On First-Meter Problems

Dynamic Governability under Endogenous Technological Change
A Technical Framework for Strategic Foresight and Policy Optionality
with Frontier AI as a Case Study

David Dosu

Research Monograph: August 21, 2026

Abstract

This monograph develops a technical foresight framework for systems in which technological change also changes the future capacity of institutions to respond. Infrastructure, financial exposure, switching costs, political coalitions, institutional benchmarks, human expertise, and outside options can become state variables that alter which later interventions remain feasible. I call this a *first-meter problem*: current movement changes both the technological state and the future response technology.

The framework has three operating levels. A qualitative mode maps mechanisms and signs when data are sparse. A bounded mode uses intervals, partial identification, and scenario ranges. An estimated mode identifies transition and response parameters and supports prospective testing. The general stochastic layer represents state-dependent response, uncertainty, and risk-constrained capacity. Under monotonicity, an exact boundary operator generates the maximal quiet path. Under positive-affine structure, the central feedback object is

$$L = C(I - H)^{-1}B,$$

where B maps action into persistent state, H propagates that state, and C maps inherited state back into future room for action. When $\rho(H) < 1$, the boundary is Schur stable exactly when $\rho(L) < 1$. The framework separates the spectral regime, the margin $1 - \rho(L)$, steady amplification $(I - L)^{-1}$, finite-horizon transient amplification, and regime-switching risk.

A coalition-enforcement layer models state-dependent response costs, heterogeneous exposure, common measurement, coordination, and pre-agreed response protocols. A strategic layer shows how a forward-looking technological actor can value present action because it improves future room for action, while institutions can value preservation because it protects future response capacity. A welfare layer defines the option value of preserving future policy choice. The empirical layer states identification conditions, uncertainty propagation, falsification tests, and a prospective forecasting protocol.

For frontier AI, a benchmark-plus-entrenchment specialization gives $L = 1 - \eta + \Gamma$. Quiet accumulation is bounded when $\Gamma < \eta$, with ceiling $K_0/(\eta - \Gamma)$, critical at equality, and supercritical above it. A global frontier-AI estimate of $\rho(L)$ is an empirical task tied to a specified state definition, time scale, action variable, response boundary, and identification design. The intended use is conditional foresight: identify state channels early, define leading indicators, state what would falsify each mechanism, stress-test regime changes, and preserve future response options before dependence makes them expensive to recover.

Keywords: dynamic governability; technological foresight; first-meter problems; path dependence; state capacity; viability theory; technological lock-in; positive systems; small-gain theorem; strategic regulation; option value; collective action; frontier AI.

Executive Summary

Institutions often operate on the assumption that they can allow a technology to develop and still retain the ability to step in and regulate it later if it becomes dangerous, disruptive, or otherwise begins to move beyond acceptable limits. A first-meter problem, as addressed in this monograph, arises when the development they permit alters the conditions of that future response itself—altering which interventions remain available, how costly or credible they are, and the institutional or economic environment in which the decision would later be made.

That assumption has sometimes held. The Montreal Protocol, for example, shows that institutions can retain the ability to tighten intervention as knowledge improves and a regulated technology evolves, especially when measurement, substitution, transition support, and pre-agreed response mechanisms remain available. Carbon infrastructure illustrates the opposite problem: once long-lived assets, production systems, and economic interests accumulate around a technology, later intervention can inherit constraints created by earlier deployment. The lesson is therefore less that institutions can never act later than that their ability to do so depends on what happens during the period in which they wait. Carrying the relaxed assumption of future control into the next wave of technological development is especially consequential when technologies can become deeply embedded in infrastructure, finance, organizational workflows, human capability, and state interests before their full effects are known.

This paper calls that class of problems **first-meter problems**. The central idea is that the importance of an early technological move can exceed the size of the move itself because the first movement changes the conditions governing every movement that follows. Research-speak for the simple idea that while institutions are waiting to see what those technologies become, deployment can steadily alter the economic, political, infrastructural, and institutional conditions that determine whether those same institutions will still be able to respond effectively later. If infrastructure can be built around a technology, debt can be issued against it, firms can reorganize around its continued use, governments can acquire fiscal or strategic interests in its success, organizations can lose alternative ways of performing important tasks, and countries can become dependent on foreign suppliers, then the technology's initial movement can alter the conditions under which later intervention is possible. Repeated tolerance changes the benchmark against which later activity is judged. And by the time intervention becomes unavoidable, the institution making that decision may therefore confront a different economic, political, and technical system from the one in which the original decision to wait was made.

The paper develops a framework around **dynamic governability**: the future ability of institutions to restrict, redirect, substitute, migrate away from, coordinate around, or otherwise respond effectively to a technology. Usually, ordinary technological forecasting asks how capability, adoption, cost, or deployment may change. First-meter analysis adds a second object of foresight by asking how that same technological path changes the future capacity of institutions to act upon it. Technological progress and governability are therefore treated as processes that can, and would have to, evolve together.

The core mechanism is a feedback process. Technological action can create persistent forms of dependence, and those dependencies can then alter the conditions of later response. Infrastructure, financial exposure, organizational reliance, political interests, human capability loss, and international dependence can survive and interact in ways that make intervention slower, more expensive, less credible, or harder to coordinate. A weaker response environment can in turn create greater room for further technological expansion, producing still more dependence and reinforcing the process. The

paper studies when this feedback weakens, when it approaches a critical boundary, when relatively small structural changes can produce large effects, and when changes between political or economic regimes create risks that a fixed view of the system would miss.

This also gives the paper a broader account of technological lock-in. Future governability can deteriorate through several moderate dependencies acting together even when no single channel appears decisive on its own. Financial exposure, scrutinized individually, may look manageable, infrastructure commitments may look reversible, organizational dependence may appear limited, and political interests may seem containable, while their interaction produces a system in which changing course becomes substantially harder. The relevant object is therefore the combined effect of technological development on the future response environment.

The strategic layer adds a further complication because technological actors can have incentives to shape that environment in efforts to achieve moats. A forward-looking actor can value deployment for both its immediate return and the effect it has on future conditions. If greater deployment creates dependencies that raises the future cost of restriction, expansion can improve the actor's later room for action as well as its present position. Institutions face the mirror incentive: they can value action today because preserving alternatives, response capacity, or institutional credibility improves the choices available to them later.

This leads to one of the paper's main policy implications: **early governance does not have to mean early restriction**. Institutions can allow technological development to continue while protecting the conditions that make later intervention possible. Interoperability can preserve migration options. Multiple suppliers can preserve substitutes. Independent scientific and technical capability can preserve the capacity to audit, recover, and operate outside a dominant system. Diversified financing can reduce the collateral cost of intervention. Contingency infrastructure and credible response protocols can preserve exit and coordination options.

The policy objective is therefore broader than deciding how much technological development to permit today; it also includes protecting the ability to make a different decision tomorrow.

The same problem appears at the international level. Better measurement can produce a clearer shared understanding of a technological risk while leaving the underlying collective-action problem unresolved. Governments may agree on the severity of a problem while facing very different economic, political, or strategic costs from responding to it. As technological dependence accumulates unevenly across countries, a coalition that could act effectively today may become more expensive or harder to assemble later. Even a formal response protocol can lose practical force if the cost of honoring it rises with dependence while the incentives supporting participation remain fixed.

Frontier AI provides the paper's main application because it could combine several of these mechanisms within one rapidly developing system. Compute and energy infrastructure can create physical commitments; financing can connect continued deployment to wider balance sheets; firms and public institutions can become dependent on particular providers and workflows; human substitution can weaken independent capability; and governments can acquire fiscal, strategic, or geopolitical interests in continued access. Institutional benchmarks and international coalitions can change at the same time. The paper treats these channels as empirical hypotheses rather than established facts, and asks which of them actually become persistent enough to alter later restriction, substitution, migration, coordination, or other forms of response.

The framework is designed to study that question at different levels of evidence. When data are sparse, analysts can map candidate mechanisms, identify the states through which they operate, track leading indicators, and specify observations that would weaken the proposed explanation.

When relationships can be bounded, the framework can support conditional ranges and robust conclusions. When the relevant transition and response effects can be identified from data, it can support estimation and prospective testing. The aim is to move from plausible stories about future lock-in toward claims that can be tested before the outcome is known.

That empirical discipline is central to the framework. A first-meter explanation should earn predictive value against simpler accounts based on concentration, leverage, sunk capital, market share, switching costs, lobbying, and conventional path dependence. If those simpler explanations consistently predict later response difficulty better, or if the proposed state channels fail to change future governability in the expected direction, the first-meter account should lose empirical support.

The contribution of the paper is therefore to organize a broad class of technological and institutional feedbacks around one question:

How does action today change the set, cost, speed, or credibility of the responses available tomorrow?

For the next wave of technological progress, this question matters because waiting is itself capable of changing the system. Institutions may learn more with time, but the period in which they are learning can also produce infrastructure, dependence, constituencies, financial exposure, and new institutional expectations that alter what they will later be able or willing to do.

The relevant question is therefore not only:

How far has the technology moved?

It is also:

What did that movement do to our ability to choose differently later?

That is the first-meter problem.

Purpose and Use

This research monograph is designed for study, technical discussion, scenario analysis, and prospective foresight. It is intended for teams that need to reason about fast technological change while the economic and institutional environment is also changing. Relevant users can include frontier-technology laboratories, governments, international organizations, central banks, standards bodies, scientific institutions, and research groups.

The framework integrates benchmark drift, coalition selection, information, response protocols, participation, historical position, stochastic control, positive systems, strategic interaction, welfare, and empirical design into one reference architecture. It is designed to support both technical audit and live foresight work.

The framework asks a specific question: *How does action today change the set, cost, speed, or credibility of responses that remain available tomorrow?* It therefore tracks technological progress and governability as a coupled system. The framework supports analysis before a full empirical model exists, but it requires the analyst to mark the difference between a mechanism hypothesis, a bounded estimate, and an identified quantitative result.

Operating rule. A number such as $\rho(L)$ is reported only after the state definition, time scale, action variable, response boundary, and identification design are stated. Before that point, the same framework is used to map signs, feedback paths, leading indicators, and conditional forecasts.

Contents

Executive Summary **2**

Purpose and Use **4**

I Framework for Foresight and Decision Analysis **11**

1 The Framework in One View **11**

1.1 The coupled object 11
1.2 The diagnostic set 11
1.3 Three levels of use 12

2 A Practical First-Meter Foresight Protocol **12**

2.1 Step 1: define the decision object 12
2.2 Step 2: build the state register 13
2.3 Step 3: map every mechanism into the transition-response interface 13
2.4 Step 4: separate the institutional layers 13
2.5 Step 5: identify anchors, moving benchmarks, and waiver rules 13
2.6 Step 6: construct the response model 14
2.7 Step 7: calculate or bound the diagnostic objects 14
2.8 Step 8: write conditional forecasts 14
2.9 Step 9: stress the path 14
2.10 Step 10: design a preservation portfolio 14
2.11 Two-day institutional exercise 15

3 The First-Meter Diagnostic Worksheet **15**

3.1 Frontier-AI indicator map 16
3.2 Four example forecast forms 17
3.3 Worked example: AI infrastructure build-out 17

4 Interpretation Rules and Failure Modes **19**

II Technical Foundations **19**

5 Technical Introduction and Scope **20**

6 Relation to the Literature and Scope of the Claim **23**
6.1 Governability, technology control, and path dependence . . . 23
6.2 Viability, constrained control, and stochastic games . . . 23
6.3 Monotone and positive systems . . . 24
6.4 Collective action and equilibrium selection . . . 24
6.5 State capacity, irreversibility, and strategic regulation . . . 24
6.6 Frontier AI as infrastructure and economic dependence . . . 25
6.7 Novelty claim . . . 25

7 The General Dynamic-Governability Model **26**
7.1 State, technological action, and institutional response . . . 26
7.2 Response correspondences and strategic institutions . . . 27
7.3 A general definition of a first-meter trap . . . 27
7.4 The plug-in contract . . . 28
7.5 History dependence and Markovization . . . 29
7.6 Partial observation and belief states . . . 29
7.7 Structural uncertainty . . . 29

8 Deterministic Monotone First-Meter Systems **30**
8.1 Fixed points and nonlinear edge cases . . . 31

9 Positive Affine Systems and the Governability Loop **31**
9.1 Vector actions and vector state . . . 31
9.2 The loop-gain matrix . . . 32
9.3 Coordinate invariance . . . 33
9.4 Complete finite-dimensional asymptotic classification . . . 33
9.5 A clean positive special case . . . 34
9.6 Transient governability . . . 34

10 Spectral Margin, Steady Amplification, and Comparative Statics **35**
10.1 Scope of cross-domain comparison . . . 36

11 Changing Regimes: Switched and Stochastic Systems **36**

12 A Coalition-Enforcement Microfoundation **38**
12.1 The stage game . . . 38
12.2 Heterogeneous exposure and coalition design . . . 38

12.3 Measurement, common knowledge, and response selection . . . 39
12.4 Finite-noise response probability . . . 40
12.5 Forward-looking institutions . . . 40

13 Strategic Entrenchment and the Governability Race **41**
13.1 The dynamic governability game . . . 41
13.2 A general strategic first-meter wedge . . . 42
13.3 The institutional mirror: preservation has continuation value . . . 42
13.4 A closed-form policy-first governability race . . . 43
13.5 Endogenous feedback architecture . . . 44
13.6 From the exact benchmark to application . . . 44

14 The Benchmark-and-Entrenchment Specialization **45**
14.1 State dynamics . . . 45
14.2 Embedding in the positive system . . . 45
14.3 Illustrative walk toward the phase boundary . . . 46
14.4 Benchmark-only corollary . . . 47
14.5 Formal anchor revision and passive normalization . . . 47
14.6 Why the adaptation rates disappear from the stable ceiling . . . 47

15 Coalition Switching, Participation, and History **48**
15.1 Endogenous coalition composition . . . 48
15.2 Participation and historical position . . . 48

16 Cumulative Hazard, Risk Budgets, and Continuous Time **49**
16.1 Cumulative hazard across periods . . . 49
16.2 Continuous-time extension . . . 49
16.3 Positive-affine continuous-time loop . . . 50

17 The Economic Value of Preserving Governability **50**
17.1 A Bellman formulation . . . 50
17.2 Structural design in the positive-affine model . . . 51
17.3 A technology-specific response-capacity stock . . . 51
17.4 Redundancy as an economic asset . . . 52
17.5 Intergenerational governability . . . 52

18 Measurement, Identification, and Falsification **52**

18.1 State-space implementation . . . 52
18.2 Identification conditions . . . 53
18.3 Point, partial, and robust identification . . . 53
18.4 Candidate identification shocks . . . 53
18.5 Testable implications . . . 54
18.6 Validation beyond retrodiction . . . 54
18.7 Competing models and falsification . . . 54
18.8 A pre-analysis template for prospective work . . . 55

19 Cross-Domain Mechanism Evidence and Validation Design 55
19.1 Carbon infrastructure . . . 56
19.2 Banking and endogenous rescue incentives . . . 56
19.3 Digital platforms and interoperability . . . 56
19.4 The Montreal Protocol as a preservation case . . . 56
19.5 What a genuine cross-domain validation would require . . . 57

20 Application to Frontier AI 57
20.1 The current institutional setting . . . 57
20.2 Compute and energy . . . 57
20.3 Financial entanglement . . . 57
20.4 Rebound demand and continual use . . . 58
20.5 Platforms, action histories, and organizational dependence . . . 58
20.6 Human capability and substitution capacity . . . 58
20.7 Fiscal and political dependence . . . 59
20.8 Compute balance of payments and sovereign dependence . . . 59
20.9 Correlated model errors . . . 59

21 Beyond Linearity: Saturation, Hysteresis, and Computed Viability 59
21.1 A saturating entrenchment example . . . 60
21.2 Hysteresis . . . 60
21.3 Viability recursion and institutional response timing . . . 60

22 Policy Design for Preserving Future Governability 62
22.1 A general structural-design problem . . . 62
22.2 A governability diagnostic set . . . 62
22.3 Scalar specialization . . . 63

22.4 Redundancy and coalition option value . . . 63
22.5 Response protocols can erode . . . 63

23 Discussion **64**
23.1 Scope of generalization . . . 64
23.2 Irreversibility and hysteresis . . . 64
23.3 Time inconsistency and commitment . . . 64
23.4 The option value of early governance . . . 64
23.5 Empirical scope of the frontier-AI interpretation . . . 65

24 Conclusion **65**

A Proofs **66**
A.1 Proof of Theorem 1 . . . 66
A.2 Proof of Proposition 1 . . . 66
A.3 Proof of Proposition 2 . . . 66
A.4 Proof of Theorem 2 . . . 67
A.5 Proof of Theorem 3 . . . 67
A.6 Proof of Proposition 3 . . . 68
A.7 Proof of Proposition 4 . . . 68
A.8 Proof of Theorem 4 . . . 68
A.9 Proof of Theorem 5 . . . 69
A.10 Proof of Proposition 7 . . . 69
A.11 Proof of Proposition 8 . . . 69
A.12 Proof of Proposition 9 . . . 70
A.13 Proof of Theorem 6 . . . 70
A.14 Derivation of Equation 91 . . . 70
A.15 Proof of Theorem 8 . . . 70
A.16 Proof of Equation 138 . . . 71
A.17 Proof of Proposition 12 . . . 71
A.18 Proof of Corollary 1 . . . 72
A.19 Proof of Proposition 13 . . . 72
A.20 Proof of Theorem 9 . . . 72
A.21 Proof of Proposition 5 . . . 73
A.22 Proof of Proposition 6 . . . 73

A.23 Proof of Proposition 10 . . . 73
A.24 Proof of Proposition 11 . . . 74
A.25 Proof of Theorem 7 . . . 74
A.26 Proof of Proposition 14 . . . 75
A.27 Proof of Proposition 16 . . . 75
A.28 Proof of Proposition 17 . . . 75
A.29 Proof of Proposition 15 . . . 75

B Additional Edge Cases and Diagnostic Quantities **76**
B.1 Critical nonlinear examples . . . 76
B.2 Reachability and observability . . . 76
B.3 Output-specific transient exposure . . . 76
B.4 Robust viability kernel . . . 76

C Numerical Verification Audit **77**

Part I

Framework for Foresight and Decision Analysis

1 The Framework in One View

1.1 The coupled object

Ordinary technology forecasting asks how capability, adoption, cost, or deployment may change. Dynamic-governability forecasting adds a second object: how the capacity to redirect, restrict, substitute, coordinate around, or exit from the technology changes as a consequence of that same path.

A first-meter mechanism has the following causal form:

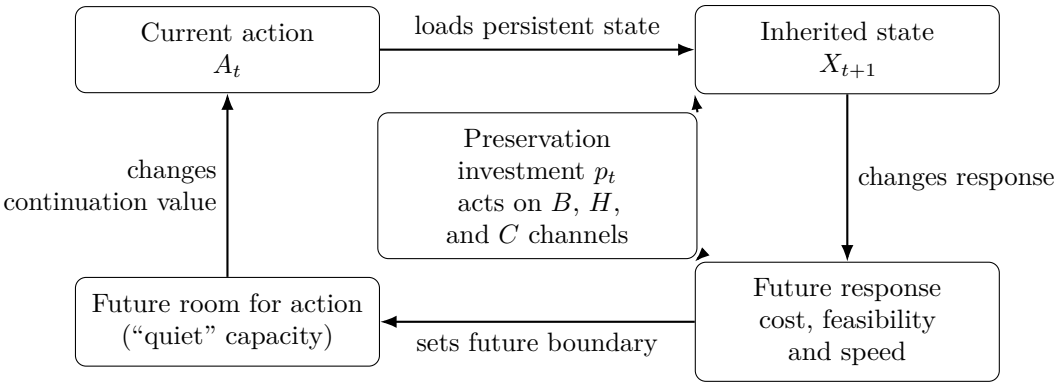


Figure 1: The first-meter loop. Current action changes the state inherited by the future response system. Preservation acts on the same transition and response channels.

In the positive-affine layer, this loop becomes

$$x_t = Hx_{t-1} + Ba_t + d, \qquad a_t \leq k + Cx_{t-1}. \tag{1}$$

The three matrices have a direct foresight interpretation:

- B asks which persistent states are created by current action.
- H asks how those states persist, decay, or propagate into one another.
- C asks how inherited state changes the future response boundary.

When $\rho(H) < 1$, their complete long-run return loop is

$$L = C(I - H)^{-1}B. \tag{2}$$

The series $(I - H)^{-1} = I + H + H^2 + \cdots$ counts every persistent propagation path through the state system before the effect returns to the response boundary.

1.2 The diagnostic set

A useful foresight report separates six objects instead of compressing them into one risk score.

Object	Formal object	Question for the analyst
Feedback regime	$\rho(L)$	Does repeated action-to-state-to-boundary feedback decay, sit near the boundary, or expand in the positive-affine model?
Spectral margin	$1 - \rho(L)$	How close is the maintained model to its feedback boundary?
Steady amplification	$(I - L)^{-1}$	How much can persistent structural forcing be amplified even while the system remains bounded?
Transient exposure	$A_T(M)$ or output-specific G_T	Can temporary loss of response capacity become large over the relevant decision horizon?
Regime-sequence risk	joint spectral radius or Lyapunov exponent	Can switching among regimes create growth that fixed-regime analysis misses?
Governability option value	counterfactual welfare difference	What is the value of preserving future response choices through redundancy, substitution, expertise, or institutional capacity?

Table 1: Core diagnostic outputs. Each object answers a different foresight question.

1.3 Three levels of use

Level 1: mechanism map. Use signs and causal paths. State which variables are candidate persistent states, which channels can increase or decrease future response cost, and which observations could distinguish the mechanisms. This level supports structured foresight when data are sparse.

Level 2: bounded model. Use intervals, partial identification, robust viability sets, and scenario ranges. Report which conclusions hold across the admissible parameter set. This level is appropriate when several parameters can be bounded but point identification is weak.

Level 3: estimated model. Estimate or recover (H, B, C) or the nonlinear response objects under a stated research design. Report sampling or identification uncertainty, test forecasts out of sample, and update the model when the state definition or response technology changes.

These levels can coexist. A team can have an estimated compute-dependence channel, a bounded financial channel, and a qualitative geopolitical channel in the same foresight exercise. The output should preserve those different evidence grades.

2 A Practical First-Meter Foresight Protocol

The protocol converts the theory into a repeatable analysis. Its purpose is to generate conditional forecasts and preservation options before a high-dependence state is reached.

2.1 Step 1: define the decision object

State the technological action, the institutional response of interest, the time horizon, and the unit of analysis. Examples include model deployment, compute construction, autonomous-agent adoption, a new scientific platform, a critical infrastructure transition, or a financial technology. Define what "effective response" means in this application: restriction, substitution, migration, audit, coordinated enforcement, technical shutdown, financial containment, or another action.

2.2 Step 2: build the state register

List only variables that carry information from the past into future technological evolution or response. For each state, record its unit, time scale, likely persistence, observability, and recovery path. Typical classes include infrastructure, financial exposure, concentration, interoperability, human capability, fiscal dependence, political organization, legal precedent, scientific measurement capacity, outside options, and coalition composition.

2.3 Step 3: map every mechanism into the transition-response interface

For each candidate mechanism, state:

(a) the state it adds or changes;
(b) how current technological action loads that state;
(c) how the state persists or propagates;
(d) how it changes future response feasibility, cost, speed, or credibility;
(e) the observable data that can identify or constrain the channel.

This is the operational form of the plug-in contract developed later in the monograph. A verbal scenario remains a hypothesis until these objects are stated.

2.4 Step 4: separate the institutional layers

Measurement, common knowledge, response incentives, response capacity, and participation solve different problems. A shared scientific assessment can improve the information state while strategic multiplicity remains. A response protocol changes incentives after a certified trigger. Capacity investment changes the future cost of response. Participation transfers and representation rules change which coalition is feasible. Formal anchor revision changes the rule itself, while passive normalization changes the benchmark generated by tolerated practice.

This separation is central for international foresight. It prevents a team from treating better measurement as a complete response architecture or treating a treaty text as fixed when the economic cost of honoring it can change with deployment.

2.5 Step 5: identify anchors, moving benchmarks, and waiver rules

Record the formal anchor against which cumulative movement is tracked. Record any adaptive benchmark generated by tolerated outcomes. State exactly how a formal revision occurs. Tolerance and formal revision should be represented as separate state transitions when the institution intends them to have different meanings.

In the scalar specialization, η measures retained weight on the formal anchor while $1 - \eta$ measures the role of the moving benchmark. The entrenchment aggregate Γ measures long-run feedback from dependence into future response cost. Their comparison gives the transparent threshold $\Gamma < \eta$ under the stated tracker model.

2.6 Step 6: construct the response model

Identify who must act, what counts as a winning response coalition, and how state variables change each actor's cost, exposure, outside option, or willingness to participate. Where coalition structure is unknown, compute several plausible response technologies. Where legitimacy or representation requirements matter, encode them inside the family of winning coalitions rather than treating them as an afterthought.

2.7 Step 7: calculate or bound the diagnostic objects

Use the strongest level supported by the evidence. Calculate L only under a specified positive-affine model with $\rho(H) < 1$. In nonlinear models, use local Jacobians, finite-horizon viability, saturation, hysteresis, or simulation as appropriate. In switching environments, test regime sequences rather than relying only on each regime's fixed matrix.

2.8 Step 8: write conditional forecasts

Every forecast should state seven fields:

Target	The future response outcome, state, margin, or exposure being forecast.
Horizon	The date or decision horizon.
Condition	The assumed technological path, policy path, and regime.
Mechanism	The B, H, C, coalition, hazard, or nonlinear channel that generates the forecast.
Leading indicators	Variables expected to move before the target outcome.
Disconfirming evidence	Observations that would weaken or reject the mechanism.
Update rule	The event, date, or new evidence that triggers revision of the forecast.

A forecast can be directional, interval-valued, or probabilistic. The output should match the identification level. This keeps scenario construction separate from empirical confidence.

2.9 Step 9: stress the path

Run at least four stress classes: faster technological action with unchanged feedback structure; stronger persistence or cross-coupling; a discrete institutional or geopolitical regime change; and a response delay. This separates "the technology moved faster" from "the feedback architecture became harder to govern." The distinction matters because faster a_t can compress calendar time while the phase object changes only when the structural feedback changes.

2.10 Step 10: design a preservation portfolio

Search for interventions that reduce persistence H, action-to-state loading B, state-to-boundary sensitivity C, baseline forcing, coalition friction, or response delay. Candidate interventions include interoperability, multiple suppliers, independent technical expertise, public alternatives, diversified

financing, common measurement, adaptive protocols, contingency infrastructure, and explicit non-waiver accounting. Evaluate each intervention against its resource cost and its effect on future policy options.

2.11 Two-day institutional exercise

A compact workshop can use the following sequence.

Block	Work	Required output
Day 1, AM	Define action, response, horizon, formal anchor, and state register.	One-page system definition and observable-state table.
Day 1, PM	Map mechanisms into $B/H/C$ or nonlinear equivalents; define coalition and information structure.	Causal loop map, evidence grades, and unknown-parameter register.
Day 2, AM	Calculate or bound diagnostics; stress faster action, entrenchment, delays, and regime switching.	Conditional forecast set, leading indicators, and failure conditions.
Day 2, PM	Design preservation interventions and pre-specify updates.	Preservation portfolio, trigger conditions, owners, and next measurement date.

Table 2: A two-day operating form of the first-meter framework.

3 The First-Meter Diagnostic Worksheet

The following worksheet is intended to be copied into a live foresight exercise.

Question	What to observe	Model object	Warning pattern
What dependence does current action create?	New infrastructure, contracts, workflow dependence, installed base, human substitution, fiscal exposure	B	A unit of action increasingly creates states that are costly to reverse.
How persistent is the dependence?	Depreciation, migration time, recovery time, asset life, skill recovery, contract duration	H	Persistence rises, recovery slows, or cross-coupling spreads dependence across sectors.
How does dependence change response?	Switching cost, coalition cost, legal or political feasibility, collateral loss, service interruption	C or $q(x, a)$	The same technological severity produces a weaker, slower, or more expensive response at higher inherited dependence.

Question	What to observe	Model object	Warning pattern
How close is the feedback boundary?	Estimated or bounded loop terms under one specification	$\rho(L)$ and $1-\rho(L)$	The margin shrinks or uncertainty intervals reach the boundary.
Can a bounded system still become hard to control temporarily?	Short-horizon propagation, leverage, bottlenecks, response delay	$A_T(M)$, G_T	Large transient exposure despite a subcritical asymptotic model.
Can regime changes interact badly?	Elections, treaties, market structure, supplier changes, conflict, legal changes	joint spectral radius / switched model	Individually stable regimes form an unstable sequence.
Is the reference point moving?	Formal threshold, tolerated practice, waiver language, benchmark updates	η, benchmark state	Passive normalization increasingly replaces cumulative accounting from the formal anchor.
Which actors must coordinate?	Winning coalitions, costs, exposures, outside options, representation constraints	$K(x)$, $\mathcal{W}$	The cheapest winning coalition becomes more expensive or disappears.
Does measurement produce action?	Evaluation coverage, publication, common reports, protocol compliance	observation kernel and response game	Evidence quality improves while action incentives remain unchanged.
What preserves future choice?	Substitutes, interoperability, spare capacity, expertise, alternative coalitions	policy p, response-capacity stock	Current cost rises modestly while future exit or response cost falls materially.
What would falsify the model?	Pre-specified competing measures and later outcomes	empirical protocol	Simpler models predict response difficulty better or proposed state channels fail to move later response.

Table 3: First-meter diagnostic worksheet.

3.1 Frontier-AI indicator map

For frontier AI, the framework suggests a register of candidate state variables rather than a single "AI risk" score.

Channel	Candidate leading indicators	Primary mapping	Foresight question
Compute and energy	provider concentration, migration time, committed power, grid bottlenecks, spare capacity	H, B, C	Does more deployment make later substitution or restriction materially more costly?
Finance	debt issuance, lease commitments, guarantees, refinancing concentration, collateral exposure	B, C	Does infrastructure finance create claims whose impairment raises the cost of intervention?
Platforms and data	switching cost, API dependence, data flywheels, workflow integration, interoperability	H, C	Does adoption make the next unit harder to exit from or easier to entrench?
Human capability	task recovery without frontier systems, retained expert capacity, training pipeline	H, C	Does substitution reduce independent capacity to operate, audit, or recover?
Fiscal and political exposure	tax base, employment, procurement, strategic reliance, organized beneficiaries	C	Does the governing side acquire a larger stake in uninterrupted deployment?
International dependence	imported AI services, physical inputs, outside options, alternative suppliers	$C, K(x)$	Which states become pivotal, dependent, or expensive members of a response coalition?
Measurement	evaluation coverage, reporting latency, independent verification, model access	O, q	Can institutions observe the relevant state before the response window closes?
Preservation	portability, public alternatives, diversified suppliers, independent expertise, protocols	policy on H, B, C, k	Which investments keep future response options open while current development continues?

Table 4: Illustrative frontier-AI indicator map. Each row requires an application-specific measurement definition before quantitative use.

3.2 Four example forecast forms

The framework supports conditional statements of the following form.

Financial-entrenchment forecast. If long-duration AI infrastructure obligations and cross-financial exposures grow while substitution remains costly, the response-sensitivity channel can increase. Under a maintained positive-affine specification, this raises relevant entries of C or B and can reduce the spectral margin.

Interoperability forecast. If portability and multi-provider substitution reduce the persistence of dependence or the cost effect of inherited concentration, relevant entries of H or C fall. Under the componentwise order conditions of the structural-preservation result, the loop gain weakly falls.

Acceleration forecast. If frontier capability or deployment grows faster while (H, B, C) remain unchanged, the system reaches future states sooner in calendar time. The positive-affine phase boundary changes when the feedback architecture changes rather than from speed alone.

Regime-switch forecast. If regulatory, geopolitical, or market regimes switch, a sequence can create greater exposure than each regime analyzed alone. Fixed-regime stability therefore supplies only one part of the foresight assessment.

3.3 Worked example: AI infrastructure build-out

This example shows the operating procedure at the mechanism-map level. It uses a hypothetical three-year horizon and leaves every quantitative parameter open for measurement.

Decision object. The technological action is additional frontier-AI infrastructure and deployment.

The response of interest is the ability of a public authority or coordinated group of authorities to restrict, redirect, migrate, or substitute a major service after a certified trigger while keeping critical-service and financial losses within a stated tolerance.

State channel	Mechanism to map	Primary object	Early-warning observation
Physical infrastructure	New deployment commits power, sites, networks, and specialized facilities that persist.	B, H	Migration time and stranded-capacity cost rise faster than available substitutes.
Financial exposure	Long-duration obligations and correlated claims connect continued operation to lenders, investors, insurers, or public balance sheets.	B, C	Intervention increasingly creates collateral, refinancing, or balance-sheet stress.
Provider dependence	Workflows, APIs, data, and operational processes become costly to move.	H, C	Recovery time after provider loss rises; portability tests deteriorate.
Human substitution	Organizations reduce parallel human or independent technical capacity.	H, C	Critical tasks take longer to perform or validate during model or provider outages.
Fiscal and political exposure	Employment, tax revenue, procurement, and strategic plans become tied to continued deployment.	$C, K(x)$	The coalition cost of restriction rises even at unchanged technical severity.
Preservation capacity	Interoperability, independent expertise, alternatives, spare capacity, and protocols preserve future choices.	policy on H, B, C, k	Exit time, substitution cost, and coalition friction fall after preservation investment.

Table 5: Mechanism map for the worked AI-infrastructure example. The table defines observables before any scalar feedback estimate is reported.

Conditional forecast card.

- **Target:** three-year change in the cost and speed of an effective migration, restriction, or substitution response.
- **Condition:** infrastructure and financial commitments continue to grow while substitute capacity and portability improve more slowly.
- **Mechanism:** action loads persistent infrastructure and financial states through B; these states persist through H and raise later response cost through C and the coalition threshold $K(x)$.
- **Leading indicators:** migration time, concentration of critical workloads, duration and concentration of financing obligations, outage-recovery performance, retained independent expertise, alternative-provider capacity, and state-specific fiscal exposure.
- **Disconfirming evidence:** scaled deployment accompanied by stable or falling migration time, broad substitutability, low correlated financial exposure, preserved independent capability, and stable response cost after comparable shocks.
- **Update rule:** re-estimate or reclassify the mechanism after a major provider outage, financing shock, interoperability mandate, merger, export restriction, treaty change, or scheduled measurement date.

Preservation portfolio. The first intervention search targets structural optionality: portable workloads, multiple providers, independent public or scientific capability, diversified financing, recovery drills, spare infrastructure, common measurement, and response protocols whose incentives scale with the future source of enforcement cost. The framework then asks whether these interventions reduce the relevant H, B, C, response delay, or coalition-cost channels enough to justify their current resource cost.

The example produces a falsifiable foresight object before a full estimate of L exists. Once transition and response variation identify the required parameters under one common specification, the same state register can move from Level 1 to Level 2 or Level 3 analysis.

4 Interpretation Rules and Failure Modes

The framework has a fallback layer for each structural failure mode. Move to the structural layer that matches the observed system.

Incomparable response sets. Early action can remove some institutional options while creating others. Strict set inclusion then gives no ordering. Section 7 therefore adds a value-based extension using an analyst-specified monotone set functional. The strong set-loss definition remains useful when one continuation set is contained in the other.

Nonlinear saturation and thresholds. A local supercritical slope can coexist with a globally bounded trajectory when feedback saturates. Convex or threshold effects can produce the reverse pattern. Use the nonlinear transition, local Jacobian, and global viability analysis when the affine approximation is weak.

Time scale. The loop matrix is invariant to exact block aggregation only under the stated constant-action and response-timing convention. A different sampling interval can define a different structural system.

Reachability and observability. A mathematical mode matters for the application only when it can be reached by relevant actions or forcing and observed in the response outcome. Weak measurement can hide an important mode; an unreachable mode can make a raw spectral diagnostic overly pessimistic.

Response-side identification. Transition data can often discipline persistence and action loading before they identify how inherited state changes intervention feasibility. A foresight team should therefore treat C as a separate empirical problem rather than infer it from adoption or infrastructure data alone.

Measurement and action. Better evidence changes the observation system. A response protocol changes incentives. Capacity investment changes what action remains feasible. These interventions should be modeled separately.

Forecast discipline. Prospective use requires a frozen state definition, a stated horizon, competing models, leading indicators, and disconfirming evidence. Retrospective fit is a starting point for measurement design; repeated out-of-sample performance is the stronger test.

Part II

Technical Foundations

5 Technical Introduction and Scope

A regulatory problem is usually written as if the object being regulated evolves while the regulator's ability to act remains available. That assumption can fail in exactly the cases where timing matters most.

An early technological change can alter the economy that will later decide whether to restrain it. Firms can reorganize around the technology. Infrastructure can be built for it. Debt can be issued against it. Governments can become dependent on its tax base or strategic value. Consumers can lose substitutes. States can become dependent on foreign suppliers. A tolerated practice can also become the reference point used to judge the next practice. The result is a coupled system: the object of governance changes, and the feasible response set changes with it.

This monograph calls that class of problems *first-meter problems*. The name is meant to emphasize a temporal asymmetry. The importance of the first movement from an institutional line extends beyond its size. Its state transition can be the dominant effect. After the first meter, the second meter may be evaluated against a different benchmark, face a different coalition, impose a different switching cost, or interact with a different financial and infrastructural system.

This idea is related to the Collingridge dilemma: early control is easier when consequences are uncertain, while later control can become difficult after a technology is embedded (Collingridge, 1980; Genus and Stirling, 2018). It is also related to increasing-returns models of technological lock-in (Arthur, 1989), carbon lock-in and techno-institutional complexes (Unruh, 2000), shifting baselines (Pauly, 1995), endogenous status quos (Dziuda and Loeper, 2016; Kennard, 2023), and the broader governability literature, which studies the fit between a governing system and the system to be governed (Bavinck et al., 2013). The contribution here is to place these intuitions inside one dynamic mathematical object and to identify the subclasses in which exact results are available.

The first goal is generality of representation. A model that lists today's mechanisms is fragile. Future mechanisms can appear that are absent from the list. The general state must therefore be an abstract sufficient statistic. At date t, let X_t contain everything from the past that is relevant for future technological evolution and institutional response. In one application it may contain capability, precedent, compute dependence, financial exposure, platform concentration, energy dependence, human skill retention, geopolitical outside options, and historical claims. In another application it can contain a completely different set of variables. The primitive law is simply

$$X_{t+1} \sim P_\theta(\cdot \mid X_t, A_t), \tag{3}$$

with a response rule or intervention hazard that also depends on the state.

The second goal is theorem-level generality. Equation (3) supplies a representation; the theory comes from structural subclasses. The monograph separates six linked layers.

1. A **general stochastic layer** defines risk-constrained capacity, robust structural uncertainty, partial observation, and a formal contract for adding new mechanisms.
2. A **deterministic monotone layer** gives an exact boundary-iteration theorem. Under order-preserving dynamics, the maximal quiet path is generated by one operator.

3. A **positive affine layer** gives closed-form finite-horizon expressions, a coordinate-invariant feedback criterion, phase, margin, amplification, critical-edge classification, and switched-system extensions.
4. A **strategic and welfare layer** endogenizes the technological actor's anticipation of future governability and the institution's investment in preserving future policy options.
5. An **empirical layer** states identification requirements, gives post-identification uncertainty propagation, maps historical mechanisms, and specifies prospective falsification tests.
6. A **frontier-AI specialization** recovers benchmark drift, endogenous coalition costs, financial and organizational entrenchment, historical participation, and strategic preservation as one transparent case.

The central finite-dimensional object is a positive feedback loop. Suppose a vector of technological choices $a_t \in \mathbb{R}^m_+$ drives a state $x_t \in \mathbb{R}^d_+$ according to

$$x_t = Hx_{t-1} + Ba_t + d, \tag{4}$$

and the largest action that remains on the quiet side of the institutional response boundary is

$$a_t \leq k + Cx_{t-1}. \tag{5}$$

Boundary behavior is then generated by

$$M := H + BC, \qquad g := d + Bk. \tag{6}$$

If $\rho(H) < 1$, define the loop-gain matrix

$$L := C(I - H)^{-1}B. \tag{7}$$

A standard positive-systems small-gain structure then gives

$$\rho(M) < 1 \quad \Longleftrightarrow \quad \rho(L) < 1. \tag{8}$$

When this holds, the exact steady boundary action is

$$a_\infty = (I - L)^{-1}\bar{k}, \qquad \bar{k} := k + C(I - H)^{-1}d. \tag{9}$$

The matrix L is a useful structural object because $(I - H)^{-1} = I + H + H^2 + \cdots$ accumulates every future propagation path through the state system. Thus L measures how one unit of technological action feeds through all persistent state channels and returns as additional future room for action.

The spectral condition is one part of the story. A stable but nonnormal M can have very large finite-horizon amplification. At the critical boundary, polynomial growth can depend on Jordan structure. If the environment switches among regimes, stability can fail even when every regime is stable in isolation; the joint spectral radius is then the relevant worst-case object. These edge cases have direct governance interpretations. They correspond to governance systems that look stable asymptotically while becoming temporarily difficult to control, systems that accumulate through chained critical dependencies, and systems whose risk comes from switching among individually benign institutional regimes.

The frontier-AI specialization makes the general structure concrete. Let y_t be a cumulative AI-relevant state, m_t an adaptive public benchmark, and $x_{j,t}$ economic-entrenchment channels. Let η be the institution's retained weight on the original anchor and let γ_j measure how much long-run

entrenchment channel j raises the minimum cost of a winning response coalition. With $\Gamma = \sum_j \gamma_j$, the scalar loop gain becomes

$$L = 1 - \eta + \Gamma. \tag{10}$$

Hence

$$\Gamma < \eta \tag{11}$$

is exactly the stable condition, and the long-run quiet ceiling is

$$Q_\infty = \frac{K_0}{\eta - \Gamma}. \tag{12}$$

The benchmark-only result is the zero-entrenchment case $\Gamma = 0$.

The main contribution is a reusable mathematical interface for conditional analysis across mechanisms. A new mechanism matters only after one specifies what state it adds, how that state evolves, how it changes the response technology, and what data observe it. If those choices satisfy stronger structural assumptions, stronger theorems become available. This gives the framework a controlled form of extensibility in which each new mechanism states its own mathematical structure.

Reader's map

The document has four practical reading paths. Foresight and policy teams can begin with Part I, then read the frontier-AI application, measurement section, and policy-design section. Economics and game-theory readers can emphasize the coalition, strategic-entrenchment, participation, welfare, and empirical sections. Control and dynamical-systems readers can emphasize the monotone, positive-affine, transient, switching, continuous-time, and nonlinear sections. Readers auditing the formal claims can use Part III, which contains proofs, edge cases, and the numerical audit.

Layer	Maintained structure	Exact output	Empirical status
General stochastic	Measurable state, transition, response, and observation objects	Risk-constrained capacity, robust and partially observed formulations	Application primitives required
Monotone	Order-preserving transition and quiet boundary	Maximal quiet path from one boundary operator	Structural assumption testable in an application
Positive affine	$H, B, C \geq 0$, $\rho(H) < 1$	Loop gain, phase, steady boundary, margin, transient and switching diagnostics	H, B, C require transition and response-side identification
Strategic	Ordered continuation values; closed-form quadratic benchmark for the two-sided game	Strategic acceleration/preservation comparisons and one exact Stackelberg benchmark	Theoretical benchmark; parameters require application data
Empirical	Observed or identified latent state plus exogenous or instrumented variation	Estimated or bounded L, uncertainty propagation, predictive tests	Application-specific research design
Nonlinear and robust	Specified nonlinear transition and institutional control timing	Local Jacobian tests and finite-horizon viability sets	Numerical model and convergence checks required

Table 6: Claim and evidence map. Each stronger conclusion uses the structural assumptions stated in its row.

6 Relation to the Literature and Scope of the Claim

6.1 Governability, technology control, and path dependence

The term *governability* has an established literature. Interactive-governance work treats governability as a property of the system to be governed, the governing system, and the interaction between them (Bavinck et al., 2013). This monograph uses the term in a narrower dynamic sense: the future feasibility and effectiveness of institutional response can itself be state-dependent.

The Collingridge dilemma is a close qualitative antecedent. Early in a technological trajectory, consequences are hard to know; later, once consequences are clearer, control can be difficult because the technology has become embedded (Collingridge, 1980; Genus and Stirling, 2018). Increasing-returns models show how adoption can lock in technological paths (Arthur, 1989). Carbon-lock-in theory emphasizes a coupled technological and institutional complex (Unruh, 2000). The framework formalizes one part of this control problem: the state transition induced by current technological movement can change the future intervention correspondence or intervention cost.

History can also move the reference point used for later choice. Shifting-baseline work gives an ecological example (Pauly, 1995); repeated coordination can make precedent focal (Crawford and Haller, 1990); and dynamic bargaining can make previous outcomes endogenous status quos (Dziuda and Loeper, 2016; Kennard, 2023). The frontier-AI specialization combines such a benchmark state with endogenous response costs.

Policy-feedback research provides another close antecedent. It studies how existing policies reshape later politics and policy development (Béland, Campbell, and Weaver, 2022), including effects on administrative capacity and organizational conditions (Moynihan and Soss, 2014). Policy-capacity research separately studies the analytical, operational, and political capabilities needed for effective policy (Wu, Ramesh, and Howlett, 2015). Recent sustainability work explicitly combines technology deployment with policy feedback and documents co-evolution between a technological innovation system and its policy environment (Dehler-Holland, 2025). The present framework focuses on a specific feedback direction within that broader family: technological movement changes the state variables that determine the future response set, and this channel is represented by a transition law and, under positive-affine structure, by a loop-gain operator.

6.2 Viability, constrained control, and stochastic games

Viability theory studies whether a controlled evolution can remain inside a constraint set and defines the viability kernel as the states from which at least one viable trajectory exists (Aubin, 2009; Aubin, Bayen, and Saint-Pierre, 2011). This is an important mathematical parent of the general framework. Constrained Markov decision processes provide a probabilistic analogue when policies optimize one objective subject to constraints on other performance measures (Altman, 1999). Measure-theoretic foundations for discrete-time stochastic control are given by Bertsekas and Shreve (1978). Robust dynamic programming treats ambiguity in transition laws, with rectangular ambiguity sets giving a time-consistent Bellman structure (Iyengar, 2005; Nilim and El Ghaoui, 2005). Stochastic games provide the natural extension when both the technological actor and governing institutions are strategic (Shapley, 1953).

These mathematical tools are established. The framework uses them to define a governance problem in which the constraint set and response technology can themselves move as a consequence of earlier

action.

6.3 Monotone and positive systems

Order-preserving dynamical systems provide global comparison tools (Smith, 1995). Positive linear systems give especially sharp results because nonnegative states and nonnegative couplings allow Perron–Frobenius and M-matrix arguments (Farina and Rinaldi, 2000). The loop-gain criterion used below is a standard positive-systems small-gain structure. The contribution here is the mapping from dynamic governability into that structure and the exact first-meter specialization.

The spectral radius describes asymptotic linear growth, but it can hide large transient amplification for nonnormal matrices. Pseudospectral analysis makes this distinction explicit (Trefethen and Embree, 2005). If a system switches among several matrices, arbitrary-switching growth is governed by the joint spectral radius (Berger and Wang, 1992; Jungers, 2009). Recent economic working papers also use spectral-radius thresholds and Perron weights to study self-reinforcing economic feedback and targeted intervention (Yamaguchi, 2026; Rajhi, 2026); concurrent work on leveraged exchange-traded products uses a loop-gain matrix and its spectral radius to measure coupled market feedback (Woo, 2026). These antecedents make the spectral mathematics itself part of the established toolkit. The contribution here lies in the response-feasibility construction $C(I - H)^{-1}B$ and its governance interpretation. The same mathematics also shows why a system can be asymptotically stable yet exhibit a large finite-time loss of control, or can be unstable under switching even when every fixed regime is individually stable.

6.4 Collective action and equilibrium selection

The frontier-AI response microfoundation is a costly coalition-enforcement game. Exact potential games (Monderer and Shapley, 1996), log-linear learning (Blume, 1993; Marden and Shamma, 2012), and Harsanyi–Selten risk dominance (Harsanyi and Selten, 1988) provide equilibrium-selection tools. Global AI governance is itself a collective-action problem because major states differ in interests, capabilities, and willingness to bear enforcement costs (Roberts et al., 2024; Olugbade, 2025).

Information structure matters separately from coalition cost. Global games study equilibrium selection when actors observe noisy private or public signals about a common state (Carlsson and van Damme, 1993; Frankel, Morris, and Pauzner, 2003; Morris and Shin, 2003). Public information can restore or preserve multiplicity under some relative-precision conditions (Hellwig, 2002; Angeletos, Hellwig, and Pavan, 2007). The framework uses these results to separate measurement quality, common knowledge, and the incentive rule that maps a certified state into coordinated response.

6.5 State capacity, irreversibility, and strategic regulation

The welfare and strategic extensions build on established work in state capacity, endogenous institutions, dynamic regulation, irreversibility, and strategic investment. The state-capacity literature explicitly models legal and fiscal capacities as forward-looking investments that constrain later policy choices (Besley and Persson, 2009). Endogenous-institutional-change theory studies self-reinforcing processes through which institutions persist or undermine themselves (Greif and Laitin, 2004). Dynamic regulation models repeated capture and the institutional constraints that shape it (Martimort, 1999). Recent work also models state capacity and societal power as jointly

evolving stocks created by strategic investment (Colonescu, 2026). The irreversibility literature shows that preserving an option can have value when decisions are difficult to reverse and future information matters (Arrow and Fisher, 1974; Henry, 1974; Dixit and Pindyck, 1994). In regulation, firms can invest strategically because current investment changes the government's later preferred policy (Glazer and Janeba, 2004; Tarui and Polasky, 2005), and voluntary restraint can preempt political intervention (Maxwell, Lyon, and Hackett, 2000).

These literatures establish endogenous institutional capacity, dynamic capture, strategic regulatory investment, and option value as prior ideas. The narrower contribution here is a technology-induced map from current action through persistent state into the future response boundary; its positive-system representation $L = C(I - H)^{-1}B$; and one interface that carries that response-feasibility mechanism from stochastic viability through phase, strategic, welfare, and empirical analysis.

6.6 Frontier AI as infrastructure and economic dependence

Compute is a useful governance input because it is physical, quantifiable, excludable, and produced through a concentrated supply chain (Sastry et al., 2024). Recent work treats AI as infrastructure as well as a discrete tool (Naser, 2026). Agent-mediated platforms can deepen incumbent lock-in (Kapoor, Kolt, and Lazar, 2025); recent position papers discuss dependence on AI at individual, social, and national levels (Kim et al., 2026); and other work adapts lock-in arguments to compute infrastructure (MacNeill, 2026). A recent economic model of digital-intelligence capital derives a structural Jevons effect and data flywheels that can generate concentration (Zhang and Zhang, 2026).

The physical and financial build-out gives these mechanisms an empirical object. The IEA projects global data-center electricity consumption to reach about 945 TWh by 2030 in its base case, with AI as the largest driver of the increase (IEA, 2025). The BIS reports rapidly rising borrowing for AI infrastructure and growing links between hyperscalers, private-credit vehicles, insurers, and banks (Eren, Krohn, and Todorov, 2026). The Bank of England reports that AI-related use of public debt, private credit, leveraged finance, and structured finance accelerated in the first half of 2026, while reporting that other firms and governments retained broad access to funding markets at the time of the report (Bank of England, 2026). These facts motivate state variables. Identification of the entrenchment parameters requires separate empirical work.

6.7 Novelty claim

The novelty claim is deliberately limited. Governability, policy capacity, policy feedback, state capacity, viability, lock-in, shifting baselines, positive feedback, spectral-radius thresholds in economic feedback systems, small gain, stochastic games, strategic regulatory investment, irreversibility, option value, and joint spectral radius all have established literatures. The claimed contribution is the following synthesis and the results that are specific to it:

1. a general definition of a first-meter problem as a state transition that changes future response feasibility or future quiet capacity;
2. a plug-in contract that represents an unknown mechanism through extensions of the state, transition, response, and observation maps;
3. an exact boundary-operator result for deterministic monotone first-meter systems;
4. a positive-affine reduction in which dynamic governability is governed by the loop-gain matrix $L = C(I - H)^{-1}B$, together with exact steady-boundary formulas, coordinate and constant-input

temporal-aggregation invariance, and a reachable-observable Jordan classification;
5. a separation of phase $\rho(L)$, spectral margin $1-\rho(L)$, Perron-direction steady amplification, nonnormal transient amplification, and switching risk;
6. a state-dependent coalition-enforcement microfoundation and a frontier-AI specialization that yields the exact $\Gamma < \eta$ phase boundary and $K_0/(\eta - \Gamma)$ ceiling;
7. an order-theoretic strategic first-meter result and a closed-form policy-first game in which the actor's value of future quiet room can induce both greater technological movement and greater institutional preservation investment;
8. a counterfactual welfare definition for the option value of preserved governability and a structural interpretation of redundancy, interoperability, and independent capacity as investments in future policy flexibility;
9. a post-identification differential and Perron-sensitivity map from (H, B, C) to L and its Perron root, paired with explicit identification conditions and a falsification protocol;
10. nonlinear and robust extensions that show where the linear phase condition remains informative and where saturation, hysteresis, switching, or structural uncertainty require different objects.

7 The General Dynamic-Governability Model

7.1 State, technological action, and institutional response

Let X be a standard Borel state space. The state $X_t \in \mathsf{X}$ is a sufficient statistic for the future. No finite dimension is imposed at this level. It can be a finite vector, a graph, a probability measure, a function, or a belief state.

Let A be a standard Borel action space and let $A(x) \subseteq \mathsf{A}$ be the nonempty feasible technological-action correspondence at state x. A nonanticipative technological policy is denoted by π.

To separate technological evolution from institutional response, let

$$P_\theta(dx' \mid x, a) \tag{13}$$

be the pre-response transition kernel. Let

$$q_\theta(x, a) \in [0, 1] \tag{14}$$

be the conditional probability that an effective institutional response occurs at the current state-action pair, conditional on survival without an earlier effective response. The parameter θ collects structural parameters.

Introduce an absorbing response state $\dagger$. The combined kernel on $\bar{\mathsf{X}} := \mathsf{X} \cup \{\dagger\}$ is

$$\bar{P}_\theta(\{\dagger\} \mid x, a) = q_\theta(x, a), \tag{15}$$

$$\bar{P}_\theta(B \mid x, a) = [1 - q_\theta(x, a)] P_\theta(B \mid x, a), \qquad B \subseteq \mathsf{X}, \tag{16}$$

$$\bar{P}_\theta(\{\dagger\} \mid \dagger, a) = 1. \tag{17}$$

Define the intervention time

$$\tau := \inf\{t \geq 0 : X_t = \dagger\}. \tag{18}$$

Let $Y : \bar{\mathsf{X}} \to \mathbb{R}_+$ be a measurable progress functional with $Y(\dagger) = 0$. In the AI application, Y can be a capability or deployment index. In another domain it can be any monotone quantity whose accumulation is relevant.

Definition 1 (Risk-constrained capacity)**.** *For horizon T and risk tolerance $\delta \in [0,1)$, define*

$$Q_T^{\delta}(x) := \sup_{\pi} \left\{ \mathbb{E}_x^{\pi}[Y(X_T)] : \mathbb{P}_x^{\pi}(\tau \leq T) \leq \delta \right\}. \tag{19}$$

Use the convention $\sup \varnothing = -\infty$. Applied models can include a reference action, such as inaction, that guarantees a nonempty feasible policy class.

This is a constrained-control object. It measures how much terminal progress is compatible with a bound on cumulative intervention probability. It handles the repeated-hazard problem automatically: many individually small response probabilities can violate the horizon-level risk budget.

Remark 1 (Post-response states)**.** *The absorbing construction measures progress that remains on the pre-response trajectory, with $Y(\dagger) = 0$ as a normalization for that event state. Applications in which intervention leaves residual output or creates a new post-response trajectory can replace $\dagger$ by an explicit post-response state space and define Y on that space. The risk constraint and state-dependent response structure remain unchanged.*

Remark 2 (Stagewise quietness)**.** *A simpler stagewise constraint $q_\theta(X_t, A_t) \leq \rho$ for every t is sometimes useful. The horizon-level constraint additionally controls cumulative hazard. The zero-noise limit is recovered when $q_\theta \in \{0,1\}$ and $\delta < 1$: every admissible path must remain in the zero-response region at every date.*

7.2 Response correspondences and strategic institutions

The reduced-form hazard q_θ can be generated by a richer institutional game. Let $U(x,a)$ be the institution's feasible response-action set and let

$$e(x,a,u) \in [0,1] \tag{20}$$

measure response effectiveness. For an effectiveness standard $\bar{e}$ and a resource budget $\bar{c}$, define the one-period governability correspondence

$$\mathcal{G}(x,a) := \{u \in U(x,a) : e(x,a,u) \geq \bar{e},\ c(x,a,u) \leq \bar{c}\}. \tag{21}$$

A strategic equilibrium selection, insurance rule, treaty protocol, or administrative process can map this correspondence into the reduced-form hazard q_θ.

If governments are forward-looking, continuation values can be included in their stage payoffs and the response rule can be generated by a Markov-perfect or other dynamic equilibrium. The general framework requires only the resulting measurable response kernel or correspondence. This avoids assuming that governments ignore the future effect of acquiescence on later governability.

7.3 A general definition of a first-meter trap

A first-meter problem should survive changes in coordinates and economic interpretation. The following definition uses future response feasibility directly.

Fix an ambient space Π_G of admissible nonanticipative institutional continuation policies. Let $\mathcal{G}_H(x) \subseteq \Pi_G$ denote the policies that, starting from x, can achieve the chosen effectiveness standard within H periods while respecting the institution's resource and legal constraints. The exact construction of $\mathcal{G}_H$ depends on the application. Using the same ambient policy space makes the set comparison below well defined.

Definition 2 (Dynamic first-meter trap)**.** *Consider a current state x and two current technological actions a and a^0, where a^0 is a reference action. In a deterministic model with next states $F(x,a)$ and $F(x,a^0)$:*

(i) action a creates a strong H-horizon first-meter loss of governability *if*

$$\mathcal{G}_{H-1}(F(x,a)) \subsetneq \mathcal{G}_{H-1}(F(x,a^0)); \tag{22}$$

(ii) let $\Psi_H : 2^{\Pi_G} \to \overline{\mathbb{R}}$ be any application-specified functional that is monotone under set inclusion. Action a creates a Ψ_H-valued first-meter loss *if*

$$\Psi_H(\mathcal{G}_{H-1}(F(x,a))) < \Psi_H\Big(\mathcal{G}_{H-1}(F(x,a^0))\Big). \tag{23}$$

In a reduced-form capacity model, action a creates an H-horizon first-meter expansion if it strictly increases the future risk-constrained capacity available to the technological actor after adjusting for the remaining risk budget.

The strong set form says that an early move removes future institutional options. The value form also handles incomparable continuation sets, such as a transition that destroys one response option while creating another. Strong set loss implies weak loss for every inclusion-monotone Ψ_H, and it implies strict loss whenever Ψ_H is strictly monotone at the compared pair. The reduced-form capacity statement is the technological-side description in models where the response boundary determines quiet feasibility. The analyst must state Ψ_H when institutional options require a substantive ranking rather than pure set inclusion.

7.4 The plug-in contract

The framework is intended to admit future mechanisms through a stable formal interface. A new mechanism therefore needs a formal interface.

Definition 3 (Admissible mechanism extension)**.** *Let a base model be*

$$\mathfrak{M} = (\mathsf{X}, \mathsf{A}, A, P, q, Y, O), \tag{24}$$

where O is an observation kernel. An admissible extension by mechanism j consists of:

(i) an added standard Borel state space Z_j;
(ii) an extended transition kernel P' on $\mathsf{X} \times \mathsf{Z}_j$ that can depend on the full current state and action;
(iii) an extended response rule q' or response correspondence $\mathcal{G}'$;
(iv) an extended observation map or kernel O';
(v) a null parameter value under which projection onto X recovers the base model.

This definition deliberately allows interaction terms. A financial mechanism can depend on platform concentration; a geopolitical mechanism can depend on energy infrastructure; a new model architecture can alter both technological progress and switching costs. Additive and separable extensions are special cases.

The practical requirement is strict: a proposed mechanism specifies the state it adds, its transition law, its effect on response, and the observables that identify or constrain its parameters. A verbal scenario remains a motivation until these objects are supplied.

7.5 History dependence and Markovization

Some future mechanisms may have long memory, changing dimension, or path-dependent effects. The general state can absorb that history.

Proposition 1 (History-state Markovization)**.** *Let the primitive state and action spaces be standard Borel, and let a discrete-time controlled process have a history-dependent transition law*

$$P_t(dx_{t+1} \mid x_{0:t}, a_{0:t}). \tag{25}$$

Define the history state

$$H_t := (x_0, a_0, x_1, a_1, \ldots, a_{t-1}, x_t). \tag{26}$$

The disjoint union of all finite history spaces is standard Borel, and the history process is Markov on that space. Hence every such discrete-time history-dependent mechanism can be represented inside (3) *by state augmentation.*

The proposition gives a formal route for hysteresis, delayed effects, learning histories, changing network structure, and endogenous populations. In many applications a lower-dimensional sufficient statistic exists and is preferable for estimation. The history-state construction supplies the general fallback representation. Time-inhomogeneous systems are covered by including calendar time in the state.

7.6 Partial observation and belief states

Suppose institutions observe R_t through

$$R_t \sim O_\theta(\cdot \mid X_t). \tag{27}$$

When X_t is hidden, the posterior belief

$$\mu_t(\cdot) := \mathbb{P}(X_t \in \cdot \mid R_{0:t}, A_{0:t-1}) \tag{28}$$

can be used as the Markov state. The state space then becomes a space of probability measures. Better measurement changes the observation kernel and therefore belief dynamics. A response protocol changes q_θ or $\mathcal{G}$. These are distinct operations.

7.7 Structural uncertainty

Ordinary shocks are uncertainty inside a known transition law. Structural uncertainty means the transition law itself is uncertain. Let

$$P_\theta(\cdot \mid x, a) \in \mathfrak{P}(x, a) \tag{29}$$

for an ambiguity correspondence $\mathfrak{P}$. Robust dynamic programming is time-consistent under standard rectangularity conditions (Iyengar, 2005; Nilim and El Ghaoui, 2005). This gives a formal treatment of structurally unidentified future mechanisms: governance can be designed for an ambiguity family of transition laws.

One conservative risk-feasible policy class is

$$\Pi^{\delta}_{\text{rob}} := \left\{ \pi : \sup_{P \in \mathfrak{P}} \mathbb{P}^{\pi}_{P}(\tau \leq T) \leq \delta \right\}. \tag{30}$$

One can then report both guaranteed capacity

$$\sup_{\pi \in \Pi^{\delta}_{\text{rob}}} \inf_{P \in \mathfrak{P}} \mathbb{E}^{\pi}_{P}[Y(X_T)] \tag{31}$$

and worst-case exposure

$$\sup_{\pi \in \Pi^{\delta}_{\text{rob}}} \sup_{P \in \mathfrak{P}} \mathbb{E}^{\pi}_{P}[Y(X_T)]. \tag{32}$$

They answer separate questions and should be reported separately.

8 Deterministic Monotone First-Meter Systems

The abstract model is intentionally broad. Exact global results require structure. This section imposes order.

Let $\mathsf{X} \subseteq \mathbb{R}^d_+$ and let technological action be $a_t \in \mathbb{R}^m_+$. Use componentwise order.

Assumption 1 (Monotone quiet system)**.** *There exist continuous maps*

$$F : \mathsf{X} \times \mathbb{R}^m_+ \to \mathsf{X}, \qquad B : \mathsf{X} \to \mathbb{R}^m_+, \qquad Y : \mathsf{X} \to \mathbb{R}_+, \tag{33}$$

with the following properties:

(i) F is order-preserving in both arguments;
(ii) B is order-preserving;
(iii) Y is order-preserving;
(iv) a weakly quiet action at x satisfies $0 \leq a \leq B(x)$ componentwise.

Define the boundary operator

$$\mathcal{T}(x) := F(x, B(x)). \tag{34}$$

Theorem 1 (Boundary domination)**.** *Under Assumption 1, let x_0 be fixed. Define the boundary path*

$$x^*_{t+1} = \mathcal{T}(x^*_t), \qquad x^*_0 = x_0. \tag{35}$$

Then every weakly quiet path satisfies

$$x_t \leq x^*_t \tag{36}$$

for every t. Consequently,

$$\max Y(x_T) = Y(\mathcal{T}^T(x_0)) \tag{37}$$

over weakly quiet paths.

Define radial strict quietness by requiring, at every date, some $\varepsilon_t \in (0,1)$ such that $0 \leq a_t \leq (1-\varepsilon_t)B(x_t)$. If $(1-\varepsilon)B(x)$ is admissible for every $\varepsilon \in (0,1)$, then the radially strict problem has the same supremum as (37).

The proof is an induction. Its importance is conceptual: under monotonicity, every detailed first-meter problem collapses to repeated application of one operator. The unknown future mechanism changes F or B; the theorem survives if the extension preserves order.

Definition 4 (Boundary expansion)**.** *A current action a at state x is one-step entrenching relative to a^0 if*

$$B(F(x,a)) \geq B(F(x,a^0)) \tag{38}$$

componentwise, with a strict inequality in at least one component.

Boundary expansion is one concrete first-meter mechanism: current movement increases the next period's maximal quiet action.

8.1 Fixed points and nonlinear edge cases

Suppose x^* is a fixed point of $\mathcal{T}$.

Proposition 2 (Local nonlinear phase test)**.** *Assume $\mathcal{T}$ is continuously differentiable near x^*. Then:*

(i) if $\rho(D\mathcal{T}(x^*)) < 1$, the fixed point is locally asymptotically stable;
(ii) if $\rho(D\mathcal{T}(x^*)) > 1$, the fixed point is unstable;
(iii) if $\rho(D\mathcal{T}(x^*)) = 1$, the linearization is inconclusive in general.

The critical case must remain open at this level. For example, scalar maps with derivative one at the origin can converge slowly, diverge slowly, or remain neutral depending on higher-order terms. A general theory that forced the critical case into one growth class would therefore be false.

A stronger global sufficient condition is contraction. If $\mathcal{T}$ maps a complete invariant set into itself and is globally Lipschitz with constant $L_T < 1$ in some complete norm, then Banach's fixed-point theorem gives a unique fixed point and geometric convergence. This is useful but more restrictive than monotonicity.

9 Positive Affine Systems and the Governability Loop

This section gives the main tractable subclass. It is broad enough to contain interacting benchmark, infrastructure, financial, organizational, geopolitical, and institutional state variables, while retaining exact formulas.

9.1 Vector actions and vector state

Let

$$x_t \in \mathbb{R}^d_+, \qquad a_t \in \mathbb{R}^m_+. \tag{39}$$

Assume

$$x_t = Hx_{t-1} + Ba_t + d, \tag{40}$$
$$a_t \leq k + Cx_{t-1}, \tag{41}$$

where

$$H \in \mathbb{R}_+^{d\times d}, \quad B \in \mathbb{R}_+^{d\times m}, \quad C \in \mathbb{R}_+^{m\times d}, \quad d \in \mathbb{R}_+^{d}, \quad k \in \mathbb{R}_+^{m}. \tag{42}$$

The vector a_t can represent several simultaneously constrained technological dimensions. The scalar first-meter model is $m = 1$.

In this layer, L is an input–output feedback object for the technological quiet boundary. The governability interpretation applies when the application links an expansion of that quiet continuation set to a contraction or higher cost of the institution's effective response set, as in Definition 2 and the coalition model below.

Boundary saturation gives

$$M := H + BC, \qquad g := d + Bk, \tag{43}$$

and

$$x_t^* = Mx_{t-1}^* + g. \tag{44}$$

Theorem 2 (Exact finite-horizon boundary path)**.** *For every $T \geq 1$,*

$$x_T^* = M^T x_0 + \sum_{r=0}^{T-1} M^r g. \tag{45}$$

The exact maximal boundary action at date $T+1$ is

$$a_{T+1}^* = k + C\left(M^T x_0 + \sum_{r=0}^{T-1} M^r g\right). \tag{46}$$

For every nonnegative weight vector $w \in \mathbb{R}_+^m$, $w^\top a_{T+1}^*$ is the maximum of $w^\top a_{T+1}$ over weakly quiet paths and the supremum over strict quiet paths under the interior condition of Theorem 1.

9.2 The loop-gain matrix

Assume now that the state subsystem is stable in the absence of feedback:

$$\rho(H) < 1. \tag{47}$$

Then $(I-H)^{-1} = \sum_{r\geq 0} H^r \geq 0$. Define

$$L := C(I-H)^{-1}B \in \mathbb{R}_+^{m\times m}, \tag{48}$$

and

$$\bar{k} := k + C(I-H)^{-1}d. \tag{49}$$

The entry L_{ij} is the total long-run effect of one unit of action dimension j on future boundary dimension i after the effect has propagated through every state path generated by H.

Theorem 3 (Positive small-gain criterion)**.** *Under (47), with $H, B, C \geq 0$,*

$$\boxed{\rho(H+BC) < 1 \quad \Longleftrightarrow \quad \rho\Big(C(I-H)^{-1}B\Big) < 1.} \tag{50}$$

If these equivalent conditions hold, the boundary system has the unique steady state

$$a_\infty = (I-L)^{-1}\bar{k}, \tag{51}$$

$$x_\infty = (I-H)^{-1}(d + Ba_\infty). \tag{52}$$

This is the central generalized phase condition. The scalar quantity $\Gamma - \eta$ that appears later is only one coordinate representation of this loop.

9.3 Coordinate invariance

A useful structural quantity is invariant under a change of state coordinates.

Proposition 3 (State-coordinate invariance)**.** *Let S be any invertible state transformation and define*

$$H' = SHS^{-1}, \qquad B' = SB, \qquad C' = CS^{-1}. \tag{53}$$

Then

$$C'(I - H')^{-1}B' = C(I - H)^{-1}B = L. \tag{54}$$

Hence the loop-gain matrix and its spectral radius are invariant to invertible state-coordinate changes. Under a general S, the positive cone is transported to $S\mathbb{R}^d_+$; positivity in the standard componentwise cone is preserved by cone-preserving transformations. If action coordinates are also changed by an invertible R, then the new loop matrix is RLR^{-1} and has the same spectral radius.

This is one reason to prefer $\rho(L)$ to a sum of parameters as the general phase object.

Proposition 4 (Exact state-subsystem aggregation under constant block action)**.** *Assume $\rho(H) < 1$ and fix an integer $h \geq 1$. If one action vector is held constant for h micro-periods, define the exact h-step state matrices*

$$H_h := H^h, \qquad B_h := \sum_{r=0}^{h-1} H^r B. \tag{55}$$

Then

$$C(I - H_h)^{-1}B_h = C(I - H)^{-1}B = L. \tag{56}$$

Thus exact aggregation of the state subsystem preserves the loop-gain matrix when one action vector is held constant inside each block and the same response loading C is evaluated at block boundaries. Within-block action variation, a different response timing convention, or a different state definition specifies a different structural system and therefore a different loop matrix.

9.4 Complete finite-dimensional asymptotic classification

Linear edge cases also include critical and transient structures. Affine forcing and Jordan structure can generate constant, polynomial, exponential-polynomial, and oscillatory terms.

Define the augmented matrix

$$\widehat{M} := \begin{pmatrix} M & g \\ 0 & 1 \end{pmatrix}, \qquad \widehat{x}_0 := \begin{pmatrix} x_0 \\ 1 \end{pmatrix}. \tag{57}$$

Then

$$\begin{pmatrix} x_T^* \\ 1 \end{pmatrix} = \widehat{M}^T \widehat{x}_0. \tag{58}$$

Any affine observable $Z_T = q^\top x_T^* + r$ can be written

$$Z_T = \widehat{q}^\top \widehat{M}^T \widehat{x}_0, \qquad \widehat{q} := \begin{pmatrix} q \\ r \end{pmatrix}. \tag{59}$$

Theorem 4 (Jordan classification of finite-dimensional edge cases)**.** *For a finite-dimensional affine boundary system, every scalar affine observable Z_T is the sum of two parts. For each nonzero eigenvalue $\lambda \in \sigma(\widehat{M})$ it has a term*

$$p_\lambda(T)\lambda^T, \tag{60}$$

where p_λ is a polynomial whose degree is at most one less than the size of the largest Jordan block for λ. Jordan blocks at $\lambda = 0$ contribute only finite-support transients and vanish once T exceeds the largest such block size. Only generalized eigenspaces reached by $\widehat{x}_0$ and observed by $\widehat{q}$ contribute nonzero coefficients.

Consequently, after zero coefficients are removed, the asymptotic envelope is determined by the largest modulus $|\lambda|$ among reachable-observable nonzero terms and, at that modulus, by the largest surviving polynomial degree. Critical systems can therefore display constant, linear, quadratic, or higher polynomial growth; supercritical systems can display $T^r\rho^T$ growth; and several dominant complex phases can generate bounded modulation or oscillation.

Theorem 4 is a linear-algebra fact, but it is important for model discipline. A statement that every critical first-meter system grows linearly would be false at this level of generality.

9.5 A clean positive special case

When M is primitive, the Perron root is simple and strictly dominant. If $\rho(M) > 1$ and the input and observable are nonzero and nonnegative, the Perron mode is both reached and observed under mild irreducibility conditions, giving pure geometric growth up to a positive constant. At $\rho(M) = 1$, affine forcing commonly creates a size-two Jordan chain in the augmented system and therefore linear growth. The benchmark-entrenchment model below has this clean rank-one structure.

9.6 Transient governability

Asymptotic stability and finite-horizon amplification are distinct properties. Define

$$\mathcal{A}_T(M; \|\cdot\|) := \max_{0 \le t \le T} \|M^t\|. \tag{61}$$

A nonnormal matrix can have $\rho(M) < 1$ while $\mathcal{A}_T$ is very large (Trefethen and Embree, 2005).

Example 1 (Stable but large transient amplification)**.** *For $0 < r < 1$ and $L_0 > 0$, let*

$$M = \begin{pmatrix} r & L_0 \\ 0 & r \end{pmatrix}. \tag{62}$$

Then $\rho(M) = r < 1$, but

$$M^t = \begin{pmatrix} r^t & tL_0 r^{t-1} \\ 0 & r^t \end{pmatrix}. \tag{63}$$

The off-diagonal term can be arbitrarily large at intermediate horizons as L_0 grows, even though $M^t \to 0$.

Thus a governance system can be asymptotically bounded and still pass through a finite period in which response becomes very difficult. A complete empirical analysis should report both long-run feedback gain and finite-horizon amplification.

For an affine forcing direction $g \geq 0$ and a nonnegative output vector q, define the output-specific transient exposure

$$\mathcal{G}_T(q, g; M) := \max_{0 \leq t \leq T} q^\top \left(\sum_{r=0}^{t-1} M^r \right) g, \tag{64}$$

with the empty sum equal to zero. This quantity records finite-horizon forced exposure in the output that the application actually values.

10 Spectral Margin, Steady Amplification, and Comparative Statics

The phase classification $\rho(L) < 1$, $\rho(L) = 1$, and $\rho(L) > 1$ is deliberately coarse. Two systems can both be subcritical while having very different distances from the boundary and very different responses to persistent forcing. This section therefore separates three objects: phase, margin, and amplification.

Definition 5 (Spectral governability margin)**.** *For a positive-affine system satisfying $\rho(H) < 1$, define*

$$\mathfrak{m}(L) := 1 - \rho(L). \tag{65}$$

The system is subcritical when $\mathfrak{m} > 0$, critical when $\mathfrak{m} = 0$, and supercritical when $\mathfrak{m} < 0$. The quantity $\mathfrak{m}$ is the spectral gap to the positive-affine phase boundary. Perturbation robustness is assessed with the sensitivity, resolvent, and transient objects below.

When $\rho(L) < 1$, define the *steady boundary amplification operator*

$$\mathcal{A}_G(L) := (I - L)^{-1} = I + L + L^2 + \cdots . \tag{66}$$

The Neumann-series representation gives the direct interpretation. The first term is baseline room for action, L is the first return through persistent state channels, L^2 is the return generated by that first return, and so on. The operator measures repeated endogenous expansion of the steady quiet boundary within the stable positive-affine model.

Proposition 5 (Perron-direction amplification)**.** *Let $L \geq 0$ be irreducible with $\rho(L) < 1$, and let $v \gg 0$ satisfy $Lv = \rho(L)v$. Define the weighted sup norm*

$$\|z\|_v := \max_i \frac{|z_i|}{v_i}. \tag{67}$$

Then

$$\|L\|_v = \rho(L), \qquad \|(I - L)^{-1}\|_v = \frac{1}{1 - \rho(L)} = \frac{1}{\mathfrak{m}(L)}, \tag{68}$$

and

$$(I - L)^{-1} v = \frac{1}{1 - \rho(L)} v. \tag{69}$$

Hence a subcritical system can have arbitrarily large steady amplification as its governability margin approaches zero from above.

The norm in Proposition 5 is adapted to the Perron direction. In a fixed Euclidean norm, nonnormality can create additional amplification, so the scalar quantity $1/\mathfrak{m}$ is an exact norm formula for the Perron-adapted norm used in Proposition 5. The distinction between resolvent amplification and transient amplification in Section 9.6 is therefore substantive.

Proposition 6 (Steady-boundary sensitivity)**.** *Let $L(\theta)$ and $\bar{k}(\theta)$ be differentiable on an open interval and suppose $\rho(L(\theta)) < 1$. Then*

$$\frac{da_\infty}{d\theta} = (I - L)^{-1}\left[\frac{d\bar{k}}{d\theta} + \frac{dL}{d\theta}a_\infty\right]. \tag{70}$$

If $L_\theta \geq 0$, $\bar{k}_\theta \geq 0$, and $a_\infty \geq 0$, then $da_\infty/d\theta \geq 0$ componentwise.

Equation (70) gives a policy-relevant decomposition. A parameter can raise long-run quiet capacity directly through $\bar{k}$, indirectly by strengthening the feedback matrix L, or through both channels. The resolvent multiplies either effect. Near the phase boundary, therefore, modest parameter changes can have large steady consequences even before stability is lost.

For the scalar frontier-AI specialization,

$$\mathfrak{m} = \eta - \Gamma, \qquad Q_\infty = \frac{K_0}{\mathfrak{m}}, \tag{71}$$

and, for $\Gamma < \eta$,

$$\frac{\partial Q_\infty}{\partial \Gamma} = \frac{K_0}{(\eta - \Gamma)^2}, \qquad \frac{\partial Q_\infty}{\partial \eta} = -\frac{K_0}{(\eta - \Gamma)^2}, \qquad \frac{\partial^2 Q_\infty}{\partial \Gamma^2} = \frac{2K_0}{(\eta - \Gamma)^3}. \tag{72}$$

The convex divergence is an early-warning result: large governability losses can occur within the formally stable region.

10.1 Scope of cross-domain comparison

The spectral radius is invariant to invertible changes of state coordinates by Proposition 3, and to invertible action-coordinate changes up to similarity. That protects the phase object from arbitrary reparameterization within a specified model. A raw value such as $\rho(L) = 0.8$ gains comparable substantive meaning across applications only under a common measurement specification or an explicit normalization. State definitions, temporal aggregation, intervention concepts, and model misspecification differ across applications. Across domains, the portable object is the theorem form and its mechanism map. An estimated phase label remains conditional on each domain's state definition and response model. Numerical magnitudes are compared within a common measurement specification or under an explicit normalization.

11 Changing Regimes: Switched and Stochastic Systems

Technology, governments, market structure, and international coalitions can change discontinuously. A fixed matrix can therefore be the wrong object.

Consider

$$x_t = M_{\sigma_t} x_{t-1} + g_{\sigma_t}, \qquad M_{\sigma_t} \in \mathcal{M}, \tag{73}$$

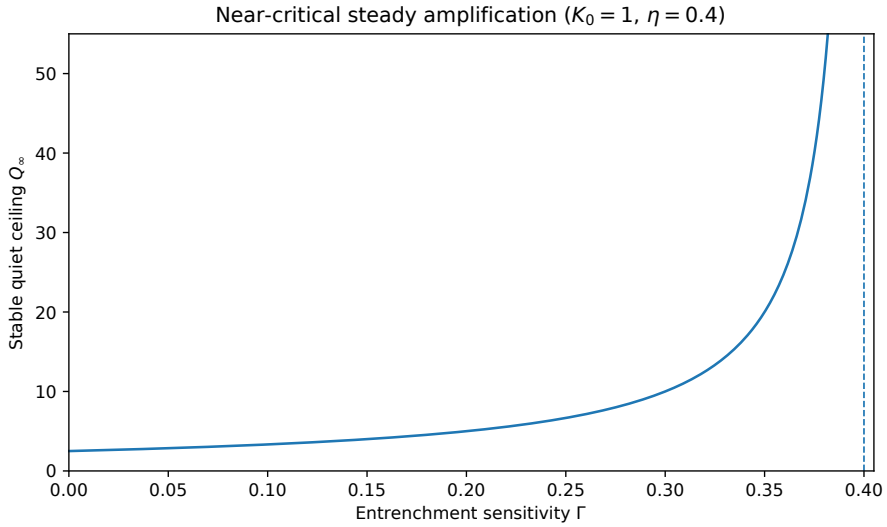


Figure 2: Illustrative scalar amplification with $K_0 = 1$ and $\eta = 0.4$. The figure uses illustrative theorem parameters. The ceiling grows sharply while the model remains subcritical.

where $\mathcal{M}$ is a bounded set of matrices and the inputs g_σ are uniformly bounded.

The joint spectral radius is

$$\widehat{\rho}(\mathcal{M}) := \limsup_{T\to\infty} \sup_{M_1,\dots,M_T\in\mathcal{M}} \|M_T \cdots M_1\|^{1/T}, \tag{74}$$

which is independent of the chosen matrix norm (Berger and Wang, 1992; Jungers, 2009).

Theorem 5 (Arbitrary-switching boundedness)**.** *If $\widehat{\rho}(\mathcal{M}) < 1$ and $\sup_\sigma \|g_\sigma\| < \infty$, then the switched affine system (73) is uniformly bounded over all switching sequences and all bounded initial conditions.*

If $\widehat{\rho}(\mathcal{M}) > 1$, there exist products of matrices in $\mathcal{M}$ with exponential norm growth. Whenever the corresponding unstable directions are reachable from the initial state or bounded forcing and observable in the capacity functional, some switching sequence generates exponential capacity growth.

At $\widehat{\rho}(\mathcal{M}) = 1$, finer product structure determines boundedness; polynomial or other subexponential growth can occur.

This produces an important policy warning. It is possible that

$$\rho(M_s) < 1 \tag{75}$$

for every regime s while

$$\widehat{\rho}(\{M_s\}) > 1. \tag{76}$$

A sequence of individually governable regimes can therefore be jointly ungovernable under switching.

For stochastic σ_t, the corresponding object is a random matrix product. Under stationarity and integrability conditions, a top Lyapunov exponent describes almost-sure exponential growth. Markov jump linear systems provide a developed theory for such cases (Costa, Fragoso, and Marques, 2005). An application-specific stochastic-switching law can be supplied when empirical structure identifies one.

12 A Coalition-Enforcement Microfoundation

The general framework allows many response mechanisms. This section provides one that is useful for international frontier-AI governance.

12.1 The stage game

Let governments be $N = \{1, \dots, n\}$, $n \geq 2$. Government i chooses $e_i \in \{0, 1\}$, where $e_i = 1$ means that it contributes to enforcement. Enforcement succeeds according to a monotone technology

$$S : \{0,1\}^n \to \{0,1\}, \tag{77}$$

with $S(0) = 0$ and at least one winning coalition. Define

$$\mathcal{W} := \{C \subseteq N : S(\mathbf{1}_C) = 1\}. \tag{78}$$

Let x be the inherited economic and political state. Government i has normalized enforcement cost $\kappa_i(x) > 0$. Let $z \geq 0$ be public effective severity. Payoffs are

$$u_i(e; z, x) = -\kappa_i(x)e_i - z[1 - S(e)]. \tag{79}$$

Define

$$K(x) := \min_{C \in \mathcal{W}} \sum_{i \in C} \kappa_i(x). \tag{80}$$

Proposition 7 (State-dependent exact potential)**.** *For every fixed (z, x), the game (79) is an exact potential game with*

$$\Phi(e; z, x) = -\sum_i \kappa_i(x)e_i - z[1 - S(e)]. \tag{81}$$

Under asynchronous log-linear learning with common inverse-noise parameter β, the stochastically stable profiles are universal acquiescence when $z < K(x)$, minimum-cost winning coalitions when $z > K(x)$, and both sets when $z = K(x)$.

For two governments under unanimity, the same threshold $K(x) = \kappa_1(x) + \kappa_2(x)$ is the Harsanyi–Selten risk-dominance threshold when both strict equilibria exist.

12.2 Heterogeneous exposure and coalition design

The common coefficient on severity in (79) is a normalization that can represent heterogeneous exposure when each government's failed-response loss is proportional to the same public severity statistic. Let primitive payoffs be

$$\widetilde{u}_i(e; z, x) = -c_i(x)e_i - \lambda_i(x)z[1 - S(e)], \qquad c_i(x) > 0, \quad \lambda_i(x) > 0. \tag{82}$$

Define the normalized cost

$$\kappa_i(x) := \frac{c_i(x)}{\lambda_i(x)}. \tag{83}$$

Proposition 8 (Weighted-potential representation under heterogeneous exposure)**.** *For fixed* (z, x)*, the primitive game* (82) *is a weighted potential game with player weight* $\lambda_i(x)$ *and potential* (81) *constructed from* $\kappa_i(x) = c_i(x)/\lambda_i(x)$*. Its Nash-equilibrium set is therefore the same as that of the normalized game. The minimum normalized cost of a winning coalition is*

$$K(x) = \min_{C \in \mathcal{W}} \sum_{i \in C} \frac{c_i(x)}{\lambda_i(x)}. \tag{84}$$

If primitive log-linear choice uses inverse-noise parameter $\widetilde{\beta}_i = \beta/\lambda_i(x)$ *for government* i*, the stochastic process is exactly the normalized common-*β *process and the stochastic-selection threshold remains* $K(x)$*.*

The proposition captures heterogeneous stakes in a common public severity. More general differences in beliefs, nonlinear losses, or actor-specific severity measures require a different game. For foresight work, (84) is useful because it shows that coalition cost depends on response cost relative to exposure rather than on response cost alone.

Weighted quotas and representation rules can be encoded directly in $\mathcal{W}$. For example, a response can require both a technical voting quota and participation from specified regions. Such legitimacy constraints should be upward closed if the maintained enforcement technology is monotone.

12.3 Measurement, common knowledge, and response selection

Private measurement, public information, and response rules perform distinct institutional functions. Private measurement determines what individual actors observe; public information determines what becomes commonly known; and response protocols change incentives after a trigger. The distinction is represented directly in the model below. Consider two governments, unanimity, equal normalized cost κ, and a common public report r about a severity state θ. Let

$$\mu(r) := \mathbb{E}[\theta \mid r]. \tag{85}$$

Conditional expected payoffs have the same form as (79) with $z = \mu(r)$.

Proposition 9 (Common measurement can leave response multiplicity)**.** *In the two-government unanimity game with common posterior mean* $\mu(r)$*:*

(i) universal acquiescence is a Bayesian Nash equilibrium for every r*;*
(ii) joint enforcement is a Bayesian Nash equilibrium when $\mu(r) \geq \kappa$*;*
(iii) under log-linear learning applied to the conditional expected-payoff game, universal acquiescence is uniquely stochastically stable when $\mu(r) < 2\kappa$*, joint enforcement is uniquely stochastically stable when* $\mu(r) > 2\kappa$*, and both maximize the potential at equality.*

This result separates observation from response selection. Better measurement can reduce uncertainty and make beliefs more common while the coordination problem remains. Global-game models show how small private-signal noise can select equilibria under additional information assumptions (Carlsson and van Damme, 1993; Frankel, Morris, and Pauzner, 2003; Morris and Shin, 2003; Hellwig, 2002; Angeletos, Hellwig, and Pavan, 2007). The present framework uses that literature as an optional information layer rather than as a universal selection rule.

A pre-agreed protocol can instead change the post-trigger incentives directly. Fix a designated winning coalition $C^\star \in \mathcal{W}$ and a certified trigger. If a designated government that fails to contribute after the trigger incurs a nonperformance cost ψ_i, its payoff becomes

$$u_i^P(e; z, x) = u_i(e; z, x) - \psi_i \mathbf{1}\{i \in C^\star,\ e_i = 0,\ \text{trigger active}\}. \tag{86}$$

Theorem 6 (Sufficient protocol implementation)**.** *When the trigger is active, if*

$$\psi_i > \kappa_i(x) \qquad \text{for every } i \in C^\star, \tag{87}$$

then contribution is strictly dominant for every designated member. The unique Nash equilibrium has every member of $C^\star$ contributing and every outsider acquiescing, so enforcement succeeds.

The condition is sufficient. Its foresight use is direct: a protocol can be credible at one inherited state and lose force later if $\kappa_i(x)$ grows past its fixed nonperformance cost. Section 22 returns to this erosion channel.

12.4 Finite-noise response probability

Let

$$c_x(e) := \sum_i \kappa_i(x) e_i, \tag{88}$$

and define

$$A_\beta(x) := \sum_{e: S(e)=1} \exp[-\beta c_x(e)], \tag{89}$$

$$L_\beta(x) := \sum_{e: S(e)=0} \exp[-\beta c_x(e)]. \tag{90}$$

The stationary probability of successful enforcement is exactly

$$\boxed{P_\beta(z, x) = \frac{A_\beta(x)}{A_\beta(x) + e^{-\beta z} L_\beta(x)}.} \tag{91}$$

When a response episode has enough revision opportunities to approximate the stationary log-linear distribution, (91) supplies one concrete response hazard q_θ. It is increasing in severity and depends on the entire state-dependent coalition-cost structure. Finite response latency can instead be represented by the corresponding finite-step Markov transition probability and included as an additional state or parameter.

12.5 Forward-looking institutions

Equation (79) is a stage-game microfoundation. If acquiescence changes future costs, rational governments can value intervention today partly because it preserves future options. Let $V_i(x)$ be a continuation value and let government action affect the next-state kernel. A dynamic version replaces stage payoff differences by

$$\Delta u_i^{\text{dyn}} = \Delta u_i^{\text{current}} + \delta_i \Delta \mathbb{E}[V_i(X_{t+1})]. \tag{92}$$

Any Markov-perfect equilibrium of that dynamic game induces a state-dependent response correspondence or hazard that plugs into the general model. The static threshold is one empirically useful microfoundation; the dynamic formulation adds the option value of preserving future governability.

13 Strategic Entrenchment and the Governability Race

The preceding sections characterize the room available to a technological actor. A forward-looking actor may also care that present movement changes future room. This strategic possibility has antecedents in the regulation literature: regulated firms can distort investment to influence later policy (Glazer and Janeba, 2004; Tarui and Polasky, 2005), and self-regulation can be used to preempt political action (Maxwell, Lyon, and Hackett, 2000). The contribution here is narrower. The first-meter framework isolates a general continuation-value channel generated by an endogenous governability state and connects it to the same feedback objects that govern the boundary dynamics.

13.1 The dynamic governability game

Let the technological actor choose $a_t \in A(x_t)$ and the governing side choose a preservation or response action $p_t \in P(x_t)$. Let

$$X_{t+1} \sim \mathsf{P}(\cdot \mid X_t, a_t, p_t), \tag{93}$$

and let stage payoffs be $u_F(x, a, p)$ for the technological actor and $u_G(x, a, p)$ for the governing side. A stationary Markov equilibrium (σ_F, σ_G) satisfies the coupled Bellman optimality conditions

$$V_F(x) = \sup_{a \in A(x)} \left\{ u_F(x, a, \sigma_G(x)) + \beta_F \mathbb{E}[V_F(X') \mid x, a, \sigma_G(x)] \right\}, \tag{94}$$

$$V_G(x) = \sup_{p \in P(x)} \left\{ u_G(x, \sigma_F(x), p) + \beta_G \mathbb{E}[V_G(X') \mid x, \sigma_F(x), p] \right\}. \tag{95}$$

Equations (94)–(95) display pure stationary best responses. Finite-state, finite-action discounted versions admit stationary equilibria in mixed strategies by standard stochastic-game results (Fink, 1964); the mixed-strategy form replaces the opponent's action by expectation under its equilibrium distribution. Continuous models require the continuity, compactness, and equilibrium-existence conditions appropriate to the application.

For a deterministic differentiable transition $x' = F(x, a, p)$, the technological actor's marginal continuation-value term is

$$\mathcal{E}_F(x, a, p) := \beta_F \nabla V_F(x')^\top F_a(x, a, p). \tag{96}$$

This is the *strategic first-meter wedge*: it is the part of the marginal value of current action that arises because current action changes the inherited future state. The governing side has the analogous continuation term

$$\mathcal{E}_G(x, a, p) := \beta_G \nabla V_G(x')^\top F_p(x, a, p). \tag{97}$$

The signs of these wedges depend on the application. Positive $\mathcal{E}_F$ creates an entrenchment incentive; a positive preservation wedge creates an incentive to invest in future response capacity.

Remark 3 (Discounted structural return)**.** *In the positive-affine state system, hold subsequent technological actions fixed after a one-unit action at the current date. The resulting increments to future quiet boundaries are $CB, CHB, CH^2B, \ldots$. If the actor values the boundary increment $r+1$ periods ahead with discount factor β_F^{r+1}, the total discounted structural return is*

$$L_{\beta_F}^{\text{cont}} := \beta_F C(I - \beta_F H)^{-1} B. \tag{98}$$

As $\beta_F \uparrow 1$ under $\rho(H) < 1$, this continuation-value object converges to the undiscounted loop matrix L. The phase criterion remains attached to the actual boundary feedback dynamics; (98) measures how a forward-looking actor values the inherited state channel.

13.2 A general strategic first-meter wedge

Let the actor have scalar action a in a totally ordered feasible set $A(x)$ and one-period payoff $u(x, a)$. Let $F(x, a)$ be the next state. Define the unique myopic action

$$a^M(x) := \arg\max_{a \in A(x)} u(x, a). \tag{99}$$

A forward-looking actor instead values

$$J_x(a) := u(x, a) + \beta V(F(x, a)), \qquad \beta \in (0, 1), \tag{100}$$

where V is the continuation value after the current action.

Proposition 10 (Strategic first-meter acceleration)**.** *Fix x. Suppose $A(x)$ is a compact interval, $u(x, \cdot)$ is continuous with unique maximizer $a^M(x)$, $F(x, a)$ is nondecreasing in a, and V is nondecreasing in the state order. Then every maximizer $a^D(x)$ of (100) satisfies*

$$a^D(x) \geq a^M(x). \tag{101}$$

If $a^M(x)$ is interior, u and $V \circ F$ are differentiable at $a^M(x)$, and

$$\frac{d}{da} V(F(x, a))\Big|_{a=a^M(x)} > 0, \tag{102}$$

then $a^D(x) > a^M(x)$.

The proposition is a monotone comparative-static benchmark. When entrenchment raises the actor's continuation value, internalizing that first-meter channel shifts every dynamic optimum weakly above the unique myopic optimum. Its empirical content therefore comes from establishing the sign and magnitude of the continuation channel $V \circ F$; the positive-affine model supplies one explicit structural route through the discounted loop in (98).

Remark 4 (When anticipation slows the actor)**.** *If the continuation value is instead decreasing in the state created by current action—for example because larger movement raises the probability of an early binding intervention—the inequality can reverse. The theory therefore predicts strategic acceleration only when the future-governability effect dominates the future-response penalty.*

13.3 The institutional mirror: preservation has continuation value

Let an institution choose a preservation action p from a compact interval. The action can represent interoperability requirements, supplier diversification, preserved human expertise, financial exposure limits, measurement capacity, or another intervention that improves future options. Let $g(x, p)$ be the institution's current payoff and let $F_G(x, a, p)$ be the next inherited state, ordered so that larger states are more entrenched. Suppose larger p weakly lowers F_G, and let W be the institution's continuation value, weakly decreasing in entrenchment.

Proposition 11 (Dynamic preservation exceeds the myopic benchmark)**.** *Suppose $g(x, \cdot)$ has a unique myopic maximizer $p^M(x)$, $F_G(x, a, p)$ is nonincreasing in p, and W is nonincreasing in the inherited entrenchment state. Then every maximizer of*

$$g(x, p) + \beta_G W(F_G(x, a, p)) \tag{103}$$

weakly exceeds $p^M(x)$. Under an interior differentiability condition with a strictly positive continuation derivative at $p^M(x)$, the inequality is strict.

This is the formal option-value channel of preserving governability. It is related to the classic irreversibility and quasi-option-value literature (Arrow and Fisher, 1974; Henry, 1974; Dixit and Pindyck, 1994), but the object being preserved here is the future policy correspondence itself. It also complements work on endogenous state capacity, where governments invest in legal and fiscal capacities that constrain future policy possibilities (Besley and Persson, 2009).

13.4 A closed-form policy-first governability race

A two-period model makes the strategic interaction transparent. At date 0, the institution chooses preservation $p \in [0, c_0]$, reducing the sensitivity of future quiet room from c_0 to $c_0 - p$. The actor then chooses current movement $a \geq 0$. Current movement creates entrenchment ba, $b > 0$. Date-1 quiet room is

$$K + b(c_0 - p)a. \tag{104}$$

The actor's objective is

$$U_F(a; p) = ra - \frac{s}{2}a^2 + \beta v \left[K + b(c_0 - p)a\right], \qquad r, s, v > 0. \tag{105}$$

The institution anticipates the actor's response and minimizes

$$J_G(p) = \lambda \left[K + b(c_0 - p)a(p)\right] + \frac{\phi}{2}p^2, \qquad \lambda, \phi > 0. \tag{106}$$

Here the date-1 quiet boundary is a reduced-form exposure proxy in the institution's objective; λ is its marginal welfare weight. The benchmark isolates strategic feedback, while Section 17 supplies the richer welfare problem. Let $q := \beta v b$. The comparative statics below vary q through the actor's continuation valuation βv, holding b and the other structural parameters fixed.

Theorem 7 (Policy-first entrenchment–preservation benchmark)**.** *Assume*

$$c_0 s\phi > \lambda br. \tag{107}$$

Then the policy-first game (105)–(106) has the unique interior Stackelberg solution

$$p^*(q) = \frac{\lambda b \left[r + 2qc_0\right]}{s\phi + 2\lambda bq}, \tag{108}$$

$$a^*(q) = \frac{1}{s}\left[r + q\frac{c_0 s\phi - \lambda br}{s\phi + 2\lambda bq}\right]. \tag{109}$$

Moreover,

$$\frac{dp^*}{dq} = \frac{2\lambda b\left(c_0 s\phi - \lambda br\right)}{(s\phi + 2\lambda bq)^2} > 0, \tag{110}$$

$$\frac{da^*}{dq} = \frac{s\phi\left(c_0 s\phi - \lambda br\right)}{s(s\phi + 2\lambda bq)^2} > 0. \tag{111}$$

Thus an increase in the actor's value of future quiet room induces both greater current technological movement and greater institutional preservation investment. In addition, under (107),

$$a^*(q) - \frac{r}{s} = \frac{q(c_0 s\phi - \lambda br)}{s(s\phi + 2\lambda bq)} > 0, \tag{112}$$

so the equilibrium action exceeds the myopic benchmark r/s for every $q > 0$. Strategic anticipation can therefore generate a two-sided entrenchment–preservation race in this benchmark.

Remark 5 (Interpretation of the interior condition)**.** *Condition* (107) *is exactly the condition that the institution's equilibrium preservation remains below the technological sensitivity* c_0*. If it fails, the constrained solution can reach a boundary and the comparative statics become piecewise.*

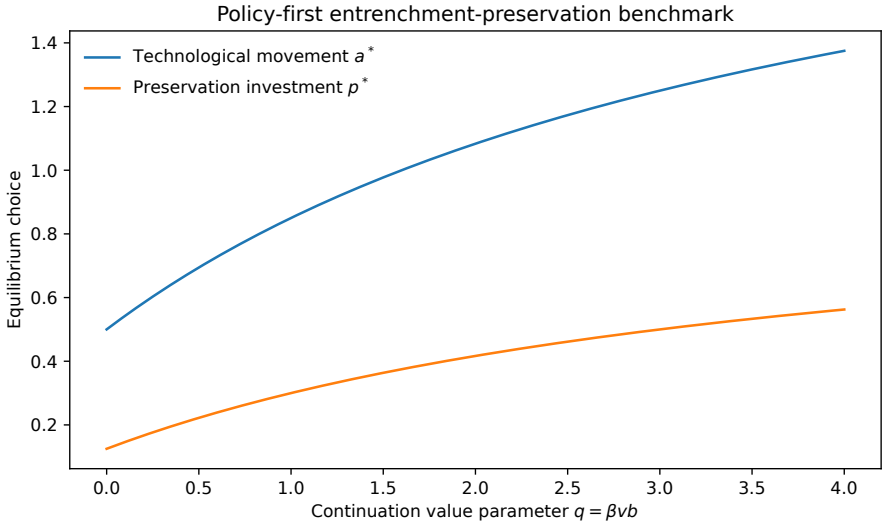


Figure 3: Closed-form policy-first benchmark for $r = 1$, $s = 2$, $c_0 = 1$, $\lambda = 0.5$, $b = 1$, and $\phi = 2$. As the actor values future quiet room more strongly, both current movement and preservation investment rise. The numbers are illustrative.

13.5 Endogenous feedback architecture

Strategic action can also change the feedback architecture itself. Let a structural state Θ_t determine

$$H_t = H(\Theta_t), \qquad B_t = B(\Theta_t), \qquad C_t = C(\Theta_t), \tag{113}$$

with

$$\Theta_{t+1} = \Psi(\Theta_t, a_t, p_t, \zeta_{t+1}). \tag{114}$$

Whenever $\rho(H_t) < 1$, define the local structural loop

$$L_t = C_t(I - H_t)^{-1} B_t. \tag{115}$$

A technological actor can then invest in persistence, dependence creation, or response sensitivity, while the governing side can invest in substitution, interoperability, diversified infrastructure, or stronger anchors. Proposition 14 later gives the first-order map from changes in (H, B, C) into changes in L and its Perron root. This makes strategic entrenchment a choice over both the technological level and the future feedback architecture.

13.6 From the exact benchmark to application

Equations (93)–(95) define the full dynamic game. Propositions 10–11 identify robust one-sided order effects, while Theorem 7 supplies an exact two-sided precommitment benchmark. Applications with richer state spaces can solve the coupled Bellman system analytically where structure permits and numerically otherwise, then feed the equilibrium response rule back into the general stochastic framework of Section 7.

14 The Benchmark-and-Entrenchment Specialization

This section gives a transparent scalar model and shows exactly how benchmark drift and endogenous entrenchment fit inside the general loop-gain theory.

14.1 State dynamics

Normalize the formal institutional anchor to zero. Let $y_t \geq 0$ be cumulative movement in a governance-relevant technological state, with

$$y_t \geq y_{t-1}, \qquad y_0 = 0. \tag{116}$$

Let m_{t-1} be a benchmark inherited from tolerated outcomes. Effective severity is

$$z_t^\eta := \eta y_t + (1-\eta)(y_t - m_{t-1}) = y_t - (1-\eta)m_{t-1}, \qquad \eta \in [0,1]. \tag{117}$$

If the outcome is tolerated,

$$m_t = (1-\alpha)m_{t-1} + \alpha y_t, \qquad \alpha \in (0,1]. \tag{118}$$

Let $x_{j,t}$, $j = 1, \ldots, J$, be entrenchment states with

$$x_{j,t} = (1-\delta_j)x_{j,t-1} + \delta_j y_t, \qquad \delta_j \in (0,1], \qquad x_{j,0} = 0. \tag{119}$$

The lagged coalition threshold is

$$K_t = K_0 + \sum_{j=1}^{J} \gamma_j x_{j,t-1}, \qquad K_0 > 0, \qquad \gamma_j \geq 0. \tag{120}$$

The lag prevents the current technological move from immediately raising its own response threshold before institutions respond.

A zero-noise quiet path satisfies

$$z_t^\eta < K_t \tag{121}$$

for every date.

14.2 Embedding in the positive system

Define

$$s_t := (m_t, x_{1,t}, \ldots, x_{J,t})^\top, \tag{122}$$

$$D := \operatorname{diag}(\alpha, \delta_1, \ldots, \delta_J), \qquad c^\top := (1-\eta, \gamma_1, \ldots, \gamma_J), \tag{123}$$

and let $\mathbf{1}$ have dimension $J+1$. Then

$$s_t = (I - D)s_{t-1} + D\mathbf{1}\, y_t, \tag{124}$$

and quietness implies

$$y_t < K_0 + c^\top s_{t-1}. \tag{125}$$

Thus the positive-system matrices are

$$H = I - D, \qquad B = D\mathbf{1}, \qquad C = c^\top, \qquad k = K_0, \qquad d = 0. \tag{126}$$

Since $I - H = D$,

$$L = C(I - H)^{-1}B = c^\top D^{-1} D\mathbf{1} = c^\top \mathbf{1} = 1 - \eta + \Gamma, \tag{127}$$

where

$$\Gamma := \sum_{j=1}^{J} \gamma_j. \tag{128}$$

Theorem 8 (Exact first-meter phase transition)**.** *Let $K_0 > 0$, $\alpha, \delta_j \in (0, 1]$, $\eta \in [0, 1]$, and $\gamma_j \geq 0$. Define the finite-horizon quiet capacity*

$$Q_T := \sup\{y_T : \textit{the length-}T\textit{ path satisfies (121)}\}. \tag{129}$$

Let

$$M = I - D + D\mathbf{1}c^\top, \qquad g = K_0 D\mathbf{1}. \tag{130}$$

Then

$$\boxed{Q_T = K_0 + c^\top \left(\sum_{r=0}^{T-2} M^r \right) g} \tag{131}$$

for $T \geq 1$, with the sum interpreted as zero at $T = 1$.

Moreover:

(i) if $\Gamma < \eta$, then $\rho(M) < 1$ and

$$\boxed{Q_\infty = \frac{K_0}{\eta - \Gamma};} \tag{132}$$

(ii) if $\Gamma = \eta$, then $\rho(M) = 1$ and Q_T grows linearly, with

$$\lim_{T\to\infty} \frac{Q_T}{T} = \frac{K_0}{c^\top D^{-1}\mathbf{1}}; \tag{133}$$

(iii) if $\Gamma > \eta$, then $\rho(M) > 1$ and Q_T has geometric growth rate $\rho(M)$ in the sense that

$$\lim_{T\to\infty} Q_T^{1/T} = \rho(M) \tag{134}$$

when the nonzero Perron mode is observed, which holds for the scalar capacity above.

The result says that fixed-anchor retention and economic entrenchment are opposing feedbacks. The relevant comparison is the strength of retained anchoring against the total feedback from tolerated development into future intervention costs.

14.3 Illustrative walk toward the phase boundary

The following numbers illustrate the theorem. Set $K_0 = 1$ and $\eta = 0.40$. If $\Gamma = 0.25$, then

$$L = 0.85, \qquad Q_\infty = \frac{1}{0.40 - 0.25} = 6.\overline{6}. \tag{135}$$

If entrenchment rises to $\Gamma = 0.35$, the system remains subcritical while

$$L = 0.95, \qquad Q_\infty = 20. \tag{136}$$

At $\Gamma = 0.40$ the model reaches the critical boundary, and at $\Gamma = 0.45$ it enters the supercritical region. The example shows why the margin and amplification objects in Section 10 carry information before the phase changes.

For measurement, the aggregate sensitivity can be decomposed into named channels such as

$$\Gamma = \Gamma_{\text{finance}} + \Gamma_{\text{infrastructure}} + \Gamma_{\text{platform}} + \Gamma_{\text{human capital}} + \Gamma_{\text{fiscal}} + \Gamma_{\text{geopolitical}} + \cdots, \tag{137}$$

provided each term is defined on the same action scale and measures its long-run contribution to the response boundary. Equation (137) is a measurement decomposition. Its terms require empirical identification.

14.4 Benchmark-only corollary

Set $J = 0$. Then $\Gamma = 0$, $M = 1 - \alpha\eta$, and Theorem 8 gives

$$Q_T = \begin{cases} K_0[1 + \alpha(T-1)], & \eta = 0, \\ \dfrac{K_0}{\eta} - \dfrac{(1-\eta)K_0}{\eta}(1-\alpha\eta)^{T-1}, & \eta > 0. \end{cases} \tag{138}$$

Thus a fully moving evaluation benchmark permits linear cumulative movement, while every positive fixed-anchor component gives a finite ceiling in the absence of economic entrenchment.

14.5 Formal anchor revision and passive normalization

A scientific or institutional anchor can require deliberate revision as evidence changes. The framework therefore separates formal revision from passive normalization. Let b_t be a formally approved anchor and let m_t remain the benchmark generated by tolerated outcomes. Severity in levels is

$$z_t = \eta(y_t - b_t) + (1-\eta)(y_t - m_{t-1}). \tag{139}$$

Between formal revisions, b_t is constant. With the coordinate shift $\tilde{y}_t = y_t - b_t$ and $\tilde{m}_t = m_t - b_t$, equation (139) reduces exactly to the normalized severity rule in (117). Tolerance updates m_t through (118); b_t changes only when the formal revision rule is invoked.

This separation gives two distinct accounting choices. A formal revision can reset both the anchor and the carried benchmark to the current physical state, which restarts the future problem at a new origin. An institution can instead revise substantive scientific thresholds while retaining a separate cumulative ledger from the earlier anchor. The second design preserves the historical accounting term while allowing scientific updating. For foresight work, the revision rule and the ledger rule should therefore appear as separate institutional state transitions.

14.6 Why the adaptation rates disappear from the stable ceiling

The rates α and δ_j change how quickly state trackers move, but every tracker converges to the same constant y along a constant long-run boundary. In the loop notation, D cancels from

$$C(I - H)^{-1}B = c^\top D^{-1} D\mathbf{1}. \tag{140}$$

This cancellation is specific to the common-target tracker (124). In a more general model with cross-channel propagation, depreciation, saturation, or heterogeneous targets, the full matrix $L = C(I - H)^{-1}B$ remains and adaptation parameters generally remain active.

15 Coalition Switching, Participation, and History

15.1 Endogenous coalition composition

Different governments can become dependent on the technology at different rates. Let an aggregate entrenchment state be $x \geq 0$ and write

$$\kappa_i(x) = a_i + b_i x, \qquad a_i > 0, \quad b_i \geq 0. \tag{141}$$

For each winning coalition C, define

$$A_C := \sum_{i \in C} a_i, \qquad B_C := \sum_{i \in C} b_i. \tag{142}$$

Then

$$K(x) = \min_{C \in \mathcal{W}} (A_C + B_C x). \tag{143}$$

Proposition 12 (Coalition switching)**.** *The threshold $K(x)$ in (143) is continuous, nondecreasing, concave, and piecewise affine. Changes in the cheapest winning coalition occur at finitely many breakpoints on every bounded interval.*

Let

$$\underline{B} := \min_{C \in \mathcal{W}} B_C. \tag{144}$$

Corollary 1 (Long-run coalition option value)**.** *Consider the scalar benchmark model (117)–(118) with one entrenchment tracker*

$$x_t = (1 - \delta)x_{t-1} + \delta y_t, \qquad \delta \in (0, 1], \tag{145}$$

and replace (120) by the coalition threshold (143), with $A_C > 0$ for every winning coalition. Then the maximal quiet path is bounded when $\underline{B} < \eta$, has asymptotically linear growth when $\underline{B} = \eta$, and has geometric growth when $\underline{B} > \eta$.

The coalition that determines long-run governability can therefore differ from the coalition that is cheapest at the initial state. A coalition with a larger initial intercept and a smaller entrenchment slope can preserve future governability.

15.2 Participation and historical position

Historical extraction or distributional claims matter only through an explicit channel. Let state i receive gross regime benefit B_i, bear compliance cost C_i, and have outside option $D_i(h_i, x)$, where h_i is a historical state and x collects current strategic conditions. Let an institution impose a restorative transfer floor $r_i(h_i) \geq 0$. The minimum transfer required for participation is

$$\tau_i(h_i, x) := \max\{0,\ D_i(h_i, x) + C_i - B_i,\ r_i(h_i)\}. \tag{146}$$

A coalition's effective cost can then be written

$$K(x,h) = \min_{C \in \mathcal{W}} \sum_{i \in C} [\kappa_i(x) + \tau_i(h_i, x)]. \tag{147}$$

Proposition 13 (History requires a channel)**.** *Changing h_i affects (147) only through current outside options, benefits, compliance costs, enforcement costs, coalition eligibility, or an explicit restorative rule. When those primitives are invariant in h_i, the coalition cost is invariant as well.*

If $D_i(h_i, x)$ and $r_i(h_i)$ are nondecreasing in h_i, holding other primitives fixed, then $K(x,h)$ is nondecreasing in each historical coordinate.

This separates a positive strategic claim from a normative restorative claim. History can matter because it changes present options and participation, or because an institution explicitly gives it legal or distributive force. A model should state which channel is active.

16 Cumulative Hazard, Risk Budgets, and Continuous Time

16.1 Cumulative hazard across periods

Suppose a deterministic path has conditional response hazards $q_t \in [0, 1)$. Then

$$\mathbb{P}(\tau > T) = \prod_{t=0}^{T-1} (1 - q_t). \tag{148}$$

Therefore

$$\mathbb{P}(\tau \le T) \le \delta \quad \Longleftrightarrow \quad \sum_{t=0}^{T-1} -\log(1 - q_t) \le -\log(1 - \delta). \tag{149}$$

The transformed quantity $-\log(1 - q_t)$ is an additive risk expenditure. Equation (149) gives a natural remaining-risk state for deterministic dynamic programming.

For stochastic state paths, the survival probability is the expectation of a product of conditional survival factors. Constrained-MDP methods or a risk-budget augmentation are then the correct tools. A rule that checks only whether each q_t is below a fixed number can severely understate cumulative intervention probability.

16.2 Continuous-time extension

Some mechanisms evolve on different time scales. In continuous time, let

$$dX_t = f(X_t, a_t)dt + \Sigma(X_t, a_t)dW_t \tag{150}$$

and let $\lambda(X_t, a_t) \ge 0$ be a response intensity. Conditional on the path,

$$\mathbb{P}(\tau > T \mid X_{[0,T]}, a_{[0,T]}) = \exp\left[-\int_0^T \lambda(X_s, a_s) ds \right]. \tag{151}$$

The continuous-time model therefore preserves the same separation between state dynamics and response hazard. Finance can react quickly, infrastructure slowly, and institutional benchmarks at another rate while allowing each mechanism to operate on its own time scale.

16.3 Positive-affine continuous-time loop

The positive feedback result has a direct continuous-time counterpart. Let

$$\dot{x}(t) = Ax(t) + Ba(t) + d, \qquad a(t) \le k + Cx(t), \tag{152}$$

where A is Metzler and Hurwitz and B, C, d, k are nonnegative. Define

$$L_c := C(-A)^{-1}B \ge 0, \qquad \bar{k}_c := k + C(-A)^{-1}d. \tag{153}$$

Theorem 9 (Continuous-time positive small gain)**.** *Under* (152)*, the saturated boundary dynamics*

$$\dot{x} = (A + BC)x + (Bk + d) \tag{154}$$

are Hurwitz stable exactly when

$$\rho(L_c) < 1. \tag{155}$$

In the stable case,

$$a_\infty = (I - L_c)^{-1}\bar{k}_c, \qquad x_\infty = (-A)^{-1}(d + Ba_\infty). \tag{156}$$

Thus the discrete-time loop $C(I - H)^{-1}B$ and continuous-time loop $C(-A)^{-1}B$ play the same structural role: each integrates the full propagation of one action through the stable state subsystem and returns it to the response boundary.

17 The Economic Value of Preserving Governability

The framework separates current restriction from investment that preserves future choice. This section gives that separation a welfare interpretation. Irreversibility and option value provide the economic foundation for valuing future flexibility (Arrow and Fisher, 1974; Henry, 1974; Dixit and Pindyck, 1994). The object here is the counterfactual welfare value of preserving a future institutional response set that technological entrenchment can erode.

17.1 A Bellman formulation

Let p_t denote governability-preserving investment and a_t technological movement. Let current social surplus be $S(X_t, a_t, p_t)$ and let the state evolve according to

$$X_{t+1} \sim P(\cdot \mid X_t, a_t, p_t). \tag{157}$$

A social planner's value is

$$V_G(x) = \sup_{a,p} \left\{ S(x, a, p) + \beta_G \mathbb{E}[V_G(X') \mid x, a, p] \right\}, \tag{158}$$

subject to any contemporaneous risk or legal constraints. The continuation term values future productive opportunities and future policy options jointly. A decomposition becomes well defined after the analyst states an explicit reference transition. The monograph therefore uses *governability option value* as a counterfactual welfare difference tied to that reference.

Definition 6 (Governability option value)**.** *Let π^p be a committed preservation rule and let $V^{\pi^p}(x)$ be social welfare when that rule is fixed and all remaining choices are optimized under the stated planner or equilibrium concept. For a reference rule π^{p^0}, define*

$$\mathrm{GOV}(x; \pi^p, \pi^{p^0}) := V^{\pi^p}(x) - V^{\pi^{p^0}}(x). \tag{159}$$

Both counterfactuals use the same technological opportunity set and welfare criterion; the preservation rule and the transition and response structure it induces are the changing objects. When technological action is strategic, V^{π^p} is evaluated at the equilibrium induced by the committed rule.

The definition makes the welfare claim falsifiable. Interoperability, redundancy, or independent expertise has positive governability option value only if the value of the future feasible response set exceeds its current resource cost in the specified counterfactual.

17.2 Structural design in the positive-affine model

Suppose a policy vector p changes (H, B, C, d, k). Define

$$L(p) = C(p)[I - H(p)]^{-1}B(p), \qquad \mathfrak{m}(p) = 1 - \rho(L(p)). \tag{160}$$

Three distinct design targets are useful:

$$\text{phase safety:} \quad \rho(L(p)) \le 1 - \epsilon, \tag{161}$$

$$\text{steady exposure:} \quad w^\top [I - L(p)]^{-1} \bar{k}(p) \le \bar{Q}, \tag{162}$$

$$\text{transient exposure:} \quad \mathcal{G}_T(q, g; M(p)) \le \bar{G}. \tag{163}$$

The three targets impose distinct requirements. A positive stability margin can coexist with large transient amplification.

17.3 A technology-specific response-capacity stock

For applications that benefit from a one-dimensional accounting device, define a stock G_t of technology-specific response capacity by

$$G_{t+1} = (1 - \delta_G)G_t + I_t^G - E_t, \tag{164}$$

where I_t^G is investment in future response capacity and E_t is erosion caused by entrenchment. The stock can enter k, C, the coalition cost function, or the observation system. This stock is an application-specific accounting device within the broader state-capacity literature. Besley and Persson's legal and fiscal capacities are general state capacities built by forward-looking governments (Besley and Persson, 2009); G_t here is narrower and technology-specific.

Examples of I_t^G include interoperable standards, multiple suppliers, independent public technical capacity, preserved human expertise, diversified financing, contingency infrastructure, and treaty mechanisms whose force scales with the source of future enforcement cost. Examples of E_t include sunk infrastructure, correlated collateral exposure, single-provider dependence, human deskilling, fiscal dependence, and coalition members whose restriction costs rise with adoption.

17.4 Redundancy as an economic asset

Static cost minimization often treats duplication as waste. Dynamic governability can reverse the ranking. Let a redundant alternative cost $c_R > 0$ today and lower a future entrenchment parameter from θ_0 to $\theta_1 < \theta_0$. The investment is efficient whenever

$$\beta_G \left[V_G(x'; \theta_1) - V_G(x'; \theta_0) \right] > c_R. \tag{165}$$

The welfare model determines the sign. Equation (165) is therefore the correct economic test: redundancy is justified when the value of the future choices it preserves exceeds its resource cost.

17.5 Intergenerational governability

When current actors receive the technological surplus while later cohorts inherit the altered response set, first-meter effects generate an intergenerational externality. Let generation 0 choose a_0 and generation 1 inherit $X_1 = F(X_0, a_0)$. If generation 0's objective omits generation 1's loss from a smaller $\mathcal{G}_H(X_1)$, private and social action differ even when the direct technological externality is fully priced. This gives a separate rationale for preserving reversibility: the missing object is the future generation's policy set.

18 Measurement, Identification, and Falsification

The theory becomes an empirical science when the transition and response objects are identified. This section treats measurement as a separate contribution and assigns numerical labels only where the data identify them.

18.1 State-space implementation

Let U_t be observed exogenous inputs or policy variables. A general empirical specification is

$$X_{t+1} = F_\theta(X_t, A_t, U_t) + \varepsilon_{t+1}, \tag{166}$$
$$R_t \sim O_\theta(\cdot \mid X_t), \tag{167}$$
$$\kappa_{i,t} = k_{i,\theta}(X_t, U_t) + \xi_{i,t}, \tag{168}$$
$$q_t = Q_\theta(\mu_t, \kappa_t, \mathcal{W}_t, \beta_t, \ell_t). \tag{169}$$

A local positive-affine approximation is

$$X_{t+1} = HX_t + BA_t + DU_t + \varepsilon_{t+1}, \tag{170}$$
$$A_t^{\max} = k + CX_t + EU_t + \nu_t. \tag{171}$$

When the objects are identified and $\rho(H) < 1$,

$$L = C(I - H)^{-1}B \tag{172}$$

becomes an estimable positive-affine boundary-feedback object conditional on the specification. Its interpretation as a loss of institutional governability uses the duality in Definition 2: greater future quiet capacity corresponds to a smaller effective institutional response set in the application.

18.2 Identification conditions

Transition persistence H and action loading B require variation in actions that is conditionally exogenous to the transition shock or an instrument/control-function design that restores the relevant moment conditions. The response sensitivity C requires variation that separates inherited dependence from contemporaneous severity. Binary intervention outcomes generally require additional restrictions or repeated observables to identify both latent state dynamics and a linear quiet boundary.

Proposition 14 (Loop-gain differential and Perron sensitivity)**.** *Let $R_H := (I - H)^{-1}$ and suppose $\rho(H) < 1$. For infinitesimal perturbations (dH, dB, dC),*

$$dL = (dC)R_H B + CR_H(dH)R_H B + CR_H(dB). \tag{173}$$

If, in addition, L is primitive with left and right Perron vectors $u, v \gg 0$ normalized by $u^\top v = 1$, then the first-order change in the Perron root is

$$d\rho(L) = u^\top (dL)v. \tag{174}$$

Proposition 14 is a post-identification sensitivity result. It propagates local parameter uncertainty after (H, B, C) have been identified or estimated under an application-specific research design. When the state is observed in (170), a standard sufficient benchmark for point identification of (H, B, D) is a full-rank population second-moment matrix for the regressors (X_t, A_t, U_t) together with the relevant orthogonality condition for ε_{t+1}. Identification of C additionally requires an observed or recoverable response boundary, or a response model with enough variation to separate inherited state from contemporaneous severity. Latent-state models require observability and normalization restrictions. Equations (173)–(174) then give a delta-method route from parameter uncertainty to uncertainty in the phase statistic. Eigenvalue multiplicities and partially identified specifications use set-valued or simulation-based inference.

18.3 Point, partial, and robust identification

Three empirical outputs should be distinguished.

1. *Point identified*: (H, B, C) and therefore L are uniquely determined by the maintained model and data-generating assumptions.
2. *Partially identified*: data imply a set Θ_I and therefore a set

$$\mathcal{L}_I := \{C(I - H)^{-1}B : \theta \in \Theta_I\}. \tag{175}$$

   A phase conclusion is robust only when every admissible $L \in \mathcal{L}_I$ lies on the same side of one.
3. *Structurally uncertain*: several transition or response families remain plausible. The output is then a robust viability or ambiguity analysis.

A useful empirical report therefore includes a confidence or identified interval for the margin, or a probability such as $\Pr[\rho(L) > 1 \mid \text{data, model}]$ when a Bayesian specification is justified.

18.4 Candidate identification shocks

Different mechanisms require different variation. Provider outages and forced migrations can identify persistence and substitution. Export restrictions can shift outside options and compute

dependence. Interoperability mandates can change switching costs. Mergers and divestitures can change concentration. Financing shocks can change the cost of maintaining capital-intensive infrastructure. Regulatory bans, court decisions, elections, and treaty triggers can shift the response correspondence. These events become useful identification shocks after the analyst states and defends the required exclusion restrictions.

18.5 Testable implications

The framework yields application-level predictions once a state and response technology are specified.

1. *Persistent-state channel.* An exogenous increase in technological action that loads a persistent state should shift later response feasibility even after contemporaneous technological severity is held fixed.
2. *Near-critical amplification.* Within a common measurement specification, the response of the steady quiet boundary to a given structural perturbation should increase as the spectral margin shrinks, according to Proposition 6 and the resolvent.
3. *Strategic acceleration.* When an inherited state raises the actor's continuation value through future quiet capacity, and the early-response penalty is controlled, stronger continuation incentives should raise current action relative to the myopic benchmark.
4. *Structural preservation.* Interventions that reduce persistence, action-to-state loading, or state-to-boundary sensitivity should weakly reduce the positive-affine loop gain under the order conditions in Proposition 17 below.
5. *Regime-sequence risk.* Sequences of regimes can generate larger long-run exposure than any fixed-regime analysis implies when the joint spectral radius exceeds the individual spectral radii.

These predictions require application-specific operational definitions and can be rejected by data.

18.6 Validation beyond retrodiction

Historical cases are useful for measurement design, but a model chosen after observing the outcome is vulnerable to hindsight fit. The validation sequence should therefore become progressively harder:

1. **Retrospective measurement feasibility.** Can the proposed states and response variables be measured without using the outcome to define them?
2. **Pseudo-out-of-sample tests.** Estimate the model using information available through date t, freeze the specification, and predict later governability outcomes.
3. **Prospective tests.** Publish the state definition, parameter estimates, and prediction before the outcome window begins.

18.7 Competing models and falsification

The framework earns empirical value only if it adds information beyond simpler measures. Benchmarks should include concentration, leverage, market share, sunk capital, switching-cost proxies, lobbying measures, and standard path-dependence variables. Evidence against the framework would include persistent failure of estimated feedback objects to predict later response difficulty, failure of entrenchment states to affect intervention feasibility, systematic superiority of simpler models, or policy interventions that move the proposed matrix entries without the predicted change in later governability.

18.8 A pre-analysis template for prospective work

For each application, a credible empirical protocol should pre-specify: (i) the state vector; (ii) the technological action; (iii) the institutional response outcome; (iv) the observation window; (v) the intervention horizon; (vi) the estimand L, margin, or viability set; (vii) the outcome used for validation; (viii) alternative models; and (ix) robustness to time aggregation and state definitions. This is the intended route from structural estimability to a genuinely predictive measurement program.

19 Cross-Domain Mechanism Evidence and Validation Design

A general theory should survive contact with domains that differ sharply in technology, institutions, and time scale. The cases below test whether the proposed mechanisms have empirical counterparts. Current public evidence supports mechanism-level tests; a harmonized estimate of $\rho(L)$ requires a common measurement system that these cases currently lack.

Domain	Observed first-meter mechanism	Mapping to the framework	What remains unidentified
Carbon infrastructure	Long-lived assets create persistent technology choices; plant-level evidence also links entry-period energy prices to lifetime energy intensity.	Strong persistence in H; new investment loads state through B; response-cost effects remain a separate C object.	A jointly identified response boundary C and a phase estimate under one common specification.
Banking / too-big-to-fail	In UK bank-level data from the 2007–09 crisis, bank size relative to the system had a large positive and nonlinear association with public intervention probability.	Direct evidence that the institutional response rule depends on inherited system state; size/interconnectedness can enter $q(x, a)$ or coalition cost.	Dynamic persistence and action-to-state matrices sufficient to identify L.
Digital platforms	EU policy explicitly identifies network effects, end-user lock-in, business-user dependence, and interoperability as contestability mechanisms.	Switching costs and network effects enter H and cross-couplings; interoperability can reduce persistence or response sensitivity.	A mature post-policy outcome window and a causal estimate of the full feedback loop.
Ozone-depleting substances	The Montreal Protocol has phased out nearly 99% of controlled ozone-depleting substances and has repeatedly adjusted controls as science and technology changed.	A contrasting case in which precommitment, measurement, substitutes, differentiated schedules, and adaptive rules preserved the feasibility of stronger later action.	A historical L constructed under a pre-specified state model; this case supplies mechanism evidence for preserved governability.

Table 7: Cross-domain mechanism evidence. The table deliberately separates evidence for state-dependent response and persistence from a validated loop-gain estimate.

19.1 Carbon infrastructure

Carbon lock-in research describes mutually reinforcing technological, institutional, and behavioral constraints on decarbonization (Seto et al., 2016). Using detailed infrastructure data, Tong et al. (2019) estimated that fossil-fuel infrastructure existing in 2018 would emit about 658 $GtCO_2$ if operated as historically, with proposed power plants adding a central estimate of about 188 $GtCO_2$. More directly, Hawkins-Pierot and Wagner (2026) use U.S. Census microdata and quasi-experimental variation in electricity prices to show persistent plant-level technology lock-in: plants entering under lower electricity prices remain more energy intensive, and a 10% increase in entry-year electricity prices implies about a 3% reduction in lifetime energy intensity. Together, these results give strong evidence for persistent state and action-loading channels. Identification of how those states change the political or institutional cost of later intervention remains a separate response-side problem. In the present notation, the evidence therefore disciplines H and B more directly than C.

19.2 Banking and endogenous rescue incentives

The too-big-to-fail mechanism is unusually close to the response side of the framework. Using bank-level data for the United Kingdom during the 2007–09 crisis, Rose and Wieladek (2012) find that bank size relative to the system had a large positive and nonlinear effect on the probability of public intervention. Here intervention takes the form of rescue; the structural point is the same: inherited economic scale changes the institution's feasible or preferred response. A full first-meter estimate would additionally require the dynamic law by which earlier balance-sheet choices created the systemically important state.

19.3 Digital platforms and interoperability

The European Union's Digital Markets Act identifies gatekeeper power with strong network effects, economies of scale, end-user lock-in, business-user dependence, and vertical integration. Its interoperability requirements are intended to let third parties access platform features on comparable terms and thereby improve contestability (European Commission, 2026). In the current framework, portability and interoperability are candidate interventions on persistence H, action-to-dependence loadings B, or boundary sensitivity C. Because the regulatory regime is recent and still evolving, this case currently supports mechanism mapping and policy-design analysis.

19.4 The Montreal Protocol as a preservation case

A general theory needs successes as well as failures. The Montreal Protocol provides a strong example of an institution that retained the ability to tighten action over time. The UN-backed 2022 scientific assessment reports that nearly 99% of banned ozone-depleting substances have been phased out (UNEP, 2023). The regime also evolved through repeated adjustments and amendments as scientific knowledge and technology changed. The case is consistent with several governability-preserving mechanisms: common measurement, pre-agreed control schedules, differentiated participation rules, financial support for transition, and technological substitution. It also illustrates strategic feedback: historical work on environmental regulation notes that firms can invest to influence later policy, including the CFC-substitution context (Tarui and Polasky, 2005). The case is consistent with the mechanism that substitute technologies and a credible adaptive response architecture preserve

future intervention feasibility as the regulated economy changes.

19.5 What a genuine cross-domain validation would require

The next empirical step is to define comparable within-domain measurement systems, estimate them before the validation window, and test whether phase, margin, and amplification improve prediction of later response difficulty. Repeated predictive success and policy-counterfactual performance would establish portability across applications.

20 Application to Frontier AI

Status of the application. The results in this section map observed and hypothesized AI mechanisms into the theory. Estimation of the current global value of $\rho(L)$ remains a separate empirical task. The scalar phase theorem is exact conditional on its tracker and response assumptions; data determine which regime describes the real system.

20.1 The current institutional setting

The international AI-governance setting increasingly separates measurement from action. The United Nations established an Independent International Scientific Panel on AI and a recurring Global Dialogue on AI Governance. The Panel released its preliminary report on 1 July 2026, and the inaugural Global Dialogue was held in Geneva on 6–7 July 2026; the report supplied a common scientific starting point for the discussions (United Nations Independent International Scientific Panel on AI, 2026; United Nations, 2026). A shared assessment can improve information. Coalition choice, cost allocation, and implementation remain separate institutional decisions. The coalition microfoundation models that downstream problem.

20.2 Compute and energy

Compute is both an input to frontier AI and a possible governance lever (Sastry et al., 2024). It is also becoming tied to large physical infrastructure. The IEA's base case projects data-center electricity consumption to more than double by 2030, with AI the largest driver of the increase (IEA, 2025). In the model, this can enter through several state variables: dependence of firms on compute services, electricity infrastructure committed to data centers, geographic concentration, and the opportunity cost of restricting facilities.

The equation permits either sign for these effects on intervention costs. Public compute, interoperability, excess grid capacity, or multiple providers can reduce dependence. The sign and magnitude are empirical.

20.3 Financial entanglement

The BIS documents increasing on- and off-balance-sheet borrowing for AI infrastructure and new links among hyperscalers, private-credit vehicles, insurers, and banks (Eren, Krohn, and Todorov, 2026). The Bank of England reports a broader expansion into public debt, private credit, leveraged

finance, and structured finance, while emphasizing that immediate stocks of outstanding AI debt were still modest at the beginning of 2026 and that evidence of crowding out remained limited at the time of its July report (Bank of England, 2026).

This motivates two distinct state channels. Financial entrenchment can raise the cost of intervention if restrictions would impair collateral values, debt service, pension assets, or financial intermediaries. The same exposure can also raise severity by creating a correlated shock channel. In the general system, one mechanism therefore changes both the response boundary and the loss function.

20.4 Rebound demand and continual use

A fall in the cost of machine cognition can increase total demand for cognition and total compute use. Recent economic modeling explicitly derives a structural Jevons effect in digital-intelligence demand (Zhang and Zhang, 2026). If lower prices cause agents to perform additional tasks that were previously uneconomic, technological action can accelerate. If the resulting use also embeds AI in workflows and data loops, it changes both the action path and the state transition matrix.

This distinction matters. Faster y_t alone compresses calendar time; the phase boundary changes when the structural feedback parameters change. Entrenchment changes the phase boundary only when it increases future quiet room or response cost.

20.5 Platforms, action histories, and organizational dependence

Agent-mediated platforms can create new switching costs and deepen incumbent control over transaction surfaces (Kapoor, Kolt, and Lazar, 2025). AI lock-in can also arise from loss of human capability or national dependence on AI services (Kim et al., 2026). Action-history data can create a feedback loop in which more deployment generates more intent-action-outcome data, which improves the system and attracts more deployment.

These mechanisms are naturally modeled as state persistence and cross-coupling. A portability requirement can lower entries of H; a provider-independent agent standard can reduce entries of C; a stronger data flywheel can increase cross-couplings that return adoption into future dependence.

20.6 Human capability and substitution capacity

Loss of human capability can create a separate exit-cost channel. One stylized state is

$$h_{t+1} = (1 - \delta_h)h_t + I_t^h - \xi A_t, \tag{176}$$

where h_t is retained human or organizational capability, I_t^h is investment in that capability, and ξA_t captures substitution or skill atrophy associated with deployment. A response-cost specification such as

$$\kappa_i(X_t) = \widetilde{\kappa}_i(X_t) + \chi_i(\bar{h} - h_t)_+ \tag{177}$$

maps capability loss into the cost of switching away from the technology. The parameters require empirical work on task substitution, recovery after outages, and the cost of maintaining parallel human systems. This channel gives preserved expertise a direct governability value.

20.7 Fiscal and political dependence

Governments can also acquire direct stakes in continued deployment through tax revenue, employment, strategic procurement, or politically organized beneficiaries. Let $f_{i,t}$ measure fiscal exposure and $\ell_{i,t}$ a measurable political-dependence state. A local coalition-cost equation can include

$$\kappa_i(X_t) = \kappa_i^0 + \gamma_i^f f_{i,t} + \gamma_i^\ell \ell_{i,t} + \text{other state terms.} \tag{178}$$

Positive coefficients represent channels through which restriction becomes more costly as dependence rises. Negative coefficients can arise when adoption builds regulatory expertise, tax capacity, or political support for stronger oversight. The signs and magnitudes are empirical. These variables enter C or the coalition threshold and create a direct link to fiscal dependence and regulatory capture hypotheses.

20.8 Compute balance of payments and sovereign dependence

If AI services become a major intermediate input, countries can differ along at least four dimensions: domestic frontier capability, dependence on imported AI services, control of critical physical inputs, and exposure to AI risk. A state with modest model capability can still have bargaining power through energy, minerals, markets, data-center sites, or scientific talent. A state that depends heavily on a foreign provider can face high domestic costs from restrictions on that provider.

The coalition threshold should therefore be estimated from state-specific costs and outside options, with development status used only when it carries identified information about those primitives.

20.9 Correlated model errors

Shared models can reduce average error while increasing correlation in residual errors across institutions. Correlation matters for systemic loss. If banks, hospitals, grids, or government agencies use closely related models, a common failure mode can generate a different severity mapping from independent human or model errors.

This mechanism enters the left side of the governance problem: the effective loss from continued deployment. Financial and organizational entrenchment can simultaneously raise the right side: the cost of intervention. The relevant empirical question is which side moves faster.

21 Beyond Linearity: Saturation, Hysteresis, and Computed Viability

The positive-affine model is valuable because it gives exact global formulas. Its phase condition applies exactly to the positive-affine model; nonlinear systems use the local Jacobian and global nonlinear structure. The local Jacobian test in Proposition 2 is the correct first-order object near a fixed point, while global nonlinearities can either worsen or cap the feedback.

21.1 A saturating entrenchment example

Consider the benchmark tracker

$$m_t = (1-\alpha)m_{t-1} + \alpha y_t \tag{179}$$

and one entrenchment tracker

$$x_t = (1-\delta)x_{t-1} + \delta y_t. \tag{180}$$

Replace the linear intervention-cost term γx_{t-1} by a bounded monotone function $\varphi(x_{t-1})$, so that the quiet boundary is

$$y_t \le K_0 + (1-\eta)m_{t-1} + \varphi(x_{t-1}), \qquad 0 \le \varphi(x) \le \bar{\varphi}. \tag{181}$$

Proposition 15 (Saturation can restore a global bound)**.** *Suppose $\eta > 0$, $\alpha, \delta \in (0,1]$, and φ satisfies (181). Then every boundary path is globally bounded. In particular,*

$$\limsup_{t\to\infty} m_t \le \frac{K_0 + \bar{\varphi}}{\eta}, \tag{182}$$

and y_t and x_t are bounded as well.

The result matters because a local affine approximation of the entrenchment channel near low x can have slope $\varphi'(0) > \eta$ and therefore produce a supercritical affine comparator, while bounded global feedback still caps the trajectory. Local stability around a nonlinear fixed point uses the derivative at that fixed point as in Proposition 2. Convex or threshold feedback can create the opposite pattern, with mild local feedback and severe global amplification. The linear Γ versus η boundary is exact for the stated tracker model and serves as a local or approximate guide outside it.

A concrete specification used in the numerical appendix is

$$\varphi(x) = \frac{\gamma x}{1 + x/s}. \tag{183}$$

With $K_0 = 1$, $\eta = 0.35$, $\gamma = 0.60$, and $s = 3$, the local entrenchment slope exceeds anchor retention ($\gamma > \eta$), while the nonlinear boundary trajectory converges because $\varphi(x) \le 1.8$.

21.2 Hysteresis

A first-meter mechanism can be irreversible even when the technological action itself reverses. Formally,

$$F(F(x,+a),-a) \neq x. \tag{184}$$

Sunk infrastructure, skill loss, legal precedent, and political coalition changes can all create this property. In a nonlinear model, hysteresis implies that a policy that returns the technological variable to its old level can leave the governability state on a different recovery path. Empirical work should therefore measure recovery paths as well as forward paths.

21.3 Viability recursion and institutional response timing

Let K be a state set in which a specified minimum level of institutional response remains feasible. Write a deterministic transition as

$$x' = F(x,a,u), \tag{185}$$

where $a \in A(x)$ is technological action and u is an institutional preservation or response action. Timing determines the correct viability quantifiers. If the institution responds after observing a, define

$$V_0^{\text{post}} := K, \tag{186}$$

$$V_{t+1}^{\text{post}} := \left\{ x \in K : \forall a \in A(x)\ \exists u \in U(x,a) \text{ such that } F(x,a,u) \in V_t^{\text{post}} \right\}. \tag{187}$$

If the institution must precommit before observing a, use

$$V_0^{\text{pre}} := K, \tag{188}$$

$$V_{t+1}^{\text{pre}} := \left\{ x \in K : \exists u \in U(x)\ \forall a \in A(x) \text{ such that } F(x,a,u) \in V_t^{\text{pre}} \right\}. \tag{189}$$

Transition ambiguity adds a universal quantifier over $F \in \mathfrak{F}$ at the relevant predecessor step. Stochastic systems replace deterministic membership with the required conditional probability.

Proposition 16 (Response-timing monotonicity)**.** *Suppose the precommitment action set $U(x)$ is a subset of the actions available after every observed a, so $U(x) \subseteq U(x,a)$ for all a. Then*

$$V_t^{\text{pre}} \subseteq V_t^{\text{post}} \qquad \textit{for every finite horizon } t. \tag{190}$$

The proposition states the option value of observing technological action before choosing the institutional response. Grid methods, level-set methods, or approximate dynamic programming can compute either kernel when closed forms fail. An application-level computation requires a stated state/action discretization, control timing, interpolation rule, and convergence study. The accompanying reproducibility scripts audit the saturating boundary example; a calibrated nonlinear viability kernel remains an application-specific empirical exercise.

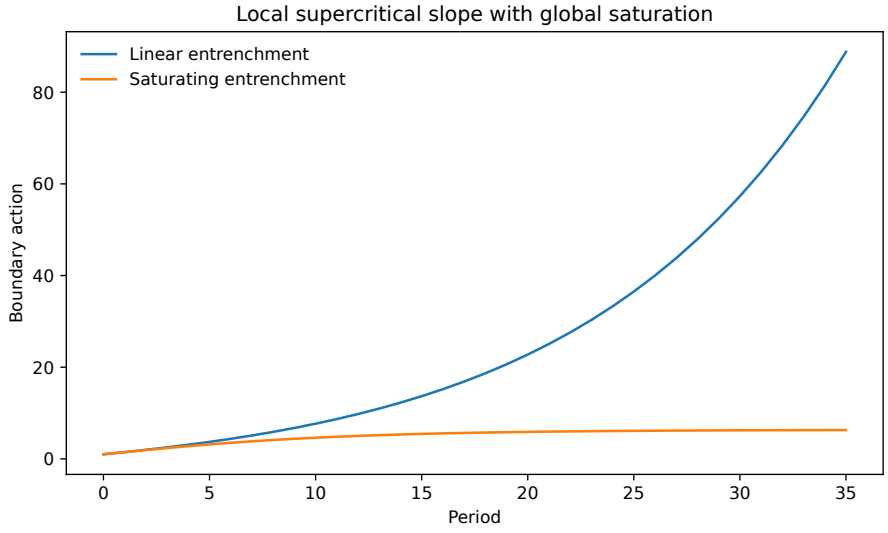


Figure 4: Boundary iteration for the same local entrenchment slope under linear and saturating feedback. Parameters are $K_0 = 1$, $\eta = 0.35$, $\gamma = 0.60$, $\alpha = 0.50$, $\delta = 0.25$, and $s = 3$. The linear model is supercritical because $\gamma > \eta$; the saturating model is globally bounded by Proposition 15.

22 Policy Design for Preserving Future Governability

The framework separates current restriction from investment in future optionality. A policy can allow present deployment while changing the transition system so that future exit, substitution, or coordinated intervention remains feasible.

22.1 A general structural-design problem

Let policy p change (H, B, C, d, k) and let $\mathcal{C}(p)$ be its cost. A natural design problem is

$$\min_{p} \ \mathcal{C}(p) \quad \text{subject to} \quad \rho\Big(C(p)[I - H(p)]^{-1}B(p)\Big) \leq 1 - \epsilon \tag{191}$$

for a chosen margin $\epsilon > 0$.

A stronger target can constrain the steady boundary itself:

$$w^{\top}[I - L(p)]^{-1}\bar{k}(p) \leq Q. \tag{192}$$

Equation (191) preserves boundedness; equation (192) controls the level of the bound.

Policies can act on different matrices. Portability and easier substitution can reduce persistence in H. Public alternatives and supplier diversity can reduce the effect of one technological action on dependence, lowering B. Strong cumulative accounting can reduce the sensitivity of future quiet room to adapted benchmarks, lowering relevant entries of C. Baseline enforcement capacity can change k. Independent scientific and administrative capacity can change both H and C.

Proposition 17 (Structural preservation dominance)**.** *Consider two positive-affine specifications indexed by $r \in \{0, 1\}$ with $\rho(H_r) < 1$. Suppose specification 1 is structurally preserving relative to specification 0 in the componentwise order:*

$$0 \leq H_1 \leq H_0, \qquad 0 \leq B_1 \leq B_0, \qquad 0 \leq C_1 \leq C_0, \tag{193}$$

and $0 \leq d_1 \leq d_0$, $0 \leq k_1 \leq k_0$. Then

$$L_1 = C_1(I - H_1)^{-1}B_1 \leq C_0(I - H_0)^{-1}B_0 = L_0, \tag{194}$$

so $\rho(L_1) \leq \rho(L_0)$. If $\rho(L_0) < 1$, both systems are subcritical and their steady boundary actions satisfy

$$a_{\infty,1} \leq a_{\infty,0}. \tag{195}$$

This gives a sufficient policy ordering with a direct interpretation. Lower state persistence, lower action-to-state loading, and lower state-to-boundary sensitivity weakly reduce the feedback loop; lower baseline forcing and quiet intercept weakly reduce the steady boundary as well.

22.2 A governability diagnostic set

For an estimated positive-affine model, four quantities answer different policy questions:

Object	Formula	Question
Phase	$\rho(L)$	Does repeated feedback decay, sit at the critical boundary, or expand?
Spectral margin	$1-\rho(L)$	How far is the positive-affine model from its phase boundary in spectral terms?
Steady boundary amplification	$(I-L)^{-1}$	How strongly does repeated feedback amplify a baseline boundary or persistent structural forcing?
Finite-horizon amplification	$\mathcal{A}_T(M)$ or an output-specific analogue	How severe can temporary loss of governability become over the policy horizon?

Each estimate should be reported with sampling or identification uncertainty. A decision rule can then use a safety margin, an exposure ceiling, and a finite-horizon constraint together.

22.3 Scalar specialization

In the benchmark-entrenchment model, suppose policy can increase anchor retention by $r \in [0, 1-\eta]$ and reduce entrenchment sensitivities by $s_j \in [0, \gamma_j]$. Then the stable ceiling is

$$\frac{K_0}{\eta + r - \left(\Gamma - \sum_j s_j\right)}. \tag{196}$$

To guarantee a target $Q \geq K_0$, the exact requirement is

$$r + \sum_j s_j \geq \Gamma - \eta + \frac{K_0}{Q}. \tag{197}$$

The scalar formula is therefore one projection of the matrix design problem.

22.4 Redundancy and coalition option value

Redundancy can look inefficient under static cost minimization. The dynamic model gives it option value. Alternative compute providers, independent infrastructure, multiple scientific institutions, or several possible international coalitions can keep future intervention costs from being tied to one highly entrenched system. In the heterogeneous coalition model, a coalition with a higher initial cost can determine the long-run growth class when its entrenchment slope is lower.

22.5 Response protocols can erode

The protocol condition in Theorem 6 is state dependent. Suppose a treaty gives designated government i a fixed nonperformance cost ψ_i after a certified trigger. If current normalized enforcement cost is $\kappa_i(x)$, the sufficient dominance condition holds only while

$$\psi_i > \kappa_i(x). \tag{198}$$

A fixed protocol can therefore lose force as economic dependence grows. Robust precommitment must be indexed to the future source of enforcement cost or be paired with policies that bound that cost.

23 Discussion

23.1 Scope of generalization

The framework is deliberately more general than the AI application, with layered results matched to the structure of each mechanism. Generality is layered.

At the stochastic level, state and kernel extension represents a broad class of mechanisms, while closed-form results require added structure. At the monotone level, the boundary operator gives a global comparison theorem. At the positive affine level, exact finite-horizon expressions and spectral criteria become available. At the scalar first-meter level, the phase boundary becomes $\Gamma = \eta$.

A future mechanism can therefore "plug in" after its mathematical properties are stated. A mechanism outside monotonicity uses the general stochastic layer; a nonlinear mechanism uses the local Jacobian and global nonlinear analysis; discontinuous regime switching uses the joint spectral radius or a set-valued viability analysis. The assumptions therefore identify the applicable conclusions.

23.2 Irreversibility and hysteresis

First-meter problems are naturally related to hysteresis. A state can fail to retrace its path when the technological action is reversed:

$$F(F(x, +a), -a) \neq x. \tag{199}$$

Human deskilling, sunk infrastructure, political coalitions, and legal precedents can all generate this property. The general state model admits hysteresis directly. Linear tracker models are only one approximation.

23.3 Time inconsistency and commitment

A government can prefer to preserve future governability ex ante and later prefer acquiescence after dependence has accumulated. That is a commitment problem. Pre-agreed protocols, independent regulators, automatic legal consequences, or financial exposure limits can be interpreted as devices that alter the future response correspondence before the high-dependence state is reached.

23.4 The option value of early governance

The model gives a reason for early governance even when current substantive restriction is unwarranted. Early policy can target the transition law rather than the current technological level. It can preserve interoperability, alternative providers, independent expertise, diversified finance, and cumulative accounting. The value comes from keeping future actions feasible.

This separates precaution from optionality. A policymaker can prefer continued technological development today while also investing in future exit and response capacity.

23.5 Empirical scope of the frontier-AI interpretation

A global frontier-AI estimate of $\rho(L)$ is a separate empirical task and must be tied to a specified state definition, time scale, action variable, response boundary, and identification design. Sovereign-debt effects, recursive self-improvement, frontier-compute concentration, and the other motivating scenarios remain empirical or forecasting hypotheses. Their role here is to identify candidate state variables, transitions, and response channels.

The formal result is conditional. When technological movement changes state variables that make later response more costly or later quiet movement easier, those feedbacks can be represented and measured. A measured loop gain below one gives a bounded positive-affine model. A loop gain above one gives unstable boundary dynamics. Nonlinear, switched, partially observed, and structurally uncertain systems use the corresponding higher-level framework.

24 Conclusion

Current technological action can change the future conditions under which institutions act. That is the first-meter mechanism. The contribution of this monograph is to make that mechanism explicit at several levels of structure and to state the assumptions needed at each level.

At the most general level, state-dependent transition and response objects define risk-constrained capacity and institutional continuation sets. Under monotonicity, one boundary operator generates the maximal quiet path. Under positive-affine structure, the central feedback object is

$$L = C(I - H)^{-1}B, \tag{200}$$

which separates state persistence, action-to-state loading, and the return from inherited state into future quiet room. Its Perron root gives the positive-affine phase; $1 - \rho(L)$ gives the spectral gap to that phase boundary; $(I - L)^{-1}$ gives steady boundary amplification; and transient and switching diagnostics capture risks that a single asymptotic eigenvalue can miss.

The strategic layer adds anticipation. A technological actor can value current movement because it improves its future state, while an institution can value preservation because it improves its future response set. The exact two-period benchmark shows one parameter region in which both incentives rise together. The welfare layer evaluates preservation through a committed-policy counterfactual: redundancy, interoperability, independent expertise, and substitute infrastructure have value when the future options they preserve exceed their current cost.

The empirical burden is equally explicit. Transition data can discipline H and B; response-side variation is needed to discipline C; latent-state applications require observability and normalization restrictions. Historical evidence in carbon, banking, digital platforms, and ozone control establishes candidate persistence and response mechanisms. A phase estimate requires those mechanisms to be identified jointly under one specification. The proposed frozen-specification, pseudo-out-of-sample, and prospective tests expose the framework to failure rather than rewarding retrospective fit.

For the scalar frontier-AI specialization,

$$L = 1 - \eta + \Gamma, \tag{201}$$

so the exact positive-affine phase boundary is $\Gamma = \eta$ and the subcritical ceiling is $K_0/(\eta - \Gamma)$. The result is a conditional theorem about a stated benchmark-plus-entrenchment model. Its empirical

use depends on measuring anchor retention and the channels through which AI dependence changes future intervention cost.

Dynamic governability therefore adds a second object to technology policy analysis: alongside the value of present technological progress sits the value of preserving future freedom to respond. The mathematical framework supplies a way to represent that object, derive exact results under transparent structure, and specify the evidence needed to measure it.

A Proofs

A.1 Proof of Theorem 1

Let $x_0 = x_0^*$. Suppose inductively that $x_t \le x_t^*$. A weakly quiet path has

$$a_t \le B(x_t) \le B(x_t^*) \tag{202}$$

by isotonicity of B. Since F is order-preserving in both arguments,

$$x_{t+1} = F(x_t, a_t) \le F(x_t^*, B(x_t^*)) = x_{t+1}^*. \tag{203}$$

Thus $x_t \le x_t^*$ for every t. Monotonicity of Y gives

$$Y(x_T) \le Y(x_T^*) = Y(\mathcal{T}^T(x_0)). \tag{204}$$

The boundary path is itself weakly quiet, so the maximum is attained.

For strict quietness, define

$$\mathcal{T}_\varepsilon(x) := F(x, (1-\varepsilon)B(x)). \tag{205}$$

For every finite T, continuity implies

$$\mathcal{T}_\varepsilon^T(x_0) \to \mathcal{T}^T(x_0) \tag{206}$$

as $\varepsilon \downarrow 0$. Every $\varepsilon > 0$ path is radially strict by construction. Hence the radially strict supremum equals the weak maximum. □

A.2 Proof of Proposition 1

A finite Cartesian product of standard Borel spaces is standard Borel, and a countable disjoint union of standard Borel spaces is standard Borel. Hence the history space

$$\mathsf{H} := \bigsqcup_{t \ge 0} (\mathsf{X} \times \mathsf{A})^t \times \mathsf{X} \tag{207}$$

is standard Borel. Given the current history H_t and current action a_t, the original history-dependent kernel determines the distribution of x_{t+1}; appending (a_t, x_{t+1}) gives a Markov transition kernel on H. Thus the augmented process satisfies a Markov law of the form (3). □

A.3 Proof of Proposition 2

This is the standard discrete-time linearization theorem. If every eigenvalue of $D\mathcal{T}(x^*)$ has modulus below one, an equivalent norm makes the derivative a local contraction and x^* is locally asymptotically stable. If an eigenvalue has modulus above one, there is a locally expanding direction and the fixed point is unstable. At spectral radius one, higher-order terms determine the result in general. □

A.4 Proof of Theorem 2

Under the weak boundary, Theorem 1 implies boundary saturation. Substitution of $a_t^* = k + Cx_{t-1}^*$ into (40) gives

$$x_t^* = (H + BC)x_{t-1}^* + (d + Bk) = Mx_{t-1}^* + g. \tag{208}$$

Iteration gives (45). Equation (46) follows from the boundary equation. Nonnegative w preserves the order comparison. The strict case follows from Theorem 1. □

A.5 Proof of Theorem 3

Let

$$M = H + BC, \qquad L = C(I - H)^{-1}B. \tag{209}$$

Because $H \geq 0$ and $\rho(H) < 1$, $(I - H)^{-1} \geq 0$.

First suppose $\rho(L) < 1$. Then $(I - L)^{-1} \geq 0$. The Woodbury identity gives

$$(I - M)^{-1} = (I - H - BC)^{-1} \tag{210}$$

$$= (I - H)^{-1} + (I - H)^{-1}B(I - L)^{-1}C(I - H)^{-1} \geq 0. \tag{211}$$

For a nonnegative matrix, existence of a nonnegative inverse $(I - M)^{-1}$ is equivalent to $I - M$ being a nonsingular M-matrix, hence $\rho(M) < 1$.

Conversely, suppose $\rho(L) \geq 1$. If $\rho(L) = 1$, choose a nonzero Perron vector $v \geq 0$ with $Lv = v$ and set

$$x = (I - H)^{-1}Bv. \tag{212}$$

Then $x \neq 0$ and

$$Mx = Hx + BCx \tag{213}$$

$$= Hx + BLv \tag{214}$$

$$= Hx + Bv \tag{215}$$

$$= x, \tag{216}$$

so 1 is an eigenvalue of M and $\rho(M) \geq 1$.

If $\rho(L) > 1$, for $s > \rho(H)$ define

$$L(s) := C(sI - H)^{-1}B. \tag{217}$$

The entries of $L(s)$ are continuous and nonincreasing in s, $\rho(L(1)) > 1$, and $\rho(L(s)) \to 0$ as $s \to \infty$. Hence there exists $s_* > 1$ with $\rho(L(s_*)) = 1$. Choose $v \geq 0$, $v \neq 0$, with $L(s_*)v = v$, and set

$$x = (s_*I - H)^{-1}Bv. \tag{218}$$

Then

$$(H + BC)x = s_*x. \tag{219}$$

Thus $\rho(M) \geq s_* > 1$. This proves the equivalence.

For the steady state, write

$$x_\infty = (I - H)^{-1}(d + Ba_\infty). \tag{220}$$

The saturated boundary gives

$$a_\infty = k + Cx_\infty \tag{221}$$
$$= k + C(I-H)^{-1}d + C(I-H)^{-1}Ba_\infty \tag{222}$$
$$= \bar{k} + La_\infty. \tag{223}$$

Since $\rho(L) < 1$, $I - L$ is invertible with nonnegative inverse, giving (51); (52) follows. □

A.6 Proof of Proposition 3

Since

$$I - SHS^{-1} = S(I-H)S^{-1}, \tag{224}$$

we have

$$(I-H')^{-1} = S(I-H)^{-1}S^{-1}. \tag{225}$$

Therefore

$$C'(I-H')^{-1}B' = CS^{-1}S(I-H)^{-1}S^{-1}SB \tag{226}$$
$$= C(I-H)^{-1}B. \tag{227}$$

An action-coordinate transformation $a' = Ra$ produces $L' = RLR^{-1}$, which is similar to L and therefore has the same spectral radius. □

A.7 Proof of Proposition 4

Let

$$S_h := I + H + \cdots + H^{h-1}. \tag{228}$$

Then $B_h = S_h B$ and, because S_h is a polynomial in H,

$$I - H^h = S_h(I-H) = (I-H)S_h. \tag{229}$$

Both $I - H$ and $I - H^h$ are invertible under $\rho(H) < 1$. Therefore

$$(I-H^h)^{-1}S_h = (I-H)^{-1}, \tag{230}$$

and left multiplication by C and right multiplication by B give (56). □

A.8 Proof of Theorem 4

Put $\widehat{M}$ into Jordan normal form over $\mathbb{C}$:

$$\widehat{M} = SJS^{-1}. \tag{231}$$

For a Jordan block $J_\lambda = \lambda I + N$ of size r,

$$J_\lambda^T = \sum_{j=0}^{r-1} \binom{T}{j} \lambda^{T-j} N^j. \tag{232}$$

For $\lambda \neq 0$, factor λ^T and absorb the remaining constants and binomial coefficients into a polynomial of degree at most $r-1$. For $\lambda = 0$, nilpotence gives $J_0^T = 0$ for every $T \geq r$, so this block contributes only a finite transient. Multiplication by the fixed vectors $\widehat{q}$ and $\widehat{x}_0$ yields the stated representation. If either vector annihilates a generalized eigenspace, its coefficient is zero. The dominant nonzero terms determine the asymptotic envelope. □

A.9 Proof of Theorem 5

If $\widehat{\rho}(\mathcal{M}) < 1$, choose γ with $\widehat{\rho}(\mathcal{M}) < \gamma < 1$. Standard joint-spectral-radius theory gives a constant C_γ such that every length-t product satisfies

$$\|M_t \cdots M_1\| \le C_\gamma \gamma^t. \tag{233}$$

With $\|g_\sigma\| \le G$,

$$\|x_T\| \le C_\gamma \gamma^T \|x_0\| + \sum_{r=0}^{T-1} C_\gamma \gamma^r G \tag{234}$$

$$\le C_\gamma \|x_0\| + \frac{C_\gamma G}{1-\gamma}, \tag{235}$$

which is uniform in the switching sequence.

If $\widehat{\rho}(\mathcal{M}) > 1$, the definition gives products with exponential growth rate greater than one along a subsequence. Reachability and observability determine whether those products enter the scalar capacity. At equality, the JSR gives the exponential rate while polynomial growth can arise from finer product structure. □

A.10 Proof of Proposition 7

For a unilateral change by government i,

$$u_i(e_i', e_{-i}; z, x) - u_i(e_i, e_{-i}; z, x) \tag{236}$$

$$= -\kappa_i(x)(e_i' - e_i) + z[S(e_i', e_{-i}) - S(e_i, e_{-i})], \tag{237}$$

which equals the change in Φ. Hence the game is an exact potential game. Under asynchronous log-linear learning, the stationary distribution is proportional to $\exp(\beta\Phi)$. Among losing profiles, $e = 0$ uniquely maximizes the potential with value $-z$. Among winning profiles, the potential is maximized by coalitions with total cost $K(x)$ and has value $-K(x)$. Comparing $-z$ and $-K(x)$ gives the three cases. □

A.11 Proof of Proposition 8

For a unilateral change by government i,

$$\widetilde{u}_i(e_i', e_{-i}; z, x) - \widetilde{u}_i(e_i, e_{-i}; z, x) = -c_i(x)(e_i' - e_i) + \lambda_i(x) z [S(e_i', e_{-i}) - S(e_i, e_{-i})] \tag{238}$$

$$= \lambda_i(x) \left[\Phi(e_i', e_{-i}; z, x) - \Phi(e_i, e_{-i}; z, x)\right], \tag{239}$$

where $\kappa_i(x) = c_i(x)/\lambda_i(x)$. This is the weighted-potential identity. Division of player i's payoff by the positive constant $\lambda_i(x)$ preserves that player's best responses, so the Nash-equilibrium set equals the normalized game's equilibrium set and the winning-coalition threshold is (84). If $\widetilde{\beta}_i = \beta/\lambda_i(x)$, then

$$\exp\{\widetilde{\beta}_i \widetilde{u}_i(e; z, x)\} = \exp\{\beta u_i(e; z, x)\},$$

so each primitive logit update equals the normalized common-β update. □

A.12 Proof of Proposition 9

Conditional on the common report, replace z in the complete-information game by $\mu(r)$. If the other government acquiesces, unilateral contribution cannot make unanimity succeed, so contribution adds cost κ and universal acquiescence is a best-response profile for every report. If the other government contributes, contribution gives payoff $-\kappa$ while acquiescence gives $-\mu(r)$, so joint enforcement is a Bayesian Nash equilibrium exactly when $\mu(r) \geq \kappa$. The conditional expected-payoff game is the exact-potential game of Proposition 7 with $K = 2\kappa$. Its log-linear stochastic-selection threshold is therefore 2κ. □

A.13 Proof of Theorem 6

Fix a designated member $i \in C^\star$ while the trigger is active. For any e_{-i}, the payoff difference between contribution and acquiescence is

$$u_i^P(1, e_{-i}; z, x) - u_i^P(0, e_{-i}; z, x) = \psi_i - \kappa_i(x) + z\,[S(1, e_{-i}) - S(0, e_{-i})]\,. \tag{240}$$

Monotonicity of S makes the last term nonnegative. Under (87), the difference is strictly positive for every e_{-i}, so contribution is strictly dominant for each designated member. Therefore every equilibrium contains all of $C^\star$. Because $C^\star$ is winning, response then succeeds regardless of outsider actions. Every outsider strictly prefers acquiescence because contribution adds positive cost without changing the successful outcome. The stated profile is therefore the unique Nash equilibrium. □

A.14 Derivation of Equation 91

The Gibbs weight of winning profile e is $\exp[-\beta c_x(e)]$. The Gibbs weight of losing profile e is $\exp[-\beta(c_x(e) + z)] = e^{-\beta z} \exp[-\beta c_x(e)]$. Summing over winning and losing profiles gives (91). □

A.15 Proof of Theorem 8

The exact finite formula follows from Theorem 2 with the stated specialization and $s_0 = 0$. The saturated sequence is feasible for the original monotone-movement constraint: each tracker is a convex combination of its previous value and the current boundary level, so $0 \leq s_{t-1,j} \leq y_t$; the boundary level $y_{t+1} = K_0 + c^\top s_t$ is therefore weakly increasing from $y_1 = K_0$.

The loop gain is (127). Therefore Theorem 3 gives

$$\rho(M) < 1 \quad \Longleftrightarrow \quad 1 - \eta + \Gamma < 1 \quad \Longleftrightarrow \quad \Gamma < \eta. \tag{241}$$

In the stable case, (51) gives

$$Q_\infty = \frac{K_0}{1 - (1 - \eta + \Gamma)} = \frac{K_0}{\eta - \Gamma}. \tag{242}$$

At criticality $L = 1$. Since $H = I - D$ has spectral radius below one, the matrix determinant lemma reduces the feedback eigenvalues outside $\sigma(H)$ to roots of $1 - c^\top (sI - H)^{-1} B = 0$. At $s = 1$ the root is simple because

$$\frac{d}{ds} c^\top (sI - H)^{-1} B \Big|_{s=1} = -c^\top (I - H)^{-2} B < 0. \tag{243}$$

Let

$$v = (I - H)^{-1}B = D^{-1}D\mathbf{1} = \mathbf{1}. \tag{244}$$

A left eigenvector $w^\top M = w^\top$ normalized by $w^\top v = 1$ is

$$w^\top = \frac{c^\top (I - H)^{-1}}{c^\top (I - H)^{-2} B} = \frac{c^\top D^{-1}}{c^\top D^{-1}\mathbf{1}}. \tag{245}$$

The row-stochastic critical matrix M is power-bounded, so Cesàro convergence gives

$$\frac{1}{T}\sum_{r=0}^{T-1} M^r \to v w^\top. \tag{246}$$

Hence

$$\lim_{T\to\infty} \frac{Q_T}{T} = c^\top v\, w^\top g \tag{247}$$

$$= 1 \cdot \frac{c^\top D^{-1}}{c^\top D^{-1}\mathbf{1}} K_0 D\mathbf{1} \tag{248}$$

$$= \frac{K_0}{c^\top D^{-1}\mathbf{1}}. \tag{249}$$

If $\Gamma > \eta$, Theorem 3 gives $\rho(M) > 1$. The nonnegative rank-one feedback construction reaches and observes its Perron mode through $g = K_0 D\mathbf{1}$ and $c \neq 0$. Perron–Frobenius theory therefore gives geometric asymptotic growth rate $\rho(M)$. □

A.16 Proof of Equation 138

With $J = 0$, the recursion is

$$m_t = (1 - \alpha\eta)m_{t-1} + \alpha K_0 \tag{250}$$

under boundary saturation, and

$$y_t = K_0 + (1 - \eta)m_{t-1}. \tag{251}$$

If $\eta > 0$,

$$m_{T-1} = \frac{K_0}{\eta}[1 - (1 - \alpha\eta)^{T-1}], \tag{252}$$

which gives the second line of (138). If $\eta = 0$, $m_t = m_{t-1} + \alpha K_0$, giving the first line. Strict quietness is obtained as a supremum by replacing K_0 with $K_0 - \varepsilon$ and taking $\varepsilon \downarrow 0$. □

A.17 Proof of Proposition 12

Each $A_C + B_C x$ is continuous, affine, and nondecreasing. The pointwise minimum of finitely many affine functions is continuous, concave, and piecewise affine. Since all slopes are nonnegative, the minimum is nondecreasing. Finitely many affine pairs have finitely many intersections on bounded intervals. □

A.18 Proof of Corollary 1

Because $\mathcal{W}$ is finite,

$$K(x) = \min_{C \in \mathcal{W}} (A_C + B_C x) \tag{253}$$

has an eventual affine branch. Let

$$\bar{A} := \min\{A_C : B_C = \underline{B}\}. \tag{254}$$

There exists $\bar{x} < \infty$ such that

$$K(x) = \bar{A} + \underline{B}x \qquad \text{for every } x \geq \bar{x}. \tag{255}$$

The boundary path has nondecreasing y_t, m_t, x_t.

If $\underline{B} < \eta$ and the boundary path were unbounded, then x_t would eventually exceed $\bar{x}$ and the system would thereafter coincide with the stable affine benchmark-entrenchment model with entrenchment coefficient $\underline{B}$. Theorem 8 gives a finite ceiling, a contradiction. Hence the path is bounded.

Now suppose $\underline{B} \geq \eta$. Since $A_C > 0$ for all C, let $A_{\min} := \min_C A_C > 0$. Then

$$K(x) \geq A_{\min} + \underline{B}x \geq A_{\min} + \eta x. \tag{256}$$

A bounded boundary path would converge along its monotone state trackers to some common level $y_\infty = m_\infty = x_\infty$. Boundary equality at the limit would give

$$y_\infty = K(y_\infty) + (1 - \eta) y_\infty, \tag{257}$$

so $\eta y_\infty = K(y_\infty) \geq A_{\min} + \eta y_\infty$, a contradiction. Thus the path is unbounded and eventually enters the affine branch $K(x) = \bar{A} + \underline{B}x$. Theorem 8 then gives linear growth at $\underline{B} = \eta$ and geometric growth at $\underline{B} > \eta$. □

A.19 Proof of Proposition 13

If h_i enters none of the listed primitives, every coalition cost in (147) is unchanged, so the minimum is unchanged.

If D_i and r_i are nondecreasing in h_i, then τ_i is nondecreasing in h_i. Every coalition containing i therefore has weakly higher cost when h_i rises, while coalitions excluding i are unchanged. The minimum across all coalitions is therefore nondecreasing. □

A.20 Proof of Theorem 9

Because A is Metzler and Hurwitz, $-A$ is a nonsingular M-matrix and $(-A)^{-1} \geq 0$. The saturated matrix $A + BC$ is Metzler. Its Hurwitz stability is equivalent to $-(A + BC) = (-A) - BC$ being a nonsingular M-matrix. The standard nonnegative rank-update criterion gives

$$(-A) - BC \text{ is a nonsingular M-matrix} \iff \rho\Big(C(-A)^{-1}B\Big) < 1. \tag{258}$$

For the equilibrium, $0 = Ax_\infty + Ba_\infty + d$ gives

$$x_\infty = (-A)^{-1}(Ba_\infty + d). \tag{259}$$

Substitution into $a_\infty = k + Cx_\infty$ yields

$$a_\infty = \bar{k}_c + L_c a_\infty, \tag{260}$$

and the stated formula follows from $\rho(L_c) < 1$. □

A.21 Proof of Proposition 5

For any vector z,

$$|Lz| \leq L|z| \leq \|z\|_v Lv = \rho(L)\|z\|_v v,$$

so $\|L\|_v \leq \rho(L)$. Equality follows by evaluating at $z = v$, hence $\|L\|_v = \rho(L)$. Since $\rho(L) < 1$,

$$(I - L)^{-1} = \sum_{r=0}^{\infty} L^r$$

and therefore

$$\|(I - L)^{-1}\|_v \leq \sum_{r=0}^{\infty} \|L\|_v^r = \frac{1}{1 - \rho(L)}.$$

But

$$(I - L)^{-1} v = \sum_{r=0}^{\infty} \rho(L)^r v = \frac{1}{1 - \rho(L)} v,$$

so the upper bound is attained.

A.22 Proof of Proposition 6

Differentiate

$$(I - L)a_\infty = \bar{k}.$$

This gives

$$-(dL)a_\infty + (I - L)da_\infty = d\bar{k}.$$

Since $\rho(L) < 1$, $I - L$ is invertible, and

$$da_\infty = (I - L)^{-1}[d\bar{k} + (dL)a_\infty].$$

For a scalar parameter this is (70). Positivity follows from $(I - L)^{-1} = \sum_{r \geq 0} L^r \geq 0$.

A.23 Proof of Proposition 10

Fix x and write $a^M = a^M(x)$. For any $a < a^M$, uniqueness of the myopic maximizer gives

$$u(x, a) < u(x, a^M).$$

Because $F(x, a)$ is nondecreasing in a and V is nondecreasing,

$$V(F(x, a)) \leq V(F(x, a^M)).$$

Hence $J_x(a) < J_x(a^M)$ for every $a < a^M$, so every dynamic maximizer lies weakly above a^M. For the strict claim, an interior differentiable myopic optimum satisfies $u_a(x, a^M) = 0$. The stated continuation derivative then gives $J_x'(a^M) > 0$, so the dynamic objective increases locally to the right of a^M. Combined with the weak-order result, every dynamic maximizer lies strictly above a^M.

A.24 Proof of Proposition 11

The proof is the same order argument with the signs reversed in the state transition. If $p < p^M$, uniqueness of the myopic optimum gives $g(x,p) < g(x,p^M)$. Since F_G is nonincreasing in p and W is nonincreasing in entrenchment,

$$W(F_G(x,a,p)) \leq W(F_G(x,a,p^M)).$$

Thus every dynamic maximizer lies weakly above p^M. The derivative condition gives strict inequality exactly as in Proposition 10.

A.25 Proof of Theorem 7

For fixed p, the actor's objective is strictly concave in a and has first-order condition

$$r - sa + q(c_0 - p) = 0.$$

Hence

$$a(p) = \frac{r + q(c_0 - p)}{s}. \tag{261}$$

Substituting into the institution's objective yields

$$J_G(p) = \lambda K + \frac{\lambda b}{s}(c_0 - p)[r + q(c_0 - p)] + \frac{\phi}{2}p^2.$$

Therefore

$$J_G'(p) = -\frac{\lambda b}{s}[r + 2q(c_0 - p)] + \phi p$$

and

$$J_G''(p) = \frac{2\lambda bq}{s} + \phi > 0.$$

The unique unconstrained minimizer is

$$p^*(q) = \frac{\lambda b[r + 2qc_0]}{s\phi + 2\lambda bq}.$$

Condition (107) is equivalent to $p^*(q) < c_0$; positivity is immediate. Thus the constrained solution is interior. Moreover,

$$c_0 - p^*(q) = \frac{c_0 s\phi - \lambda br}{s\phi + 2\lambda bq},$$

which gives (109). Differentiation gives

$$\frac{dp^*}{dq} = \frac{2\lambda b(c_0 s\phi - \lambda br)}{(s\phi + 2\lambda bq)^2} > 0$$

and

$$\frac{da^*}{dq} = \frac{s\phi(c_0 s\phi - \lambda br)}{s(s\phi + 2\lambda bq)^2} > 0.$$

A.26 Proof of Proposition 14

Let $R_H = (I - H)^{-1}$. Since

$$dR_H = R_H(dH)R_H,$$

the product rule applied to $L = CR_HB$ gives (173). For a simple eigenvalue, standard eigenvalue perturbation gives

$$d\rho = \frac{u^\top (dL)v}{u^\top v}.$$

The normalization $u^\top v = 1$ gives (174). Primitivity ensures that the Perron root is simple and strictly dominant.

A.27 Proof of Proposition 16

The result follows by induction. At $t = 0$, both sets equal K. Suppose $V_t^{\text{pre}} \subseteq V_t^{\text{post}}$ and let $x \in V_{t+1}^{\text{pre}}$. Then there exists one $u \in U(x)$ such that, for every $a \in A(x)$, $F(x, a, u) \in V_t^{\text{pre}}$. By $U(x) \subseteq U(x, a)$ and the induction hypothesis, for each observed a the same u is feasible and sends the state into V_t^{post}. Hence $x \in V_{t+1}^{\text{post}}$.

A.28 Proof of Proposition 17

Because $0 \leq H_1 \leq H_0$ and both spectral radii are below one,

$$(I - H_1)^{-1} = \sum_{r \geq 0} H_1^r \leq \sum_{r \geq 0} H_0^r = (I - H_0)^{-1} \tag{262}$$

componentwise. Multiplication by nonnegative matrices preserves the componentwise order, so

$$L_1 = C_1(I - H_1)^{-1}B_1 \leq C_0(I - H_0)^{-1}B_0 = L_0. \tag{263}$$

Monotonicity of the Perron root for nonnegative matrices gives $\rho(L_1) \leq \rho(L_0)$. The same Neumann-series argument and the ordering of (d, k) imply

$$\bar{k}_1 = k_1 + C_1(I - H_1)^{-1}d_1 \leq k_0 + C_0(I - H_0)^{-1}d_0 = \bar{k}_0. \tag{264}$$

If $\rho(L_0) < 1$, then $\rho(L_1) < 1$ and

$$(I - L_1)^{-1} = \sum_{r \geq 0} L_1^r \leq \sum_{r \geq 0} L_0^r = (I - L_0)^{-1}. \tag{265}$$

Therefore $a_{\infty,1} = (I - L_1)^{-1}\bar{k}_1 \leq (I - L_0)^{-1}\bar{k}_0 = a_{\infty,0}$.

A.29 Proof of Proposition 15

Boundary saturation and $\varphi \leq \bar{\varphi}$ imply

$$y_t \leq K_0 + (1 - \eta)m_{t-1} + \bar{\varphi}.$$

Therefore

$$\begin{aligned} m_t &= (1-\alpha)m_{t-1} + \alpha y_t \\ &\le (1-\alpha\eta)m_{t-1} + \alpha(K_0 + \bar{\varphi}). \end{aligned}$$

Because $\alpha\eta > 0$, this scalar affine recursion is stable and

$$\limsup_{t\to\infty} m_t \le \frac{K_0 + \bar{\varphi}}{\eta}.$$

The bound on y_t then follows from the quiet-boundary inequality. Finally, $x_t = (1-\delta)x_{t-1} + \delta y_t$ is a stable tracker driven by a bounded input, so x_t is bounded.

B Additional Edge Cases and Diagnostic Quantities

B.1 Critical nonlinear examples

The condition $\rho(D\mathcal{T}(x^*)) = 1$ is genuinely indeterminate. Near zero, the scalar maps

$$\mathcal{T}_-(x) = x - x^3, \qquad \mathcal{T}_0(x) = x, \qquad \mathcal{T}_+(x) = x + x^3 \tag{266}$$

all have derivative one at zero. The first is locally attracting on the positive side, the second is neutral, and the third is repelling. Higher-order terms therefore determine the nonlinear critical boundary.

B.2 Reachability and observability

A matrix can contain an unstable mode that the technological process never excites or that the capacity metric never observes. Theorem 4 handles this exactly by deleting zero coefficients. Empirically, this means a raw spectral radius can be too pessimistic if the corresponding mode is structurally disconnected from the relevant inputs and outputs. Conversely, weak empirical observability can make a dangerous mode difficult to detect even when it is reachable.

B.3 Output-specific transient exposure

Equation (64) can be reported alongside $\rho(M)$ or $\rho(L)$. The transient object measures finite-time forced exposure in a specified output; the spectral objects measure asymptotic feedback.

B.4 Robust viability kernel

For a safe or governable state set $K \subseteq \mathsf{X}$ and transition ambiguity family $\mathfrak{P}$, one robust finite-horizon viability set is

$$\mathrm{Viab}_T^{\mathrm{rob}}(K) := \{x \in K : \exists \pi \ \forall P \in \mathfrak{P}, \mathbb{P}_P^{\pi}(X_t \in K \ \forall t \le T) = 1\}. \tag{267}$$

This is stronger than a chance constraint. It is appropriate when the institutional requirement is to preserve governability under every transition law in the ambiguity family.

C Numerical Verification Audit

Analytic proofs remain primary. The numerical verification suite checks algebra, implementation, comparative-static identities, and institutional extensions across randomized parameter draws. The final suite contains fifteen independent test families.

Table 8: Numerical verification suite for the monograph.

Check	Trials	Maximum error / result
Positive small-gain classification	20,000	zero classification disagreements
State-coordinate invariance	5,000	3.055×10^{-13}
Exact temporal aggregation	5,000	1.776×10^{-14}
Scalar finite-horizon capacity	10,000	2.328×10^{-10}
Benchmark-only closed form	10,000	3.411×10^{-13}
Policy-first Stackelberg benchmark	10,000	2.242×10^{-7}
Loop-gain differential	3,000	max. relative error 2.830×10^{-8}; median 3.619×10^{-11}
Structural-preservation ordering	5,000	zero order violations
Continuous-time positive small gain	10,000	zero classification disagreements
Saturating-entrenchment global bound	5,000	zero bound violations
Heterogeneous-exposure weighted potential	20,000	1.155×10^{-14}
Common-measurement equilibrium conditions	10,000	all equilibrium inequalities verified
Certified-trigger response protocol	10,000	all dominance inequalities verified
Cumulative-hazard additive risk budget	10,000	2.220×10^{-16}
Perron-direction steady amplification	3,000	5.670×10^{-14}

The suite covers the central positive-system equivalence, coordinate and temporal invariance, the scalar benchmark specialization, the policy-first strategic benchmark, sensitivity formulas, continuous-time stability, nonlinear saturation, structural-preservation ordering, cumulative hazard, heterogeneous exposure, common measurement, and certified-trigger response protocols. All fifteen test families passed. The accompanying verification script and audit summary provide the reproducible implementation record.

These checks test implementation and algebra across broad parameter draws. The analytic proofs in Appendix A provide the mathematical basis for the formal claims.

References

Altman, E. (1999). *Constrained Markov Decision Processes*. Chapman & Hall/CRC.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394), 116–131.

Aubin, J.-P. (2009). *Viability Theory*. Birkhäuser.

Aubin, J.-P., Bayen, A. M., and Saint-Pierre, P. (2011). *Viability Theory: New Directions*. Springer.

Bank of England (2026). *Financial Stability Report, July 2026*. Financial Policy Committee.

Berger, M. A., and Wang, Y. (1992). Bounded semigroups of matrices. *Linear Algebra and its Applications*, 166, 21–27.

Bertsekas, D. P., and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3), 387–424.

Collingridge, D. (1980). *The Social Control of Technology*. Frances Pinter.

Costa, O. L. V., Fragoso, M. D., and Marques, R. P. (2005). *Discrete-Time Markov Jump Linear Systems*. Springer.

Crawford, V. P., and Haller, H. (1990). Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica*, 58(3), 571–595.

Dziuda, W., and Loeper, A. (2016). Dynamic collective choice with endogenous status quo. *Journal of Political Economy*, 124(4), 1148–1186.

Eren, E., Krohn, I., and Todorov, K. (2026). Financing the AI infrastructure boom: On- and off-balance sheet borrowing. *BIS Quarterly Review*, March 2026.

Farina, L., and Rinaldi, S. (2000). *Positive Linear Systems: Theory and Applications*. Wiley.

Genus, A., and Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, 47(1), 61–69.

Harsanyi, J. C., and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. MIT Press.

International Energy Agency (2025). *Energy and AI*. IEA, Paris.

Iyengar, G. N. (2005). Robust dynamic programming. *Mathematics of Operations Research*, 30(2), 257–280.

Jungers, R. M. (2009). *The Joint Spectral Radius: Theory and Applications*. Springer.

Kapoor, S., Kolt, N., and Lazar, S. (2025). Position: Build agent advocates, not platform agents. *Proceedings of the 42nd International Conference on Machine Learning*, PMLR 267, 81617–81633.

Kennard, A. (2023). Who controls the past: Far-sighted bargaining in international regimes. *American Journal of Political Science*, 67(3), 553–568.

Kim, J., Lee, S., Lee, J., and Lee, C. (2026). Position: AI lock-in is in progress, and we must be prepared. *arXiv:2608.14565*.

Bavinck, M., Chuenpagdee, R., Jentoft, S., and Kooiman, J. (eds.) (2013). *Governability of Fisheries and Aquaculture: Theory and Applications*. Springer.

MacNeill, T. (2026). Extractive stagnating diffusion: Compute lock-in, the productivity paradox, and the political economy of AI. SSRN Working Paper 7009218.

Marden, J. R., and Shamma, J. S. (2012). Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior*, 75(2), 788–808.

Monderer, D., and Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, 14(1), 124–143.

Naser, M. Z. (2026). Governing AI as infrastructure. *AI and Ethics*, 6, Article 396.

Nilim, A., and El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5), 780–798.

Olugbade, O. (2025). In search of a global governance mechanism for artificial intelligence: A collective action perspective. *Global Public Policy and Governance*, 5, 139–161.

Pauly, D. (1995). Anecdotes and the shifting baseline syndrome of fisheries. *Trends in Ecology & Evolution*, 10(10), 430.

Roberts, H., Hine, E., Taddeo, M., and Floridi, L. (2024). Global AI governance: Barriers and pathways forward. *International Affairs*, 100(3), 1275–1286.

Sastry, G., Heim, L., Belfield, H., Anderljung, M., Brundage, M., Hazell, J., O'Keefe, C., Hadfield, G. K., Ngo, R., Pilz, K., Gor, G., Bluemke, E., Shoker, S., Egan, J., Trager, R. F., Avin, S., Weller, A., Bengio, Y., and Coyle, D. (2024). Computing power and the governance of artificial intelligence. *arXiv:2402.08797*.

Shapley, L. S. (1953). Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10), 1095–1100.

Smith, H. L. (1995). *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. American Mathematical Society.

Trefethen, L. N., and Embree, M. (2005). *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press.

United Nations Independent International Scientific Panel on Artificial Intelligence (2026). *Preliminary Report: Evidence-Based Assessment of Opportunities, Risks and Impacts of AI*. United Nations.

United Nations (2026). *Global Dialogue on Artificial Intelligence Governance: Inaugural Session, Geneva, 6–7 July 2026*. United Nations.

Unruh, G. C. (2000). Understanding carbon lock-in. *Energy Policy*, 28(12), 817–830.

Rajhi, T. (2026). Network design and the threshold for explosive growth. SSRN Working Paper 6559450.

Woo, J. (2026). The loop-gain matrix: Coupled rebalancing feedback and the blind spots of scalar stability monitoring. *arXiv:2608.22768*.

Yamaguchi, A. (2026). Persistence, dominance, and targeted intervention in economic feedback systems. SSRN Working Paper 6584243.

Zhang, Y., and Zhang, T. (2026). The economics of digital intelligence capital: Endogenous depreciation and the structural Jevons paradox. *arXiv:2601.12339*.

Arrow, K. J., and Fisher, A. C. (1974). Environmental preservation, uncertainty, and irreversibility. *The Quarterly Journal of Economics*, 88(2), 312–319.

Besley, T., and Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, 99(4), 1218–1244.

Dixit, A. K., and Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton University Press.

European Commission (2026). Interoperability under the Digital Markets Act and first DMA review materials. Digital Markets Act portal, European Commission. https://digital-markets-act.ec.europa.eu/.

Fink, A. M. (1964). Equilibrium in a stochastic n-person game. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)*, 28(1), 89–93.

Glazer, A., and Janeba, E. (2004). Strategic investment by a regulated firm. *International Tax and Public Finance*, 11(2), 123–132.

Henry, C. (1974). Investment decisions under uncertainty: The "irreversibility effect." *American Economic Review*, 64(6), 1006–1012.

Maxwell, J. W., Lyon, T. P., and Hackett, S. C. (2000). Self-regulation and social welfare: The political economy of corporate environmentalism. *Journal of Law and Economics*, 43(2), 583–618.

Rose, A. K., and Wieladek, T. (2012). Too big to fail: Some empirical evidence on the causes and consequences of public banking interventions in the UK. *Journal of International Money and Finance*, 31(8), 2038–2051.

Seto, K. C., Davis, S. J., Mitchell, R. B., Stokes, E. C., Unruh, G. C., and Ürge-Vorsatz, D. (2016). Carbon lock-in: Types, causes, and policy implications. *Annual Review of Environment and Resources*, 41, 425–452.

Tarui, N., and Polasky, S. (2005). Environmental regulation with technology adoption, learning and strategic behavior. *Journal of Environmental Economics and Management*, 50(3), 447–467.

Tong, D., Zhang, Q., Zheng, Y., Caldeira, K., Shearer, C., Hong, C., Qin, Y., and Davis, S. J. (2019). Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target. *Nature*, 572, 373–377.

United Nations Environment Programme (2023). *Scientific Assessment of Ozone Depletion: 2022*. UNEP, Nairobi.

Colonescu, C. (2026). Fragility and regime stability in a state–society dynamic model. *Journal of Economic Behavior & Organization*, 243, 107439.

Greif, A., and Laitin, D. D. (2004). A theory of endogenous institutional change. *American Political Science Review*, 98(4), 633–652.

Hawkins-Pierot, J. T., and Wagner, K. R. H. (2026). Technology lock-in and costs of delayed climate policy. *Journal of Political Economy Microeconomics*, advance article. https://doi.org/10.1086/740232.

Martimort, D. (1999). The life cycle of regulatory agencies: Dynamic capture and transaction costs. *The Review of Economic Studies*, 66(4), 929–947.

Béland, D., Campbell, A. L., and Weaver, R. K. (2022). *Policy Feedback: How Policies Shape Politics*. Cambridge University Press.

Dehler-Holland, J. (2025). From virtuous to vicious cycles—towards a life cycle model of technology deployment policies. *Research Policy*, 54(7), 105267.

Moynihan, D. P., and Soss, J. (2014). Policy feedback and the politics of administration. *Public Administration Review*, 74(3), 320–332.

Wu, X., Ramesh, M., and Howlett, M. (2015). Policy capacity: A conceptual framework for understanding policy competences and capabilities. *Policy and Society*, 34(3–4), 165–171.

Carlsson, H., and van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, 61(5), 989–1018.

Frankel, D. M., Morris, S., and Pauzner, A. (2003). Equilibrium selection in global games with strategic complementarities. *Journal of Economic Theory*, 108(1), 1–44.

Morris, S., and Shin, H. S. (2003). Global games: Theory and applications. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. 1, 56–114. Cambridge University Press.

Hellwig, C. (2002). Public information, private information, and the multiplicity of equilibria in coordination games. *Journal of Economic Theory*, 107(2), 191–222.

Angeletos, G.-M., Hellwig, C., and Pavan, A. (2007). Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3), 711–756.