (\pi_j^k E_{i,c,t}^j)^{-1}\Big).$$

If $\pi_j^k E_{i,c,t}^j = 0$*, no informative signal on* (j,k) *is observed (equivalently, the posterior equals the prior on that dimension from that channel). Signals are independent across* j *and* k *conditional on* (m,h)*.*

Citizen i's information set in (c,t) is

$$\mathcal{F}_{i,c,t} \;=\; \sigma\Big(E_{i,c,t}^S,\, E_{i,c,t}^P,\, \tilde{x}_{S,i,c,t}^m,\, \tilde{x}_{S,i,c,t}^h,\, \tilde{x}_{P,i,c,t}^m,\, \tilde{x}_{P,i,c,t}^h\Big).$$

Microfoundations. (i) The noise in $\tilde{x}_{j,i,c,t}^k$ is mean-zero. This reflects that arguments circulating in the public sphere are citable and contestable: fact claims can be checked, and counter-arguments are voiced. With many micro-senders and heterogeneous views, aggregation (or averaging over Poisson arrivals) washes out idiosyncratic distortions, yielding no systematic drift in either dimension. (ii) $E_{i,c,t}^j$ independent repetitions of a basic Gaussian signal (for m and for h), averaged, so precision adds to $\pi_j^k E_{i,c,t}^j$; or (iii) a continuous-time information flow with rate π_j^k observed over a window of length $E_{i,c,t}^j$, yielding cumulative precision $\pi_j^k E_{i,c,t}^j$. More $\pi_j^k E_{i,c,t}^j$ moves beliefs toward the truth (variance falls).

2.2 Analysis

2.2.1 Static analysis

Posterior with exposure scaling. Define prior precisions $\tau_m \equiv \sigma_m^{-2}$, $\tau_h \equiv \sigma_h^{-2}$, and the total effective precision on dimension $k \in \{m,h\}$,

$$Q_{k,i,c,t} \;\equiv\; \sum_{j\in\{S,P\}} \pi_j^k\, E_{i,c,t}^j.$$

Thus, reach scales informativeness linearly: each additional unit of exposure to channel j adds π_j^k units of precision on dimension k.

Lemma 1 *Under Assumption 3, posteriors decouple by dimension and are Gaussian with precision* $\tau_k + Q_{k,i,c,t}$*:*

$$m \mid \{\tilde{x}_{j,i,c,t}^m\}_j \sim \mathcal{N}\left(\frac{\tau_m\mu_m + \sum_j (\pi_j^m E_{i,c,t}^j)\,\tilde{x}_{j,i,c,t}^m}{\tau_m + \sum_j \pi_j^m E_{i,c,t}^j},\ \Big(\tau_m + \sum_j \pi_j^m E_{i,c,t}^j\Big)^{-1}\right),$$

$$h \mid \{\tilde{x}_{j,i,c,t}^h\}_j \sim \mathcal{N}\left(\frac{\tau_h\mu_h + \sum_j (\pi_j^h E_{i,c,t}^j)\,\tilde{x}_{j,i,c,t}^h}{\tau_h + \sum_j \pi_j^h E_{i,c,t}^j},\ \Big(\tau_h + \sum_j \pi_j^h E_{i,c,t}^j\Big)^{-1}\right).$$

Proof 1 (Sketch) *Conditional on each dimension* k*, signals are independent Gaussians with precisions* $\pi_j^k E_{i,c,t}^j$*. Conjugacy implies precisions add and the posterior mean is a precision-weighted average.*

Benchmark: within-channel bundling (no cross-channel fit). Fixing (i, c, t) and suppressing subscripts, for each channel $j \in \{S, P\}$ and dimension $k \in \{m, h\}$, the effective precision (or intensity) is

$$Q_j^k \;\equiv\; \pi_j^k \, E^j \;\geq 0.$$

In the benchmark, endorsement requires a *matched pair* of reasons (m, h) that is assembled *within the same channel.* A convenient Poisson microfoundation is that, in channel j, opportunities to encounter a persuasive merit argument arrive with precision Q_j^m and opportunities to encounter a persuasive implementation argument arrive with precision Q_j^h, independently. The probability that channel j delivers a complete (matched) bundle is then

$$\Lambda_j(Q_j^m, Q_j^h) \;\equiv\; \left(1 - e^{-Q_j^m}\right)\left(1 - e^{-Q_j^h}\right).$$

Let $\Lambda \equiv \Lambda_S + \Lambda_P$ denote the total arrival precision (or intensity) of complete bundles from either channel, and let endorsement be

$$s(E^S, E^P) \;=\; u(\Lambda),$$

where $u : \mathbb{R}_+ \to [0, 1)$ is $\mathcal{C}^2$, increasing and concave ($u' > 0$, $u'' \leq 0$). A canonical case is $u(\Lambda) = 1 - e^{-\Lambda}$.

Proposition 1 (Within-channel bundling implies substitutability across channels) *Under the benchmark above, endorsement is weakly increasing in each exposure and submodular:*

$$\frac{\partial s}{\partial E^S} \geq 0, \qquad \frac{\partial s}{\partial E^P} \geq 0, \qquad \frac{\partial^2 s}{\partial E^S \, \partial E^P} \leq 0.$$

Proof 2 (Sketch) *Since $\Lambda = \Lambda_S(E^S) + \Lambda_P(E^P)$, we have $\partial s / \partial E^j = u'(\Lambda)\, \Lambda_j'(E^j) \geq 0$. Moreover,*

$$\frac{\partial^2 s}{\partial E^S \, \partial E^P} = u''(\Lambda)\, \Lambda_S'(E^S)\, \Lambda_P'(E^P) \;\leq\; 0,$$

because $u'' \leq 0$ and $\Lambda_j' > 0$ whenever $\pi_j^m + \pi_j^h > 0$ and $E^j > 0$.

Interpretation. Monotonicity ($u' > 0$) captures that additional reach increases the precision of complete argument bundles and hence endorsement. Concavity ($u'' \leq 0$) reflects diminishing returns from saturation, limited attention, and audience overlap. In the benchmark, $\Lambda = \Lambda_S(E^S) + \Lambda_P(E^P)$ is additively separable across channels, so the cross–partial is negative only through the curvature of u: $s_{SP} = u''(\Lambda)\Lambda_S'(E^S)\Lambda_P'(E^P) \leq 0$. This is therefore a prediction about how *bundles* combine at the endorsement stage, not about exposure mechanics per se. I normalize $u'(0) = 1$.

Producer choice (benchmark only). In the benchmark, complete argument bundles are assembled within channels. I therefore model precision choices $\{\pi_j^m, \pi_j^h\}_{j \in \{S,P\}}$ as determined by within-channel optimization, and treat them as predetermined when studying cross-channel fit. This timing isolates the mechanism of interest: holding content mixes fixed, cross-channel fit changes the comparative statics of endorsement with respect to reach.

Channel $j \in \{S, P\}$ has aggregate reach $E_{c,t}^j \geq 0$ and chooses efforts $e_j^m, e_j^h \geq 0$, which generate precisions $\pi_j^k = \upsilon_j^k(e_j^k)$ for $k \in \{m, h\}$. Under within-channel bundling, the intensity of complete bundles from channel j is

$$\Lambda_j \;=\; \left(1 - e^{-\pi_j^m E_{c,t}^j}\right)\left(1 - e^{-\pi_j^h E_{c,t}^j}\right).$$

With $\Lambda \equiv \Lambda_S + \Lambda_P$ and endorsement $s = u(\Lambda)$, $u' > 0$, $u'' \le 0$, channel j chooses (e_j^m, e_j^h) to maximize an internalized benefit net of production costs:

$$\max_{e_j^m, e_j^h \ge 0} \ \beta_j \, B \, u(\Lambda) \ - \sum_{k \in \{m,h\}} C_j^k(e_j^k), \qquad j \in \{S, P\}.$$

Proposition 2 (Benchmark producer optimality conditions) *Fix (c,t) and suppress indices. Let $\Lambda = \Lambda_S + \Lambda_P$, with*

$$\Lambda_j(\pi_j^m, \pi_j^h; E^j) \ = \ (1 - e^{-\pi_j^m E^j})(1 - e^{-\pi_j^h E^j}), \qquad j \in \{S, P\}.$$

Channel j chooses $e_j^m, e_j^h \ge 0$, which generate $\pi_j^k = \upsilon_j^k(e_j^k)$, to maximize

$$\beta_j B \, u(\Lambda) \ - \sum_{k \in \{m,h\}} C_j^k(e_j^k).$$

At any interior optimum $(e_j^m, e_j^h) \gg 0$, the first-order conditions are, for $k \in \{m,h\}$ and $\ell \ne k$,

$$\beta_j B \, u'(\Lambda) \, E^j \, e^{-\pi_j^k E^j} \, (1 - e^{-\pi_j^\ell E^j}) \, \upsilon_j^{k\prime}(e_j^k) \ = \ C_j^{k\prime}(e_j^k). \tag{1}$$

Equivalently, defining the marginal cost of precision as $MC_j^k(e) \equiv C_j^{k\prime}(e)/\upsilon_j^{k\prime}(e)$, (1) can be written as

$$MC_j^k(e_j^k) \ = \ \beta_j B \, u'(\Lambda) \, E^j \, e^{-\pi_j^k E^j} \, (1 - e^{-\pi_j^\ell E^j}).$$

Interpretation. Condition (1) equates the marginal cost of improving precision in dimension k to its marginal benefit through bundle formation. Because the benefit term contains $(1 - e^{-\pi_j^\ell E^j})$, raising precision in one dimension increases the marginal value of raising precision in the other: within-channel bundling induces complementarity between m and h on the producer side. The internalization weight β_j scales incentives to invest in precision, while reach E^j scales the payoff to precision because more citizens are affected.

For small E^j, $(1 - e^{-\pi_j^\ell E^j}) = \pi_j^\ell E^j + o(E^j)$, so the marginal benefit in (1) scales locally as $(E^j)^2 \pi_j^\ell$, which is the force behind Proposition 3.

Proposition 3 (Comparative advantage yields (local) specialization pressure) *Consider the benchmark producer problem with within-channel bundling and normalize $u'(0) = 1$. Let each channel $j \in \{S, P\}$ internalize a fraction $\beta_j \in (0, 1]$ of the common benefit. Under Assumption 1, symmetry on the m-dimension, $\beta_S = \beta_P$, and interior solutions, there exists $\bar{E} > 0$ such that for reach levels $E^S = E^P \in (0, \bar{E}]$, the state-agent channel chooses weakly higher implementation precision than peers:*

$$\pi_S^h \ \ge \ \pi_P^h,$$

and hence a (weakly) more implementation-tilted mix:

$$\frac{\pi_S^h}{\pi_S^m} \ \ge \ \frac{\pi_P^h}{\pi_P^m}.$$

Proof 3 (Sketch) *For small x, $1 - e^{-x} = x + o(x)$. Thus, for small reach, within-channel bundling satisfies $\Lambda_j = (1 - e^{-\pi_j^m E^j})(1 - e^{-\pi_j^h E^j}) = (\pi_j^m E^j)(\pi_j^h E^j) + o((E^j)^2)$. With $u'(0) = 1$, the local marginal benefit of increasing π_j^h is proportional to $(E^j)^2 \pi_j^m$, and analogously the marginal benefit of raising π_j^m is proportional to $(E^j)^2 \pi_j^h$. Under symmetry on the m-dimension, channels face the same marginal cost and technology for raising π^m, while state agents have strictly lower marginal cost of raising π^h by Assumption 1. The resulting interior first-order conditions therefore imply $\pi_S^h \ge \pi_P^h$ for small enough reach, and thus the stated tilt.*

Interpretation. Because endorsement requires a complete (m,h) bundle, precision on either dimension has little value if the other is missing. At low reach, complete bundles are scarce and $\Lambda_j \approx (E^j)^2\pi_j^m\pi_j^h$, so the marginal return to increasing π_j^h is scaled by $(E^j)^2\pi_j^m$. In this scarcity region, a comparative advantage in implementation precision translates directly into higher chosen π^h for the state-agent channel (and hence a more implementation-tilted mix). The result is stated locally because the Poisson technology saturates: when $Q_j^h = \pi_j^h E^j$ is large, $e^{-Q_j^h}$ becomes small and the marginal gain from further increasing π_j^h collapses, weakening specialization pressure.

In what follows, I take the precision vector $\{\pi_j^m, \pi_j^h\}$ as given (determined by the benchmark within-channel production problem) and study how specialization shapes the mapping from exposures (E^S, E^P) to endorsement through cross-channel fit.

Extension: cross-channel fit under specialization. When channels specialize, citizens may assemble a matched pair across channels: a merit reason from one channel and an implementation reason from the other. A natural extension adds a cross-channel bundle intensity

$$\Lambda_\times \;\equiv\; \left(1-e^{-Q_S^m}\right)\left(1-e^{-Q_P^h}\right) \;+\; \left(1-e^{-Q_S^h}\right)\left(1-e^{-Q_P^m}\right),$$

and sets $s(E^S,E^P) = u(\Lambda_S + \Lambda_P + \Lambda_\times)$. Under strong specialization (e.g., $Q_S^m \simeq 0$ and $Q_P^h \simeq 0$), $\Lambda_\times$ dominates within-channel bundling, and the cross-partial can turn positive, capturing complementarity in exposures driven by cross-channel fit.

Lemma 2 (Cross-channel fit generates local complementarity) *Fix (i,c,t) and suppress subscripts. Let $f(x) \equiv 1-e^{-x}$ and $Q_j^k \equiv \pi_j^k E^j$ for $j \in \{S,P\}$ and $k \in \{m,h\}$. Let Λ_S and Λ_P be the within-channel bundle intensities defined above, and define the cross-channel fit intensity*

$$\Lambda_\times \;\equiv\; f(Q_S^m)f(Q_P^h) + f(Q_S^h)f(Q_P^m).$$

Let total bundle intensity be $\Lambda \equiv \Lambda_S + \Lambda_P + \Lambda_\times$ and endorsement be $s(E^S,E^P) = u(\Lambda)$, where $u : \mathbb{R}_+ \to [0,1)$ is $\mathcal{C}^2$ with $u'(0) > 0$. Then

$$\left.\frac{\partial^2 s}{\partial E^S\,\partial E^P}\right|_{E^S=E^P=0} \;=\; u'(0)\,\kappa, \qquad \kappa \;\equiv\; \pi_S^m\pi_P^h + \pi_S^h\pi_P^m.$$

In particular, if $\kappa > 0$ then there exists $\varepsilon > 0$ such that $\partial^2 s/\partial E^S \partial E^P > 0$ for all $(E^S,E^P) \in [0,\varepsilon]^2$.

Proof 4 (Sketch) *At $E^S = E^P = 0$, we have $Q_j^k = 0$ and $f(0) = 0$, $f'(0) = 1$. Thus Λ_S depends only on E^S and Λ_P only on E^P, so they contribute no cross-partial at the origin. The cross term satisfies*

$$\left.\frac{\partial^2 \Lambda_\times}{\partial E^S\,\partial E^P}\right|_0 = \pi_S^m\pi_P^h + \pi_S^h\pi_P^m = \kappa.$$

Moreover, $\partial\Lambda/\partial E^S\big|_0 = \partial\Lambda/\partial E^P\big|_0 = 0$, so by the chain rule, $s_{SP}|_0 = u'(0)\,\Lambda_{SP}|_0 = u'(0)\kappa$. Continuity of s_{SP} yields positivity on a neighborhood when $\kappa > 0$.

Interpretation. The benchmark assumes that citizens obtain a complete case for endorsement – a persuasive *why* (m) and a persuasive *how* (h) – from within the same channel. Specialization breaks this: if one channel is relatively better at supplying feasibility and the other at supplying normative justification, then complete bundles may be assembled *across* channels. The term $\Lambda_\times$

captures precisely this possibility: $f(Q_S^h)f(Q_P^m)$ is the probability of encountering an implementation argument via the state channel *and* a merit argument via the peer channel (and symmetrically $f(Q_S^m)f(Q_P^h)$ for the opposite pairing). Under strong specialization ($Q_S^m \simeq 0$, $Q_P^h \simeq 0$), most complete bundles arrive through cross-channel matching rather than within-channel bundling.

The parameter

$$\kappa \;\equiv\; \pi_S^m \pi_P^h + \pi_S^h \pi_P^m$$

summarizes the scope for cross-channel fit: it is large when the channels' content mixes are *non-aligned* (one loads on h while the other loads on m), and it collapses when mixes are proportional. Lemma 2 shows that, at low exposure, cross-channel fit induces local complementarity in reach: increasing E^S raises the marginal return to E^P because it makes it more likely that a citizen who hears a merit argument from peers also encounters an implementation argument from state agents, and conversely.

The emphasis on the origin reflects that, when exposures are small, $f(x) = 1 - e^{-x} \approx x$ and bundle-arrival probabilities are approximately linear in effective precisions; this yields a transparent closed-form cross-partial $u'(0)\kappa$.

Corollary 1 (Strong specialization) *Under strong specialization with $\pi_S^h > 0$, $\pi_P^m > 0$, and $\pi_S^m = \pi_P^h = 0$,*

$$\left.\frac{\partial^2 s}{\partial E^S\,\partial E^P}\right|_{E^S=E^P=0} \;=\; u'(0)\,\pi_S^h \pi_P^m \;>\; 0.$$

If, moreover, $u'(0) = 1$ (normalization), this equals $\pi_S^h \pi_P^m$.

Interpretation (Instrumental specialization as a neutral amplifier). Consider a role-specialized Weberian bureaucracy spilling over into the public sphere, with $\pi_S^h \gg \pi_S^m \simeq 0$. The state-agent channel then contributes feasibility (h) with little moral content (m). The cross-fit term $\kappa = \pi_S^m \pi_P^h + \pi_P^m \pi_S^h$ is dominated by the off-diagonal product $\pi_P^m \pi_S^h$, so feasibility talk from state agents mechanically amplifies whatever moral reasons peers supply. Normatively, such *instrumental specialization is a content-agnostic amplifier*: it increases the effectiveness of prevailing m, whether benign or malign[1]. Avoiding over-specialization is therefore a robustness concern, not merely an efficiency loss, especially at low exposures. Section 2.2.3 studies when this amplification is socially costly and which guardrails preserve informational gains while limiting amplification risk.

2.2.2 Dynamics: Feedback and regime shifts

The previous results characterize within-period effects. I now allow endorsement to feed back into future reach through public-sphere propagation, generating norm dynamics.

Let aggregate endorsement at date t be $S_t \in [0, 1]$. Public-sphere propagation implies that reach in the next period is increasing in current endorsement:

$$E_{t+1}^j \;=\; E^j(S_t), \qquad j \in \{S, P\}.$$

Given reach (E_{t+1}^S, E_{t+1}^P), total bundle precision is $\Lambda_{t+1} \equiv \Lambda(E_{t+1}^S, E_{t+1}^P)$, and endorsement updates through u:

$$S_{t+1} \;=\; u(\Lambda_{t+1}) \;\equiv\; \Phi(S_t), \qquad \Phi(S) \;=\; u\Big(\Lambda(E^S(S), E^P(S))\Big).$$

[1] This formalizes Horkheimer and Adorno (1944)'s concern that instrumental reason, when detached from reflection on ends, can be mobilized toward oppressive projects.

Contrast with preference-falsification accounts. A canonical mechanism for rapid norm change is preference falsification: individuals switch their *public* stance when the perceived social acceptability of a position shifts, generating threshold cascades (Kuran (1991), Kuran (1995), Bursztyn et al. (2020)). A reduced-form representation is

$$y_{i,t} = \mathbf{1}\{\, p_i + \mu_t - c_i(S_{t-1}) + \varepsilon_{i,t} \geq 0 \,\}, \qquad S_t = \mathbb{E}[y_{i,t} \mid \mathcal{F}_t],$$

where p_i is the private stance, $c_i(\cdot)$ decreases in perceived support, and μ_t is an exogenous acceptability shock (e.g., an election outcome or salient elite statement). In this class of models, sender identity and argument structure play no role: norm change operates through μ_t shifting expression thresholds. By contrast, my mechanism operates through persuasion: reach and institutional context determine exposure ($E^j(\cdot)$) and producer internalization (β_j); content precision and bundling (π_j^k, within-channel bundles, and cross-channel fit) govern how endorsement responds to discourse.

Assumption 4 (Exposure maps) *For $j \in \{S, P\}$, $E^j : [0,1] \to [0,\infty)$ is twice differentiable, increasing ($E^{j\,\prime} \geq 0$) with bounded derivatives; $E^{j\,\prime}$ may eventually decline (saturation).*

Lemma 3 (Curvature decomposition) *For $\Phi(S) = u(\Lambda(S))$,*

$$\Phi'(S) = u'(\Lambda)\,\Lambda'(S), \qquad \Phi''(S) = u''(\Lambda)\,[\Lambda'(S)]^2 \;+\; u'(\Lambda)\,\Lambda''(S).$$

Moreover, in the specialization extension, $\Lambda''(S)$ contains the strictly positive cross-fit term

$$2\,E^{S\,\prime}(S)\,E^{P\,\prime}(S)\left[\pi_S^m\pi_P^h\,e^{-\{Q_S^m(S)+Q_P^h(S)\}} \;+\; \pi_S^h\pi_P^m\,e^{-\{Q_S^h(S)+Q_P^m(S)\}}\right],$$

which is absent in the benchmark (where $\Lambda_\times \equiv 0$).

Proof 5 (Sketch) *The first display is the chain rule. For the second, note $f'(x) = e^{-x} > 0$ and write $\Lambda_\times(S) = f(Q_S^m(S))f(Q_P^h(S)) + f(Q_S^h(S))f(Q_P^m(S))$. Differentiating twice yields a term $2\,f'(Q_S^\ell)\,f'(Q_P^{\ell'})\,Q_S^{\ell\,\prime}(S)\,Q_P^{\ell'\,\prime}(S)$ for each off-diagonal pair $(\ell,\ell') \in \{(m,h),(h,m)\}$, with $Q_j^{k\,\prime}(S) = \pi_j^k E^{j\,\prime}(S)$, giving the stated expression.*

Proposition 4 (When the endorsement update map is S-shaped) *Under Assumption 4 and $u' > 0$, $u'' \leq 0$, a sufficient condition for local convexity at some interior $S^\star \in (0,1)$ is*

$$u'(\Lambda^\star)\,\Lambda''(S^\star) \;>\; -u''(\Lambda^\star)\,[\Lambda'(S^\star)]^2, \qquad \Lambda^\star \equiv \Lambda(S^\star).$$

If, in addition, $E^{j\,\prime}(S)$ eventually decreases and $u'(\Lambda(S)) \to 0$ as $S \to 1$, then $\Phi''(S) < 0$ for S sufficiently close to 1. Hence, for sufficiently strong cross-channel fit (e.g., large off-diagonal products $\pi_S^m\pi_P^h + \pi_S^h\pi_P^m$), the map Φ is S-shaped.

Proof 6 (Sketch) *By Lemma 3, $\Phi'' = u''(\Lambda)(\Lambda')^2 + u'(\Lambda)\Lambda''$. Cross-channel fit raises Λ'' through the strictly positive term in Lemma 3, making the inequality hold at intermediate S when the off-diagonal products are large enough. Near $S = 1$, saturation ($E^{j\,\prime}\downarrow$) and $u'(\Lambda)\downarrow$ imply $\Phi'' < 0$.*

Example (linear reach, exponential link). Let $E^j(S) = a_j + b_j S$ with $a_j, b_j > 0$ and $u(\Lambda) = 1 - e^{-\Lambda}$. Under strong specialization $(\pi_S^m, \pi_P^h) = (0,0)$,

$$\Lambda(S) = \big(1 - e^{-\pi_S^h E^S(S)}\big)\big(1 - e^{-\pi_P^m E^P(S)}\big).$$

Differentiating shows $\Lambda''(S)$ contains the strictly positive term $2\,\pi_S^h\pi_P^m\,E^{S\,\prime}(S)E^{P\,\prime}(S)\,e^{-(\pi_S^h E^S + \pi_P^m E^P)}$, which can dominate at intermediate S, while saturation implies concavity for large S. Hence $\Phi(S) = 1 - e^{-\Lambda(S)}$ can be S-shaped.

Corollary 2 (Multiplicity and regime switches) *Let* $S_{t+1} = \Phi(S_t)$ *with* Φ *increasing and S-shaped on* $[0, 1]$. *If* $\max_{S\in[0,1]} \Phi'(S) > 1$, *then there exist three fixed points* $0 < S_L < S_M < S_H \leq 1$ *with* S_L, S_H *locally stable and* S_M *unstable. Small shocks to reach* $E^j(\cdot)$ *or to cross-channel fit (through the off-diagonal products* $\pi_S^m \pi_P^h + \pi_S^h \pi_P^m$*) near* S_M *can trigger discontinuous transitions between low- and high-endorsement regimes.*

Interpretation (double-edged specialization). Stronger specialization increases off-diagonal products such as $\pi_S^h \pi_P^m$, raising the cross-fit contribution to $\Lambda''(S)$ (Lemma 3) and steepening the update map around intermediate endorsement levels (efficiency at low exposure). But it also raises $\max_S \Phi'(S)$, making the system more sensitive to small shocks and more prone to sharp regime switches (Corollary 2).

Illustrations. The illustrations below provide stylized mappings from the model to historical episodes.

Illustration 1: Interwar Germany as a low-reach, specialization–driven regime switch. Interwar Germany can be interpreted as an environment in which baseline receptiveness to particularist narratives was nontrivial – placing the system closer to an intermediate threshold than to a low-endorsement corner – consistent with the long-run intellectual currents emphasized by Kohn (1950) and Mosse (1964)[2]. In the model, a bureaucracy with a strong legal-rational tradition (Arendt (1963), Heldring (2023)) spills over into the public sphere and is captured by comparatively high implementation precision in the state-agent channel (high π_S^h). This increases the scope for cross-channel fit when paired with moral content supplied in the peer channel. Programmatic materials and administrative planning can be viewed as raising the availability of implementation talk alongside moral narratives, further strengthening cross-fit. Baseline reach can be thought of as relatively low in the sense of weak public-sphere integration. Despite dense associational life, civic engagement often occurred within segmented subpublics rather than cross-cutting fora, and party politics remained fragmented, limiting common arenas in which competing groups encountered shared reasons (Berman (1997)). In the model, this places the system in a low-exposure region in which complete bundles are rare and the marginal effects of increases in reach are large when cross-channel fit is active (Lemma 2). Mass persuasion technologies in the interwar period (e.g., radio propaganda; Adena et al. (2015)) plausibly raised effective reach in both channels, increasing $E^S(S)$ and $E^P(S)$. With cross-channel fit active, these forces can steepen the endorsement update map around intermediate endorsement levels (Proposition 4), so that a sufficiently large shock to reach tips the system from the vicinity of the unstable threshold S_M into the high-endorsement regime S_H. Thus, the surge reflects the interaction of preexisting particularism, strong implementation specialization of state agents, a weakly integrated public sphere, and a late increase in reach.

Illustration 2: The French Revolution as broad reach with within-channel bundling and an internalization shock. A natural reading of late–1780s France is that, prior to the political rupture itself, peer discourse already bundled the *why* and the *how* – normative claims accompanied by concrete institutional designs – consistent with an Enlightenment culture of "useful knowledge" (Mokyr (2010), Mokyr (2016)). In the model, this corresponds to substantial within-channel bundling

[2] Suggestive evidence of persistence in beliefs about cultural superiority comes from PewResearchCenter (2011), which reports that 47% of German respondents (and 49% of Americans) agree that "our culture is superior to others," compared with 44% in Spain, 32% in Britain, and 27% in France.

in the peer channel (high π_P^m and π_P^h), so that peer-delivered complete bundles

$$\Lambda_P \;=\; \big(1 - e^{-\pi_P^m E^P(S)}\big)\,\big(1 - e^{-\pi_P^h E^P(S)}\big)$$

can be generated without relying primarily on cross-channel pairing. Prominent contributions such as Condorcet (1785) and Sieyès (1789) illustrate this coupling of general-interest criteria with institutional proposals. Against this backdrop, the late-1780s pamphlet and newspaper boom, together with dense associational life (salons and *sociétés populaires*), plausibly raised peer reach (E^P) by expanding the scale and frequency of public argument exchange (Habermas (1989), Darnton (1982), Popkin (1990)). As the Estates–General convened and parliamentary procedure emerged through the National (then Constituent) Assembly, the public sphere became institutionally consequential. In the model, I capture this as an *internalization shock*: the perceived returns to producing, processing, and relaying public reasons increase when public endorsement becomes tied to formal political choice, so that $\beta_S \uparrow$ and $\beta_P \uparrow$, particularly for implementation talk[3]. Thus, the regime change is driven less by sharp cross-channel fit tied to strong specialization, and more by the combination of broad increases in reach and already-strong within-peer bundling, reinforced by higher incentives to supply precise public reasons once endorsement becomes consequential. This interpretation coheres with de Tocqueville (1998)'s emphasis on pre–Revolution administrative centralization meeting a public sphere empowered by the Estates–General, and with Habermas (1989)'s account of deliberation migrating from print and salons into parliamentary procedure.

2.2.3 Welfare and policy: amplification risk and guardrails

The positive analysis highlights a trade-off. Because endorsement is bundle-based – it requires a matched *why* (m) and *how* (h)– precision investments can improve the informational basis of legitimacy. But under strong role specialization, the state-agent channel supplies mainly implementation talk (h) while contributing little merit discourse (m). This reduces contestation on the m-dimension and can turn h into a content-agnostic amplifier: it scales up whichever merit discourse prevails rather than disciplining it. This section separates (i) a planner wedge that governs overall precision provision from (ii) guardrails that target the specific configuration generating amplification risk.

Planner benchmark (internalization wedge). Fix (c, t) and suppress indices. Channel $j \in \{S, P\}$ chooses effort (e_j^m, e_j^h), generating precisions $\pi_j^k = \upsilon_j^k(e_j^k)$, and endorsement depends on bundle precision Λ through $s = u(\Lambda)$, with $u' > 0$ and $u'' \le 0$. In the decentralized problem, each channel internalizes only a fraction $\beta_j \in (0, 1]$ of the legitimacy benefit, so its objective is

$$\max_{e_j^m, e_j^h \ge 0} \; \beta_j \, B \, u(\Lambda) \;-\; \sum_{k \in \{m,h\}} C_j^k(e_j^k).$$

A utilitarian planner instead internalizes the full benefit and chooses $\{e_j^m, e_j^h\}_{j \in \{S,P\}}$:

$$\max_{\{e_j^m, e_j^h\}_j} \; B \, u(\Lambda) \;-\; \sum_{j \in \{S,P\}} \sum_{k \in \{m,h\}} C_j^k(e_j^k).$$

When $\beta_j < 1$, decentralized channels under-provide precision relative to the planner, motivating policies that raise β_j or subsidize precision investments.

[3]The *cahiers de doléances* ("books of grievances") were lists of complaints and reform proposals compiled across France in early 1789 by local assemblies of the three estates to instruct their deputies to the Estates–General.

Amplification risk (composition wedge under specialization). The model isolates a composition risk. Under comparative advantage (Assumption 1), the state-agent channel is pushed toward higher implementation precision h relative to peers (Proposition 3). When this tilt becomes extreme ($\pi_S^h \gg \pi_S^m \simeq 0$), cross-channel fit creates *local* complementarity in reach: under strong specialization the interaction in exposures is positive at low reach (Corollary 1). Because h is content-agnostic – it raises endorsement for whichever merit discourse prevails – this configuration is socially costly precisely when the merit case is weakly disciplined. In that case, additional implementation talk scales up endorsement without disciplining the underlying m-content.

Policy levers as guardrails. The amplification risk in the model arises when three conditions coincide: (i) strong specialization on implementation in the state-agent channel ($\pi_S^h \gg \pi_S^m \simeq 0$); (ii) low or fragmented reach (small effective E^j and weak common fora); and (iii) weakly disciplined merit discourse (low-quality or weakly contested m). The following levers map directly into these primitives. For clarity, I separate *planner-wedge* policies (raising internalization / relaxing costs) from *composition and public-sphere* guardrails that target (i)–(iii).

1. **Composition guardrail: mix some m into the state-agent channel (mitigate (i)).** Policies that raise π_S^m and/or reduce an excessive tilt toward π_S^h move the state-agent channel away from the instrumental-amplifier configuration $\pi_S^h \gg \pi_S^m \simeq 0$. In the model, this increases the availability of within-channel bundles in S and reduces reliance on cross-channel pairing (Corollary 1), thereby lowering the scope for content-agnostic amplification when peer merit discourse is weakly disciplined. Institutionally, this corresponds to embedding substantive professional norms (e.g. universalistic service standards, rights-compatibility, fairness constraints) alongside legal-rational expertise (mission-oriented bureaucracy).

2. **Merit-discipline guardrail: raise media quality (mitigate (iii), and partly (ii)).** Higher media quality can be represented as higher precision for a given reach (higher π_P^m, and potentially π_P^h) and/or a higher marginal cost of low-quality claims (steeper C_P^m and C_P^h for unverified content). These changes increase contestability and reduce the probability that m is one-sided, making implementation talk h more likely to complete a disciplined merit case.

3. **Merit-discipline guardrail: invest in civic deliberative capacity (mitigate (iii)).** Civic-deliberation investments (education, civic training, debate and discussion infrastructures) can be modeled as lowering the cost of producing and processing public reasons in the peer channel (downward shifts in C_P^m and C_P^h) and/or improving precision technologies $\upsilon_P^m(\cdot)$ and $\upsilon_P^h(\cdot)$. This increases the availability of contestable reasons and, in turn, the likelihood that persuasive bundles are merit-disciplined rather than one-sided.

4. **Public-sphere integration guardrail: strengthen common fora (mitigate (ii)).** Policies that increase cross-cutting exposure and shared reference points – e.g. institutions or design choices that foster audience integration, common informational baselines, and deliberative fora – can be represented as raising baseline reach and reducing fragmentation in the exposure maps $E^j(\cdot)$. In the model, this shifts the environment away from the low-reach region in which complete bundles are rare and marginal effects of reach are large under cross-channel fit (Lemma 2). Conceptually, this lever targets the "weakly integrated public sphere" condition directly.

5. **Epistemic-infrastructure guardrail: stable funding for verification and accountability (mitigate (iii)).** Stable financing of verification and accountability capacity – independent journalism, fact-checking institutions, public archives – can be modeled as relaxing resource

constraints behind peer-side costs (lower C_P^m and C_P^h) and/or improving peer-side precision technologies (higher υ_P^m and υ_P^h). Such funding therefore limits capture-driven skew.

6. **Planner-wedge lever: increase internalization (raise β_j).** A utilitarian planner internalizes the full legitimacy benefit B, whereas decentralized channels internalize only $\beta_j B$. Institutions that make endorsement politically consequential (e.g. direct-democratic procedures) raise β_j, increasing equilibrium precision provision. This lever corrects underprovision of precision *in aggregate*; it does not by itself eliminate amplification risk, because comparative advantage can still channel additional provision toward h on the state-agent side. Hence it is most effective when combined with levers (1)–(5), which discipline merit discourse and limit extreme h-only specialization.

7. **Planner-wedge lever: civic dividend / time endowment (cost relief, and possibly reach).** An unconditional per-capita transfer T relaxes time and attention constraints and lowers the opportunity cost of producing and processing public reasons. In the model, this is naturally captured as cost relief on both content dimensions and in both channels:

$$C_j^k(e;T) \quad \text{with} \quad \frac{\partial}{\partial T} C_j^{k\prime}(e;T) < 0, \qquad j \in \{S,P\}, \quad k \in \{m,h\}.$$

   This increases equilibrium precisions π_j^m and π_j^h. Optionally, it may also raise participation and exposure (e.g. $E^j(S;T)$ with $\partial E^j/\partial T > 0$), but its clean planner interpretation is that it relaxes the precision-underprovision wedge. By itself, however, it does not remove amplification risk: if m remains weakly contested and the state channel is highly h-specialized, additional feasibility talk can still scale up a one-sided merit case.

Takeaway. Amplification risk is highest when (i) the state-agent channel specializes heavily in h, (ii) reach is low or fragmented, and (iii) merit discourse is weakly disciplined, so that cross-channel fit (Corollary 1) makes h a content-agnostic amplifier. Guardrails that discipline merit and strengthen deliberative capacity and common fora (levers 2–5), together with limits on extreme h-only specialization (lever 1), reduce this risk. Planner-type levers that raise internalization or relax costs (levers 6–7) then increase the supply of precision and legitimacy *without* mechanically increasing the likelihood of harmful regime switches.

For reference, Table 1 collects the main notation used throughout.

3 Conclusion

Beyond its positive and normative results, the paper contributes a language for connecting classic debates about Enlightenment rationality to institutional change. In the model, endorsement is bundle-based by construction, so the central question is whether feasibility completes a contested merit case or converts a one-sided one. This captures, in a single framework, both the democratic promise of a public sphere in which reasons are contestable and the risk that instrumental implementation capacity can be mobilized toward domination when detached from contestation.

Several extensions are natural. Endogenizing reach, allowing strategic manipulation of content quality, and introducing network structure and heterogeneous attention would sharpen the conditions under which low-reach environments generate amplification risk. Introducing multiple competing norms would allow the framework to speak to selection among institutional pathways. Finally, while the paper is theory-only, the model delivers clear comparative statics that can guide empirical work: the sign of exposure interactions should depend on whether bundles are assembled within or across channels; complementarity should be most pronounced at low reach and under stronger specialization; and institutional reforms that increase internalization should raise precision provision but require guardrails to avoid scaling up one-sided merit discourse.

The overarching message is that institutional emergence is not only a matter of preferences or coordination, but also of *public reasons* and the *channels* through which those reasons are assembled into persuasive bundles. Understanding when feasibility talk disciplines legitimacy formation versus amplifies prevailing narratives is essential for explaining institutional change and for designing reforms that preserve informational gains while limiting amplification risk.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya (2015) "Radio and the Rise of The Nazis in Prewar Germany *," *The Quarterly Journal of Economics*, 130 (4), 1885–1939, 10.1093/qje/qjv030.

Arendt, Hannah (1963) *Eichmann in Jerusalem: A Report on the Banality of Evil*, New York: The Viking Press.

Berman, Sheri (1997) "Civil Society and the Collapse of the Weimar Republic," *World Politics*, 49 (3), 401–429, 10.1353/wp.1997.0008.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin (2020) "From Extreme to Mainstream: The Erosion of Social Norms," *American Economic Review*, 110 (11), 3522–48, 10.1257/aer.20171175.

Castoriadis, Cornelius (1987) *The Imaginary Institution of Society*, Cambridge, Massachusetts: The MIT Press.

Condorcet, Marie-Jean-Antoine-Nicolas de Caritat (1785) *Essay on the Application of Analysis to the Probability of Majority Decisions (Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix)*: Paris.

Darnton, Robert (1982) *The Literary Underground of the Old Regime*: Harvard University Press.

Habermas, Jürgen (1989) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*: MIT Press.

Heldring, Leander (2023) "Bureaucracy as a Tool for Politicians: Evidence from Germany," working paper.

Horkheimer, Max and Theodor W. Adorno (1944) *Dialektik der Aufklärung: Philosophische Fragmente*, Amsterdam: Querido Verlag.

Kohn, Hans (1950) "Romanticism and the Rise of German Nationalism," *The Review of Politics*, 12 (4), 443–472.

Kuran, Timur (1991) "The East European Revolution of 1989: Is it Surprising that We Were Surprised?" *The American Economic Review*, 81 (2), 121–125.

——— (1995) *Private Truths, Public Lies*: Harvard University Press.

Mokyr, Joel (2010) *The Enlightened Economy*: Yale University Press.

——— (2016) *A Culture of Growth*: Princeton University Press.

Mosse, George L. (1964) *The Crisis of German Ideology: Intellectual Origins of the Third Reich*, New York: Grosset and Dunlap.

PewResearchCenter (2011) "The American-Western European Values Gap,"Technical report, Pew Research Center's Global Attitudes Project, Retrieved from http://pewrsr.ch/vIHVsk.

Popkin, Jeremy D. (1990) *News and Politics in the Age of Revolution: Jean Louis Carrier and the Révolutions de Paris, 1789–1797*: Cornell University Press.

Sieyès, Emmanuel-Joseph (1789) *Qu'est-ce que le Tiers-État?*: Paris.

de Tocqueville, Alexis (1998) *The Old Regime and the Revolution*: University of Chicago Press.

Weber, Max (1922) *Economy and Society*, 4th ed. Tübingen.

Table 1: Notation

Object	**Definition / interpretation**
Indices and sets	
i	Citizen (agent).
(c, t)	Public-sphere context c at date t.
$j \in \{S, P\}$	Communication channel: S = state-agent channel; P = peer channel.
$k \in \{m, h\}$	Content dimension: m = merit/"why"; h = implementation/"how".
Latent states, priors, and signals	
m, h	Latent "merit" and "implementation" states relevant for persuasion.
μ_k, σ_k^2	Prior mean and variance for dimension $k \in \{m, h\}$.
$\tau_k \equiv \sigma_k^{-2}$	Prior precision for dimension k.
$\tilde{x}_{j,i,c,t}^k$	Signal received by i from channel j about dimension k in (c, t).
$\mathcal{F}_{i,c,t}$	Citizen i's information set in (c, t): $\mathcal{F}_{i,c,t} = \sigma\big(E_{i,c,t}^S, E_{i,c,t}^P, \tilde{x}_{S,i,c,t}^m, \tilde{x}_{S,i,c,t}^h, \tilde{x}_{P,i,c,t}^m, \tilde{x}_{P,i,c,t}^h\big)$.
$\sigma(\cdot)$	Sigma-algebra generated by the listed random variables (information available to the agent).
Exposure / reach	
$E_{i,c,t}^j$	Realized exposure of citizen i to channel j in (c, t) (nonnegative).
$E_{c,t}^j$	Aggregate reach (average exposure intensity) of channel j in public sphere (c, t).
$\eta_{i,c,t}^j$	Idiosyncratic reach component: $E_{i,c,t}^j = E_{c,t}^j + \eta_{i,c,t}^j$, $\mathbb{E}[\eta_{i,c,t}^j \mid c, t] = 0$.
Precision and effective precision	
e_j^k	Producer effort by channel j to generate precision on dimension k.
$\upsilon_j^k(\cdot)$	Precision technology: $\pi_j^k = \upsilon_j^k(e_j^k)$ (increasing, concave).
π_j^k	Precision supplied by channel j on dimension k.
Q_j^k	Effective precision / intensity from channel j on dimension k: $Q_j^k \equiv \pi_j^k E^j$ (suppressing indices when convenient).
$Q_{k,i,c,t}$	Total effective precision for citizen i on dimension k in (c, t): $Q_{k,i,c,t} \equiv \sum_{j\in\{S,P\}} \pi_j^k E_{i,c,t}^j$.
Bundling and endorsement	
$f(x)$	Poisson "arrival" map: $f(x) \equiv 1 - e^{-x}$.
Λ_j	Within-channel bundle intensity: $\Lambda_j \equiv f(Q_j^m)\, f(Q_j^h)$.
$\Lambda_\times$	Cross-channel fit (extension): $\Lambda_\times \equiv f(Q_S^m) f(Q_P^h) + f(Q_S^h) f(Q_P^m)$.
Λ	Total bundle intensity: $\Lambda \equiv \Lambda_S + \Lambda_P$ (benchmark) or $\Lambda \equiv \Lambda_S + \Lambda_P + \Lambda_\times$ (extension).
s	Endorsement probability / share: $s = u(\Lambda)$.
$u(\cdot)$	Endorsement link, increasing and concave ($u' > 0$, $u'' \le 0$); canonical case $u(\Lambda) = 1 - e^{-\Lambda}$.
κ	Off-diagonal "fit" index (local complementarity at the origin): $\kappa \equiv \pi_S^m \pi_P^h + \pi_S^h \pi_P^m$.
Producer payoffs and costs	
$C_j^k(\cdot)$	Cost of effort e_j^k for channel j on dimension k (convex).
B	Social benefit scale from endorsement/legitimacy ($B > 0$).
β_j	Internalization weight: channel j internalizes fraction $\beta_j \in (0, 1]$ of B.
δ	Implementation comparative-advantage wedge (Assumption 1).
Dynamics (when used)	
S_t	Aggregate endorsement at date t (e.g., population average of s).
$E^j(S)$	Exposure map: reach as a function of current endorsement S (public-sphere propagation).
$\Phi(S)$	Update map: $\Phi(S) \equiv u(\Lambda(S))$ so $S_{t+1} = \Phi(S_t)$.