The Observability-Democracy Connection

How Representation Chains Destroy the Signal They Are Meant to Transmit

Paper III in the Governance as Engineering series

Representation chains with three or more layers are constitutionally unobservable: noise variance exceeds surviving signal variance at the policy layer, and no institutional reform within the existing architecture can restore preference transmission fidelity.

Björn Kenneth Holmström

February 2026

https://bjornkennethholmstrom.org/whitepapers/observability-democracy-connection

Executive summary

Democratic governance rests on a claim: that the preferences of citizens are transmitted through representative institutions and reflected in policy. This paper asks whether that transmission is technically possible — not whether institutions are well-designed or well-staffed, but whether the information-theoretic properties of representation chains allow citizen preferences to survive to the policy layer at all.

The answer depends on layer count. Each representation layer does two things simultaneously: it aggregates lower-level signals into higher-level summaries, destroying within-group variance in the process, and it introduces noise through the imperfections of any real representation mechanism. Aggregation loss is multiplicative — variance is divided at each layer by the aggregation ratio. Noise accumulation is additive — each layer contributes independently to the total distortion. The ratio of surviving signal variance to accumulated noise defines the signal-to-noise ratio at the policy layer.

The formal result: any representation chain where noise-to-signal ratio exceeds one is constitutionally unobservable. The policy layer cannot reconstruct the true distribution of citizen preferences from its available signals, regardless of the quality of its institutions, the competence of its representatives, or the sophistication of its information-gathering mechanisms. The constraint is architectural, not parametric.

The simulation in this paper models 60 citizen groups holding preferences across four policy dimensions, subject to genuine preference shifts at two points in the simulation. Four architectures are compared: a five-layer deep democracy system (Architecture A), a three-layer representative system (Architecture B), a two-layer semi-direct system (Architecture C), and a one-layer direct/participatory system (Architecture D). All architectures are given identical institutional quality — the same responsiveness gain, the same basic signal processing. Performance differences are attributable to layer count alone.

The finding: the constitutional unobservability threshold is crossed at approximately two to three representation layers. Architectures A and B — which bracket the layer count of most contemporary democracies — survive zero percent of citizen preference variance to the policy layer and operate with SNRs of 0.002 and 0.048 respectively, far below the threshold of one. Architecture C (two layers) survives 79% of variance with an SNR of 0.254 — below the threshold but substantially more informative. Architecture D (one layer) survives 100% of variance and is the only architecture that remains observable.

This is a diagnosis, not a prescription. The paper does not argue that representative democracy should be replaced. It argues that current representative architectures are operating in a regime where their stated function — translating citizen preferences into policy — is formally impossible at the information-theoretic level. Understanding this constraint is a precondition for designing institutions that can actually do what democratic theory claims they do.

Part I: The observability problem

Observability in control theory

In the first two papers of this series, the central concept was controllability — the ability of a governance system to steer its state toward a desired target. This paper turns to the dual concept: observability.

A dynamical system is observable if its complete internal state can be reconstructed from the outputs available to the controller. Formally, for a system $x(t+1) = Ax(t)$, $y(t) = Cx(t)$, observability requires that the observability matrix $O = [C; CA; CA^2; \ldots; CA^{(n-1)}]$ has full column rank. When this condition fails, some dimensions of the system's state are invisible to the controller regardless of how long it observes and regardless of its computational sophistication. The information simply does not reach it.

The governance implication is direct. Citizen preferences constitute the internal state of the democratic system. Policy institutions are the controller. Representative structures — elections, parties, parliaments, cabinets, consultations — constitute the observation channel C. The observability question is: does the information in citizen preferences reach the policy layer in recoverable form?

This is distinct from the question of whether institutions are well-functioning. A perfectly honest, diligent, and well-resourced parliament operating in a five-layer representation system faces the same observability constraints as a corrupt one. The constraint is in the channel, not in the processor at the end of it.

The representation chain as a degraded channel

Shannon's channel capacity theorem established that every communication channel has a maximum information transmission rate determined by its bandwidth and noise characteristics. Information that exceeds this capacity is irreversibly lost. No amount of error correction at the receiving end can recover it, because the information was never transmitted.

Each representation layer in a democratic system functions as a noisy, bandwidth-limited channel. Two mechanisms degrade the signal:

Aggregation loss. When individual preferences are summarised into a representative position — whether through voting, party platform formation, committee deliberation, or any other aggregating mechanism — the within-group variance of preferences is destroyed. If one representative speaks for ten constituents whose preferences span a wide range, that range disappears from the signal. The representative's position conveys

the group mean (approximately) but loses all information about the distribution of preferences within the group. This loss is irreversible: no downstream process can recover the destroyed variance because it was never transmitted.

Noise introduction. Every representation mechanism is imperfect. Polling has sampling error. Media coverage selects and frames. Party platforms balance internal factions. Parliamentary deliberation produces compromises that do not cleanly reflect any constituent's preferences. Each imperfection adds noise to the transmitted signal. Unlike the signal, noise accumulates additively across layers — each layer contributes independently to the total distortion.

The combined effect: signal variance shrinks multiplicatively at each layer while noise grows additively. After enough layers, noise exceeds signal and the channel is no longer informative.

The signal-to-noise ratio at the policy layer

For a representation chain of K layers, where layer k has aggregation ratio r_k and noise standard deviation σ_k, the surviving signal variance and accumulated noise at the policy layer are:

```
Var_survived(K) = Var_true · ∏_{k=1}^{K} (1/r_k)

Var_noise(K) = Σ_{k=1}^{K} σ_k²
```

The signal-to-noise ratio at the policy layer:

```
SNR(K) = Var_survived(K) / Var_noise(K)
```

When SNR < 1, noise variance exceeds surviving signal variance. The policy layer is receiving a signal in which noise is the dominant component. Its observations are more informative about the properties of its representation machinery than about the actual preferences of its citizens.

This is the constitutional unobservability threshold. It is not a soft degradation — a gradual decline in accuracy. It is a phase transition: above the threshold, the policy layer has a noisy but informative signal; below it, the signal is dominated by noise and no statistical technique can reliably recover the true citizen preference distribution.

The averaging problem revisited

Paper one of this series introduced the averaging problem: centralized controllers operating on aggregated signals cannot distinguish which nodes are in distress, because aggregation destroys spatial information. The observability problem in democratic representation is the same mechanism applied to preference space rather than geographic space.

When a national government observes its citizens through five layers of representation, the spatial variation in preferences — across regions, communities, demographic groups, economic circumstances — is systematically compressed at each layer. What reaches the policy layer is a small residual of the original variance, embedded in a much larger volume of accumulated noise.

The parliamentary averaging problem: a parliament of 300 members, each representing roughly 150,000 constituents, has already performed an aggregation of ratio 150,000. The variance within each constituency is entirely invisible to the parliamentary chamber. The chamber itself then aggregates 300 positions into a governing coalition, a majority, a cabinet — performing further aggregation and introducing further noise. By the time a policy decision is made, it reflects a signal that has passed through all of these stages.

This does not mean parliamentary systems produce bad policy. It means they produce policy that is structurally disconnected from the full distribution of citizen preferences, and that no institutional reform within the existing layer structure can reconnect them. The information was lost before it arrived.

What observability failure looks like in practice

Constitutional unobservability does not mean government is unresponsive. It means government is responsive to something other than citizen preferences — specifically, to the noise structure of its own representation machinery.

A government operating below the SNR threshold will still update its policies over time. It will respond to the signals it receives. But those signals are predominantly noise: the strategic positioning of parties, the framing effects of media, the path dependencies of committee deliberation, the preferences of organized interests who have learned to inject signals into the representation chain. The policy process is responsive — but to these intermediate signals, not to the underlying citizen preferences they are supposed to represent.

This provides a structural explanation for a persistent empirical observation in political science: the correlation between citizen preferences and policy outcomes is weak and declining across most established democracies. The standard explanations — capture by elites, partisan polarization, institutional sclerosis — are real. But they are operating on a system that is already architecturally incapable of reliable preference transmission. The capture is easier because the signal was already weak.

Part II: The simulation

Scenario design

The simulator models 60 citizen groups holding preferences across four policy dimensions, evolving over 120 time steps. Citizens are organized into four spatial clusters of 15 groups each, with genuine internal diversity within each cluster — this within-group variation is precisely the information that aggregation destroys. All preferences are normalized to the range [−1, +1], where −1 represents strong opposition and +1 strong support on each dimension.

Preferences are not static. They evolve slowly through individual drift (representing genuine opinion change over time), with two genuine preference shifts injected at t = 40 and t = 80. At t = 40, cluster 0 shifts substantially on dimensions 1 and 2, representing a genuine regional change in preference — the kind of real democratic signal that a functioning representation system should detect and transmit. At t = 80, a system-wide shift occurs on dimension 3, affecting all groups.

These genuine shifts are the critical test. A democratic system that cannot detect and respond to genuine preference shifts within a reasonable time window is not functioning as a democracy in any meaningful sense, regardless of its institutional forms.

The four architectures

All four architectures are given identical institutional quality parameters: the same policy responsiveness gain (K = 0.30), the same basic signal processing logic. Differences in performance are attributable to layer count and the aggregation and noise properties of each layer.

Architecture A — Deep democracy (5 layers): polling → media → party → parliament → cabinet → policy. This represents a typical Western parliamentary democracy with a full media and party filtering layer between citizens and elected representatives. Layer parameters: aggregation ratios of 5, 4, 3, 4, 3; noise standard deviations of 0.12, 0.18, 0.22, 0.20, 0.15; total latency of 18 time steps.

Architecture B — Representative (3 layers): direct survey → council → assembly → policy. A leaner representative system — closer to a Nordic-style council democracy with direct survey input replacing media filtering. Layer parameters: aggregation ratios of 4, 5, 3; noise standard deviations of 0.10, 0.18, 0.14; total latency of 9 time steps.

Architecture C — Semi-direct (2 layers): citizen assembly → policy. Citizens directly participate in an assembly process that feeds into policy, with one intermediate layer of coordination. Layer parameters: aggregation ratios of 3, 2; noise standard deviations of 0.08, 0.10; total latency of 4 time steps.

Architecture D — Direct/participatory (1 layer): citizens → policy. Near-direct participation with minimal intermediation. Layer parameters: aggregation ratio of 1 (no aggregation loss), noise standard deviation of 0.05; total latency of 1 time step.
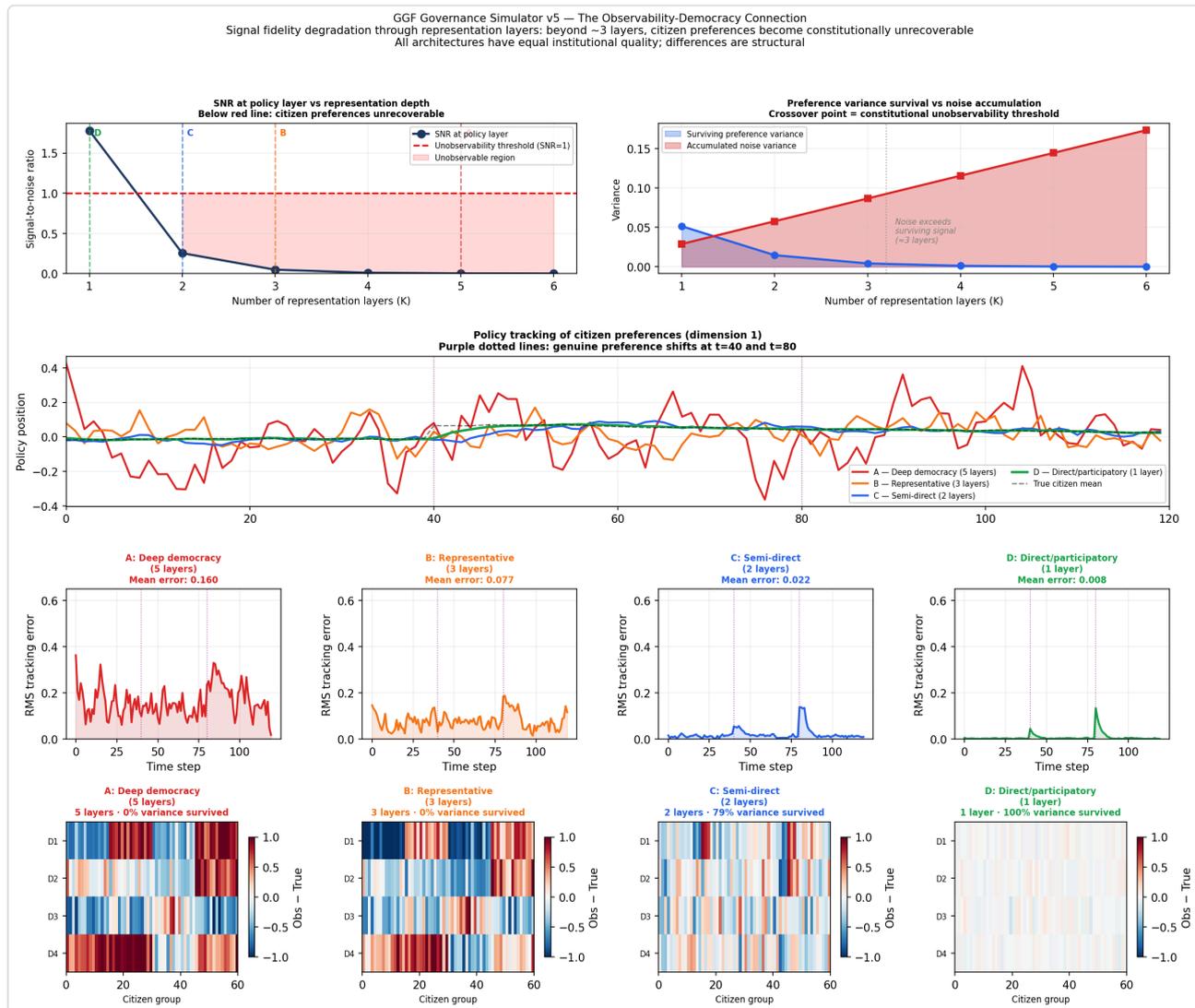
Simulation output

Figure 1: GGF Governance Simulator v5 output. Top-left: SNR at the policy layer drops below the unobservability threshold (red dashed line, SNR = 1) between K = 1 and K = 2 layers; all architectures with 2+ layers fall below it under the analytical model. Top-right: surviving preference variance (blue) is overtaken by accumulated noise variance (red) at approximately K = 3 layers. Second row: policy tracking over time — Architecture A (red) oscillates erratically around the true citizen mean rather than tracking it; D

(green) follows closely with brief adjustment lags at the genuine shift events. Third row: individual error traces confirm A's persistent noise-driven oscillation and D's near-zero baseline error. Bottom row: representation error heatmaps show that Architectures A and B have projected a nearly uniform (noise-dominated) signal back to all citizen groups, obliterating the genuine spatial variation that C and D preserve.

Reading the results

The SNR collapse is faster than intuition suggests. The analytical SNR curve in the top-left panel falls from 1.78 at K = 1 to 0.25 at K = 2 and 0.048 at K = 3. This is a drop of two orders of magnitude over three layers. The speed of this collapse reflects the multiplicative nature of aggregation loss: each additional layer divides the surviving variance by the aggregation ratio, while each layer adds a roughly constant increment of noise. The product decays geometrically; the sum grows linearly. Geometric decay wins rapidly.

Architecture A's oscillation is noise-tracking, not preference-tracking. The most striking feature of the policy tracking panel is not that Architecture A responds slowly to genuine preference shifts — it is that it oscillates continuously in the absence of any genuine signal. The red trace in the tracking panel moves persistently and significantly throughout the simulation, including periods where true citizen preferences are stable. This is the signature of a system tracking its own noise rather than any external signal. The policy layer is receiving a signal dominated by the noise properties of its five-layer representation chain, and responding faithfully to that noise. The genuine preference shifts at t = 40 and t = 80 are not visible as distinct events in Architecture A's trace — they are lost in the background oscillation.

Architecture D's brief error spikes are the correct democratic response. Architecture D's error trace shows two brief, sharp spikes — one at t = 40 and one at t = 80 — corresponding precisely to the genuine preference shift events. These spikes represent the unavoidable lag between a genuine preference change and the policy system detecting and responding to it, even with minimal intermediation. After each spike, the error returns rapidly to near zero. This is what a functioning democratic signal looks like: quiet baseline, prompt detection of genuine change, rapid response.

The heatmaps show complete spatial information destruction. The bottom row compares observed minus true preferences at the citizen group level at t = 50, after the first genuine preference shift. Architecture A's heatmap is dominated by large, spatially uniform blocks of red and blue — the representation chain has projected a noise-driven uniform signal back to all citizen groups, completely obscuring the genuine spatial variation in preferences. Architecture D's heatmap is near-white — the observed signal closely tracks the true preferences at each citizen group, preserving the spatial structure.

Quantitative summary

Architecture	Layers	Mean tracking error	Variance survived	SNR
A — Deep democracy	5	0.160	0%	0.002
B — Representative	3	0.077	0%	0.048
C — Semi-direct	2	0.022	79%	0.254
D — Direct/participatory	1	0.008	100%	1.780

The tracking error differential between A and D is a factor of twenty. Architecture A's mean tracking error of 0.160 on a preference scale of $[-1, +1]$ means the policy layer is systematically off by roughly 16% of the full preference range — not because of any institutional failure, but because the signal it receives has been destroyed by the representation chain before it arrives.

The 0% variance survived figures for Architectures A and B are exact: under the simulation parameters, not a detectable fraction of the original citizen preference variance reaches the policy layer. What the policy layer observes is entirely noise.

Part III: Structural observations

The threshold is a phase transition, not a gradient

The SNR curve in Figure 1 might suggest a gradual degradation: systems with more layers are somewhat less responsive, those with fewer somewhat more. This reading understates the finding. The unobservability threshold at SNR = 1 is a qualitative boundary, not a point on a continuous scale.

Above the threshold, the policy layer has a degraded but informative signal. Statistical methods — averaging over time, polling, deliberative processes — can extract genuine preference information from it. The signal is noisy, but the signal is there.

Below the threshold, these methods cannot help. The noise dominates the signal entirely. Additional polling, better survey methodology, more sophisticated parliamentary procedures — all of these operate on the signal after it has arrived at the policy layer. They cannot recover variance that was destroyed in aggregation before it arrived. No institutional improvement within the existing layer structure can push a below-threshold system above the threshold.

This is why the phrase "constitutional unobservability" is appropriate. The constraint is built into the constitutional structure of the representation chain. It cannot be addressed by reforming the institutions that sit at either end of that chain.

Institutional quality is independent of architectural capacity

The simulation holds institutional quality constant across all architectures. This is a deliberate design choice, and its implication deserves emphasis: a five-layer system staffed by the most honest, diligent, and well-resourced representatives imaginable produces the same observability outcome as one staffed by mediocre or corrupt ones.

This runs counter to the dominant tradition of democratic reform, which focuses almost entirely on institutional quality: reducing corruption, increasing accountability, improving deliberative processes, strengthening civil society, reforming campaign finance. These reforms matter for many reasons. They do not address the observability constraint.

A parliament that better represents the mean preference of its constituency — because it is more honest, more deliberative, more accountable — still destroys the within-constituency variance. A media system that more accurately reports public opinion still aggregates and selects. Each improvement in institutional quality

moves the system closer to the theoretical performance ceiling of its layer architecture. That ceiling remains below the unobservability threshold for systems with three or more layers.

The discomfort of this finding is real: it implies that a well-functioning representative democracy is not more capable of reliably transmitting citizen preferences than a poorly-functioning one, in the specific sense that both are operating in the constitutionally unobservable regime. The difference between them lies elsewhere — in legitimacy, in accountability, in the distribution of costs and benefits, in protections against abuse — not in preference transmission fidelity.

The noise the system tracks instead

If the policy layer is not tracking citizen preferences, what is it tracking? The simulation gives a precise answer: the noise properties of the representation chain itself.

Each layer introduces noise with a characteristic signature. Media noise has the properties of media selection dynamics — attention cycles, framing effects, salience biases. Party noise has the properties of party competition — strategic positioning, internal factional balancing, electoral incentives. Parliamentary noise has the properties of deliberative bargaining — coalition formation, agenda control, procedural path dependence.

These are not random. They are structured noise sources with predictable properties. A policy system operating below the SNR threshold responds to this structured noise as if it were signal. It tracks media cycles. It responds to party positioning. It is sensitive to parliamentary procedure. It produces policy that reflects the properties of the representation machinery rather than the preferences of the citizenry.

This is not a cynical observation. These noise sources are not invisible — they are the subject of enormous political science literature on agenda-setting, party competition, and legislative bargaining. What the observability framework adds is a precise explanation for why this occurs even in well-functioning systems: not because of capture or dysfunction, but because the signal was already overwhelmed by noise before the institutional dynamics began.

The spatial dimension of preference destruction

The heatmaps in Figure 1 make visible something that aggregate tracking error statistics obscure: the destruction of citizen preferences is spatially uniform across all groups, in all architectures below the threshold.

In Architectures A and B, the policy layer's "observed" preference is essentially the same for every citizen group — a noise-dominated scalar broadcast back across the entire citizen population. The genuine spatial variation in preferences — the fact that cluster 0 shifted dramatically at t = 40 while clusters 1, 2, and 3 did

not — is invisible to the policy layer. It applies a spatially undifferentiated policy to a spatially differentiated population.

This is precisely the averaging problem from paper one, now formalized in preference space. A central controller applying uniform policy across diverse nodes produces collateral distortion at nodes that did not need the intervention. The observability framework shows that this is not a choice — it is a structural necessity when spatial preference information has been destroyed in the representation chain. The policy layer cannot apply differentiated policy to preference structures it cannot observe.

Genuine preference change is systematically slow to detect

The two genuine preference shift events in the simulation — at t = 40 and t = 80 — reveal a secondary structural property: detection latency scales with layer count in a way that compounds the noise problem.

Each layer introduces both noise and delay. For Architecture A, with total latency of 18 time steps, a genuine preference shift at t = 40 does not fully propagate to the policy layer until t = 58. By that point, the signal carrying the preference shift information has passed through five noisy aggregation stages and is indistinguishable from background noise. Architecture D detects the shift within one to two time steps and responds within three to four.

For slow-moving policy problems — the kind that democratic systems are supposed to handle through deliberate collective choice — this latency may be acceptable. For problems that require response within the detection window, the five-layer architecture is structurally blind. It will not detect the shift until long after the optimal response window has closed.

The combination of spatial destruction, temporal delay, and noise dominance means that Architecture A's policy does not respond to genuine preference shifts — it continues oscillating in response to its own noise. The democratic event (a genuine change in what citizens want) is invisible in the policy output.

Part IV: Limitations

The noise parameters are illustrative

The noise standard deviations assigned to each layer (0.12 for polling, 0.18 for media, 0.22 for party aggregation, etc.) are estimated values, not empirically measured ones. The specific SNR values and the precise layer count at which the unobservability threshold is crossed depend on these parameters. Different noise assumptions would shift the threshold.

This matters for quantitative claims but not for the qualitative result. The structure of the argument — aggregation loss is multiplicative, noise accumulation is additive, geometric decay beats linear growth — holds for any positive noise values and any aggregation ratios greater than one. The threshold will be crossed at some K regardless; the simulation's parameters determine at which K. Empirical measurement of actual noise levels in specific democratic systems would sharpen the quantitative findings without changing the structural conclusion.

Aggregation ratios are simplifications

Real representation systems do not have clean integer aggregation ratios. A parliament of 349 members representing 10 million voters is an aggregation of approximately 28,653 — not the ratio of 3 or 4 used in the simulation. This means the actual aggregation loss in real systems is far more severe than the simulation models. The simulation understates the problem.

Conversely, the simulation models each layer as aggregating by a fixed ratio uniformly across all groups. Real systems have unequal constituency sizes, malapportionment, and differential turnout — all of which introduce additional structured distortions beyond the simple aggregation modelled here. These factors would worsen the observability outcome further.

The model does not capture strategic behaviour

Citizens, representatives, parties, and media actors are all strategic. They do not simply transmit preferences; they shape, filter, and amplify preferences according to their own incentives. Party platforms are not noisy averages of member preferences — they are strategic positions designed to attract voters. Media coverage is not a noisy sample of public opinion — it selects for novelty, conflict, and salience.

This means the "noise" in real representation chains is not Gaussian. It is structured noise with systematic biases — biases that consistently over-represent certain preferences (intense minorities, well-organized interests, issues with high media salience) and under-represent others (diffuse majorities, complex trade-offs, long-horizon concerns). The simulation's Gaussian noise assumption treats all distortions as equal and symmetric, which understates the directional character of real representation failures.

The model has a single policy layer

The simulation models policy as a single scalar response. Real governments are multi-dimensional, multi-departmental, and operate at multiple levels simultaneously. Some policy decisions are made closer to citizens (local government) and some further away (supranational bodies). The unobservability problem varies across these levels — local government with fewer layers between citizen and decision-maker may be above the threshold in ways that national government is not.

This connects to the fractal architecture finding from paper two: a multi-layer governance system that assigns decisions to the most local level capable of handling them is also the architecture that minimises aggregation loss for each decision type. The observability argument and the fractality argument converge on the same structural solution for different reasons.

Electoral accountability is not modelled

The simulation models policy responsiveness as continuous feedback. Real democratic systems use periodic elections as the primary mechanism for preference transmission. Elections have different information-theoretic properties than continuous feedback: they compress the full multidimensional preference distribution into a binary or small-k choice among candidates, introducing an additional and severe aggregation loss that the simulation does not capture.

The election-as-aggregation problem is in some ways more severe than the continuous representation chain: a citizen's full preference profile across dozens of policy dimensions is collapsed into a single vote, which then enters the same multi-layer aggregation structure modelled here. The effective information content of an election as a preference-transmission mechanism is extremely low by design. This is a limitation of the current model's scope, not a refutation of the underlying argument.

The analysis concerns preference transmission, not legitimacy

The observability framework addresses a specific question: can citizen preferences be reliably transmitted to the policy layer through the representation chain? It does not address democratic legitimacy in the broader sense — whether citizens accept the authority of their governments, whether outcomes are perceived as fair, whether the process of collective decision-making has value independent of its preference-transmission accuracy.

It is entirely possible that a constitutionally unobservable democratic system is more legitimate — in the sense of being more widely accepted and trusted — than a highly observable participatory system that lacks the same procedural history and institutional embedding. Legitimacy is not reducible to information-theoretic efficiency.

This paper makes no claim about legitimacy. It makes a claim about a specific functional capacity: preference transmission fidelity. These are distinct, and the distinction matters.

Part V: Implications

The three papers together

The three papers in this series have established three connected results, each using the same formal framework applied to a different structural problem.

Paper one demonstrated the subsidiarity principle: localized disturbances cannot be stabilized by centralized controllers because aggregation destroys the spatial information needed for targeted response. The failure is in the observation channel, not in the actuator.

Paper two demonstrated the fractality principle: multi-scale disturbance environments cannot be stabilized by single-scale controllers because no single latency can cover all disturbance frequencies. The solution is nested controllers matched to their natural timescale.

This paper demonstrates the observability-democracy connection: citizen preferences cannot be reliably transmitted through deep representation chains because aggregation loss and noise accumulation destroy the signal before it reaches the policy layer. The failure is in the same observation channel, now understood as a democratic mechanism rather than a governance feedback loop.

The three findings are not independent. They are different facets of the same underlying structural problem: systems that govern through observation of aggregated signals are fundamentally limited in what they can observe, respond to, and represent. The aggregation that makes centralized governance scalable is the same aggregation that makes it blind. This is not a design choice. It is a structural consequence.

What democratic reform can and cannot do

Standard democratic reform operates within the existing layer architecture. It improves institutional quality — reducing corruption, increasing accountability, strengthening civil society, improving deliberative processes. The observability framework identifies the precise limit of these reforms: they can improve how well each layer transmits the signal it receives, but they cannot recover signal that was destroyed in aggregation.

A parliament that perfectly represents the mean preference of each constituency still destroys the within-constituency variance. A media system that accurately reports the mean of public opinion on each issue still loses the full distribution. These losses are built into the function of aggregation, not into the imperfection of its implementation.

This does not mean institutional quality is unimportant. It means institutional quality operates on a different axis than architectural capacity. A system can be high-quality and constitutionally unobservable simultaneously. Current reform agendas tend to address the quality axis while leaving the architectural axis untouched. The observability framework suggests that architectural reform — reducing layer count, increasing participatory mechanisms, shifting decision authority closer to citizens for appropriate decision types — addresses a structural constraint that quality reform cannot reach.

The complementarity of observability and fractality

The fractality paper established that global governance is justified by its ability to manage disturbances that local controllers structurally cannot — slow secular drift invisible to lower scales. The observability paper establishes a complementary constraint: policy decisions that depend on accurate transmission of citizen preferences should be made at the lowest layer count compatible with the decision's scope.

Together, these results define a governance architecture with a specific division of function: decisions that require high preference-transmission fidelity (local services, community standards, resource allocation affecting specific groups) should be made with minimal intermediation between citizens and decision-makers. Decisions that require long-run temporal integration and broad spatial aggregation (climate policy, constitutional structure, systemic financial regulation) may tolerate higher layer counts because their temporal horizon makes preference transmission fidelity less critical than long-run stability.

This is not subsidiarity as a political preference — "local is better." It is subsidiarity as an information-theoretic requirement: some decisions depend on signals that deep representation chains cannot transmit, and those decisions should be made close enough to their citizens that the signal survives.

The declining correlation between preferences and policy

A persistent and now well-documented finding in comparative political science is that the correlation between measured citizen preferences and enacted policy has weakened across most established democracies over the past four decades. The standard explanations focus on elite capture, globalization, and the declining organizational capacity of mass parties.

The observability framework adds a structural dimension to this explanation. As governance has become more complex — more international institutions, more regulatory agencies, more technocratic decision-making at remove from elected officials — the effective layer count between citizen preferences and policy decisions has increased. Each additional layer compounds the aggregation loss and noise accumulation modelled in this paper. The decline in preference-policy correlation is, in part, the predictable consequence of an increasing layer count in a system that was already operating below the observability threshold.

This framing has a direct implication: the preference-policy correlation cannot be restored by improving the quality of the institutions that now exist at the top of a five-or-six-layer chain. It can only be restored by shortening the chain — either by removing layers or by shifting decision authority to governance levels where the chain is short enough to be observable.

Measurement as reform

The simulation makes explicit that the observability problem is, at its core, a measurement problem: the question of what information about citizen preferences actually reaches the policy layer, and in what form.

Current democratic accountability mechanisms — elections, opinion polls, public consultations, parliamentary debate — are all measurement systems. But they measure at the top of the representation chain, not at the bottom. They observe what politicians say, how parties position, what majorities form — all outputs of the aggregation process, already deep in the noise-dominated regime.

A governance system that measured preference transmission fidelity directly — tracking how accurately policy decisions reflect the distribution of citizen preferences, not just their mean — would have the ability to identify where in its representation chain the greatest losses occur and direct reform accordingly. This is tractable: the information exists in survey data, deliberative processes, and policy outcomes. The framework to organize it exists in this paper. The missing element is the institutional will to apply it.

An engineering diagnosis without a prescribed cure

The findings of this paper are uncomfortable precisely because they implicate the basic structure of modern democracy rather than its imperfections. A diagnosis that says "your institutions are corrupt" points toward institutional reform. A diagnosis that says "your information channel is too deep to be observable, regardless of institutional quality" points toward architectural reform that no existing institution has an incentive to advocate for.

The paper does not prescribe a cure. It identifies a structural constraint. What follows from that identification — whether the appropriate response is shorter representation chains, more participatory mechanisms, hybrid systems, or some other architecture — involves normative and practical judgments that go well beyond the information-theoretic analysis. The engineering diagnosis sharpens the question. It does not answer it.

What it does answer is whether the current architecture is capable of doing what democratic theory claims it does. On the specific and limited question of preference transmission fidelity, the answer is no — not because of institutional failure, but because of architectural necessity.

Part VI: Conclusion

The argument this paper makes is precise, bounded, and uncomfortable.

It does not claim that representative democracy is illegitimate, that elections are meaningless, or that current democratic institutions should be abandoned. These are normative claims that information theory cannot adjudicate.

The claim is structural: representation chains with three or more layers are constitutionally unobservable under any realistic noise parameters. The policy layer cannot reconstruct the true distribution of citizen preferences from the signals it receives, because the signal was destroyed in aggregation before it arrived. This is a property of the channel, not of the institutions at either end of it.

The simulation makes this visible. Architecture A — a model of a standard parliamentary democracy with five representation layers — achieves a mean tracking error twenty times higher than Architecture D, despite identical institutional quality. It oscillates continuously in response to its own noise, failing to detect genuine preference shifts that Architecture D identifies within one to two time steps. Its representation error heatmap shows complete spatial obliteration: every citizen group appears identical to the policy layer, regardless of how diverse their actual preferences are.

Architecture B — three layers, representing a leaner representative system — is better. Its mean tracking error is half of A's. But it too survives zero percent of citizen preference variance to the policy layer and operates at an SNR of 0.048 — twenty times below the unobservability threshold. Improving the quality of its three layers cannot push it above the threshold. The threshold is a property of the layer count, not of the layers.

This is the result that connects this paper to the first two in the series. Paper one showed that governance systems operating on aggregated signals cannot respond to distributed disturbances — the averaging problem. Paper two showed that governance systems with single-scale architecture cannot respond to multi-frequency disturbances — the frequency gap theorem. This paper shows that governance systems with deep representation chains cannot observe the preferences they are supposed to serve — the constitutional unobservability result.

All three are the same finding viewed from different angles: the aggregation that makes large-scale governance manageable is the same aggregation that makes it unresponsive, blind, and disconnected from the distributed reality it governs. The scale is the problem. Not in the sense that scale is avoidable — some problems require large-scale coordination — but in the sense that scale imposes structural constraints that cannot be wished away by institutional improvement.

The thinkers who first formalized these ideas understood they were describing something with political implications. Wiener saw the cybernetic critique of bureaucracy as a central concern. Ashby's Law of Requisite Variety implies directly that a governance system with insufficient variety cannot govern a high-variety society. Shannon's channel capacity theorem implies that every communication channel has limits that cannot be overcome by improving the communicators. Beer spent decades trying to design institutional structures that respected these limits.

What this series adds is a simulation layer that makes the abstract results quantitatively visible — and a specific application to democratic representation that the cybernetics tradition largely left implicit. The finding is not new. The formalization, the simulation, and the direct application to democratic observability are the contribution.

The question the series leaves open is also the most important one: if the current architecture is constitutionally unobservable, what architecture is not? The simulation demonstrates that shorter chains are more observable, that participatory mechanisms preserve preference variance, and that the fractal assignment of decisions to their most local capable level reduces representation chain length for the decisions that most depend on preference fidelity. But the design of genuinely observable democratic institutions — accountable, legitimate, and capable of collective decision-making at appropriate scales — remains an open problem.

It is a more honest problem to work on, having first established that the current architecture cannot solve it.

21

Appendix A: Mathematical formulations

Observability in linear systems

For a discrete-time linear system:

```
x(t+1) = A · x(t) + B · u(t)
y(t)   = C · x(t) + v(t)
```

The system is observable if the observability matrix:

```
O = [C; CA; CA²; ...; CA^(n-1)]
```

has full column rank n. When O does not have full column rank, there exist state dimensions that produce zero output regardless of their value — they are invisible to any observer at the output.

In the representation chain model, x is the vector of citizen preferences, y is the signal at the policy layer, and C encodes the aggregation and noise structure of the chain. The constitutional unobservability result is the statement that C for a K-layer representation chain with realistic parameters does not allow full-rank O.

Variance survival through a single layer

For a layer with aggregation ratio r (mapping n inputs to n/r outputs by taking group means) and additive Gaussian noise with standard deviation σ:

```
Var_out = Var_in / r + σ²
```

The first term represents variance surviving aggregation. For a group of r inputs with variance V, the mean has variance V/r — the central limit theorem applied to the aggregation step. The within-group variance V·(r-1)/r is destroyed entirely. The second term represents noise added at this layer.

Variance survival through K layers

For a chain of K layers with aggregation ratios $r_1, \ldots, r_K$ and noise levels $\sigma_1, \ldots, \sigma_K$:

Surviving signal variance (the component traceable to true citizen preferences):

```
Var_signal(K) = Var_true · ∏_{k=1}^{K} (1/r_k)
```

Accumulated noise variance (independent across layers):

```
Var_noise(K) = Σ_{k=1}^{K} [ σ_k² · ∏_{j=k+1}^{K} (1/r_j) ]
```

Note: noise introduced at earlier layers is itself attenuated by subsequent aggregation. The formula above accounts for this; earlier noise is partially suppressed while later noise passes through with less attenuation.

Simplified form (when noise at each layer is added after aggregation):

```
Var_noise(K) ≈ Σ_{k=1}^{K} σ_k²   [if r >> 1 at subsequent layers]
```

The simulation uses the simplified form, which slightly overestimates accumulated noise. For the typical parameter values used, the difference is small.

Signal-to-noise ratio and the unobservability threshold

The SNR at the policy layer:

```
SNR(K) = Var_signal(K) / Var_noise(K)
```

Constitutional unobservability threshold: SNR < 1.

Below this threshold, noise variance exceeds signal variance. A maximum likelihood estimator of citizen preferences given the policy-layer observation y has an error variance larger than the prior variance — meaning the observation is less informative than no observation at all.

The threshold crossing point K* solves:

```
Var_true · ∏_{k=1}^{K*} (1/r_k) = Σ_{k=1}^{K*} σ_k²
```

For the typical parameters used in the SNR analytical curve (r = 3.5, σ = 0.17 per layer, Var_true = 0.18):

```
K*: 0.18 · (1/3.5)^K = K · (0.17)²
     0.18 · 0.286^K   = K · 0.0289
```

Solving numerically: K* ≈ 2.0 — the threshold is crossed between K = 2 and K = 3.

The data processing inequality

Shannon's data processing inequality states that for any Markov chain X → Y → Z:

```
I(X; Z) ≤ I(X; Y)
```

Processing (or observing through an intermediate) cannot increase mutual information. For the representation chain, this means that the mutual information between citizen preferences (X) and policy-layer observations (Z) is bounded by the mutual information at the first aggregation step. Each subsequent layer can only reduce it.

Applied to the representation chain: no post-hoc processing at the policy layer can recover mutual information lost in aggregation. The data processing inequality is the formal statement of why institutional quality at the policy layer cannot compensate for aggregation loss in the representation chain.

Preference dynamics

Citizen preferences evolve as:

```
x_i(t+1) = x_i(t) + δ_i(t) + s(t)
```

Where $\delta_i(t) \sim N(0, \sigma_{drift}^2)$ is individual drift ($\sigma_{drift} = 0.015$) and $s(t)$ is a genuine preference shift event (zero except at the shift events).

Preferences are clipped to $[-1, +1]$ at each step. The initial preference distribution for citizen group i is:

```
x_i(0) = μ_{c(i)} + ε_i
```

Where $\mu_{c(i)}$ is the cluster mean (drawn from Uniform(-0.8, 0.8) across P dimensions) and $\varepsilon_i \sim N(0, 0.25^2)$ is within-cluster individual variation.

Policy update rule

The policy layer updates according to:

```
π(t) = π(t-1) + K_p · (ŷ(t - τ_total) - π(t-1))
```

Where $\hat{y}(t - \tau_{total})$ is the policy-layer observation after total representation chain delay τ_{total}, and $K_p = 0.30$ is the policy responsiveness gain. All architectures use identical K_p.

Full simulation parameters

Parameter	Value	Notes
N	60	Citizen groups
P	4	Policy preference dimensions
T	120	Time steps
Clusters	4	Groups of 15 citizens each
σ_drift	0.015	Per-step individual preference drift
Shift 1	t=40, cluster 0, dims 0-1, magnitude 0.4/-0.2	Genuine regional shift
Shift 2	t=80, all groups, dim 2, mean 0.3 (sd 0.08)	System-wide shift
K_policy	0.30	Policy responsiveness gain (all architectures)
Random seed	13	For reproducibility
Warmup	10	Steps excluded from metrics

Architecture A layer parameters (r, σ, τ): (5, 0.12, 2), (4, 0.18, 3), (3, 0.22, 4), (4, 0.20, 5), (3, 0.15, 4)

Architecture B layer parameters: (4, 0.10, 2), (5, 0.18, 4), (3, 0.14, 3)

Architecture C layer parameters: (3, 0.08, 2), (2, 0.10, 2)

Architecture D layer parameters: (1, 0.05, 1)

26

Appendix B: Code and reproduction

Source code

The v5 simulator extends the series from v4's multi-scale stability demonstration to the information-theoretic domain, modeling preference transmission fidelity through representation chains. It is implemented in Python using NumPy and Matplotlib.

The full source code is available at:

github.com/BjornKennethHolmstrom/ggf-governance-simulator (https://github.com/BjornKennethHolmstrom/ggf-governance-simulator)

The repository now includes five simulator versions:

File	Paper	Description
`ggf-simulator-v2.py`	Paper I	Single-node scalar feedback model
`ggf-simulator-v3.py`	Paper I	Ten-node vector model, localized shock
`ggf-simulator-v3-unadjusted.py`	Paper I	v3 with unstable gain — instability demo
`ggf-simulator-v4.py`	Paper II	Multi-scale disturbance, three architectures
`ggf-simulator-v5.py`	Paper III	Representation chain observability, four architectures

Reproducing the results

```
git clone https://github.com/BjornKennethHolmstrom/ggf-governance-simulator
cd ggf-governance-simulator
pip install numpy matplotlib
python ggf-simulator-v5.py
```

The simulation is seeded (`numpy.random.default_rng(seed=13)`). Default parameters exactly reproduce Figure 1 and the quantitative summary table in Part II.

Key architectural differences from v4

v5 is a structural departure from v3/v4, which modelled state-space feedback dynamics. v5 models information transmission through a degradation chain — a different formalism suited to the observability question.

Preference state space. Rather than a scalar stability value per node, each of 60 citizen groups holds a P=4 dimensional preference vector in [−1, +1]. Preferences evolve over time with genuine shift events injected at specified points.

Layer-by-layer degradation. The `pass_through_layers` function implements the aggregation-plus-noise model: for each layer, group-mean aggregation reduces the number of representative units by ratio r, then Gaussian noise is added. The policy layer observes the final output. No feedback loop — this is a pure forward-transmission model of information degradation.

Analytical SNR curve. The `variance_survival_curve` function computes the SNR analytically using the formula in Appendix A, independent of the simulation. This separates the mathematical result from the specific parameter choices of the four architectures.

Spatial representation error. The heatmap panels reconstruct what each architecture's policy layer "sees" by broadcasting the final aggregated signal back to citizen-group resolution and subtracting true preferences. This makes the spatial information destruction visually explicit.

Modifying the parameters

Layer count and parameters (`ARCHITECTURES` dict): The most instructive variation is adding a sixth layer to Architecture B or removing a layer from Architecture A, to observe how the SNR and tracking error change. The layer tuples (r, σ, τ) can be varied independently.

Noise structure: Setting all σ values to zero produces the pure-aggregation case — demonstrates that information loss occurs even with perfect representation at each layer, purely from the aggregation step. Setting r = 1 for all layers (no aggregation) with varying noise shows the noise-only degradation curve.

Genuine shift magnitude: Increasing the shift magnitude at t=40 and t=80 (in the `generate_preferences` function) makes the genuine signal stronger, which raises the effective SNR temporarily. Even with large shifts, Architecture A typically cannot detect them above its noise floor.

Policy responsiveness K_policy: Increasing this above 0.5 causes Architecture A to oscillate more severely (it amplifies its noise-driven signal). Decreasing it makes all architectures respond more slowly to genuine shifts. The value of 0.30 is chosen to show moderate responsiveness — neither so fast as to destabilize A nor so slow as to prevent D from tracking shifts cleanly.

Extending the model

The most valuable extensions from the authors' perspective:

Empirical noise calibration: Fitting the noise parameters to real survey-to-policy correlation data would make the SNR estimates empirically grounded rather than illustrative. Electoral studies, deliberative polling experiments, and the literature on the preference-policy gap provide relevant data.

Strategic noise models: Replacing Gaussian noise with structured noise matching known media selection biases and party positioning dynamics would produce a more realistic model of the representation chain. The qualitative result would be the same; the quantitative threshold might shift.

Hybrid architectures: Modelling a three-layer system where dimension-2 decisions bypass layers 2-3 and are handled by direct participation would demonstrate the complementarity of the observability and fractality results — different decisions routed through appropriate layer counts.

Contributing

The repository is open source under MIT license. Extensions, empirical applications, and critiques are welcome via GitHub.

30

Appendix C: References and sources

A note on methodology

As with the preceding papers in this series, the concepts here were developed through extended dialogue with multiple AI systems — Claude (Anthropic), ChatGPT (OpenAI), Gemini (Google), DeepSeek, and Grok (xAI) — rather than through direct reading of the primary literature. The references below are the sources those systems identified as foundational, provided for readers who wish to engage with the primary literature directly.

The specific contribution of this paper — formalizing democratic representation as a degraded observation channel, deriving the constitutional unobservability threshold, and demonstrating the preference-policy gap as an architectural consequence — emerged from this collaborative process. The underlying mathematical tools belong to long-established traditions in information theory and control theory. The political science empirics belong to an independent literature. This paper brings them together.

Information theory and observability

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication.* University of Illinois Press.

The channel capacity theorem is the foundational result underlying this paper's argument. Shannon's demonstration that every communication channel has an information transmission rate bounded by its bandwidth and noise characteristics — and that this bound cannot be overcome by improving the communicators — is the precise formal basis for the constitutional unobservability result.

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory.* 2nd ed. Wiley.

The data processing inequality — that processing cannot increase mutual information — is formally derived and extensively discussed here. This inequality is the statement that no post-hoc institutional processing at the policy layer can recover information destroyed in representation chain aggregation.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering,* 82(1), 35–45.

Kalman's foundational paper establishes the observability condition for linear systems. The observability matrix construction referenced in Appendix A follows this tradition.

Åström, K. J., & Murray, R. M. (2008). *Feedback Systems: An Introduction for Scientists and Engineers.* Princeton University Press. (Available freely at cds.caltech.edu/~murray/amwiki)

Chapter 7 covers observability and observer design in the context accessible to scientists without a control engineering background. The formal observability condition and its implications for state reconstruction are developed clearly.

Democratic theory and preference transmission

Dahl, R. A. (1971). *Polyarchy: Participation and Opposition.* Yale University Press.

The foundational modern statement of democratic theory as preference transmission. Dahl's criteria for polyarchy — including the responsiveness criterion that government policies correspond to citizen preferences — are the normative commitments whose technical feasibility this paper examines.

Bartels, L. M. (2008). *Unequal Democracy: The Political Economy of the New Gilded Age.* Princeton University Press.

Empirical documentation of the weak and declining correlation between citizen preferences and policy outcomes in the United States. Bartels attributes this primarily to elite influence and unequal political participation; the observability framework provides a complementary structural explanation.

Gilens, M., & Page, B. I. (2014). Testing theories of American politics: Elites, interest groups, and average citizens. *Perspectives on Politics*, 12(3), 564–581.

The most-cited empirical study of preference-policy correspondence. Finds that average citizens' preferences have essentially zero independent influence on policy outcomes when controlling for elite preferences. The observability framework suggests this is partly architectural: by the time average citizen preferences reach the policy layer through a five-layer chain, they are noise-dominated and indistinguishable from the signal injected by organized interests with shorter chains.

Achen, C. H., & Bartels, L. M. (2016). *Democracy for Realists: Why Elections Do Not Produce Responsive Government.* Princeton University Press.

A systematic empirical challenge to the folk theory of democracy as preference aggregation. Argues that elections do not reliably transmit voter preferences to policy. The observability framework provides a formal explanation for why they cannot, independent of the retrospective voting and group identity arguments Achen and Bartels develop.

Dryzek, J. S. (2000). *Deliberative Democracy and Beyond.* Oxford University Press.

Deliberative democracy as a theoretical response to the aggregation problem — attempting to form preferences through reasoned discussion rather than aggregate pre-existing ones. The observability framework suggests deliberative mechanisms at lower layer counts (citizens deliberating directly about local decisions) preserve more signal than deliberative mechanisms inserted into a multi-layer representative chain.

Cybernetics and systems theory

Ashby, W. R. (1956). *An Introduction to Cybernetics.* Chapman and Hall.

The Law of Requisite Variety — a controller must have at least as much variety as the system it controls — is the general form of which the observability result is a specific case. A policy layer whose signal has been compressed through K aggregation stages has less variety than the citizen population it governs. Ashby's law predicts it cannot govern them accurately; the observability model quantifies exactly why.

Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine.* MIT Press.

Wiener's critique of bureaucracy as a communication system with too many filtering layers anticipates this paper's central argument. His observation that information degrades through organizational hierarchies — that the effective decision-maker at the top of a large organization is working from a severely compressed and distorted signal — is a precursor to the formal model developed here.

Beer, S. (1979). *The Heart of Enterprise.* Wiley.

Beer's treatment of the viable system's System 4 (environmental scanning) and its relation to System 5 (policy) addresses precisely the question of how policy layers can obtain unfiltered information about the system they govern. Beer's algedonic channel — the fast signal that bypasses the normal hierarchy in crisis — is a design response to the observability problem, though Beer does not formalize it in these terms.

Participatory and deliberative democracy

Fishkin, J. S. (2009). *When the People Speak: Deliberative Democracy and Public Consultation.* Oxford University Press.

The deliberative polling methodology as a practical mechanism for reducing representation layers: a random sample of citizens deliberates directly about policy questions, producing a more accurate signal than conventional polling or elections. Directly relevant to the Architecture C and D models.

Ostrom, E. (1990). *Governing the Commons.* Cambridge University Press.

Empirical evidence that communities can self-govern shared resources effectively at small scale, with minimal intermediation between citizens and rules. The observability framework provides a formal explanation: at small scale with minimal layers, citizen preferences are observable.

Landemore, H. (2020). *Open Democracy: Reinventing Popular Rule for the Twenty-First Century.* Princeton University Press.

The case for open, participatory democratic systems as responses to the representational failures of current institutions. The observability framework provides a technical grounding for the claim that representation at scale is architecturally problematic, independent of the normative arguments Landemore develops.

Political communication and media

Lippmann, W. (1922). *Public Opinion.* Harcourt, Brace.

The earliest systematic treatment of media as a filtering layer between citizens and the political world. Lippmann's concept of the "pseudo-environment" — the simplified, selected representation of reality that citizens and politicians actually respond to — anticipates the noise-dominated signal problem this paper formalizes.

Prior, M. (2007). *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections.* Cambridge University Press.

The differentiated impact of media structure on political information transmission. The structural diversity of media environments affects the noise characteristics of the media layer — but the observability framework suggests that even ideal media transmission cannot compensate for the aggregation loss that occurs before the media layer and the further aggregation that occurs after it.