When Systems Change Their Minds

Distributed Agency, Institutional Attractors, and the Reconfiguration of Economic Priorities

Flyxion

June 2026

Abstract

Economic history is often narrated through great individuals, political parties, or ideological movements. This essay proposes an alternative interpretation. Rather than viewing economic transformations as the direct product of centralized planners, we examine how large-scale institutions function as distributed attention systems whose priority structures evolve through selection, synchronization, and memory decay. Through the transition from the Bretton Woods era to the contemporary financialized economy, we investigate how changing priority structures produce emergent forms of agency without requiring any single actor to possess complete systemic control.

The essay develops a formal apparatus comprising one central definition and five interconnected theorems describing the dynamics of institutional transformation. The primary contribution is the concept of *Operational Fiber Death*: the condition in which a system retains declarative knowledge of a prior configuration while losing the procedural pathways required to reconstruct it. From this definition, the paper derives theorems on reconstructability loss, anticipatory collapse, approximate repair, opacity selection, and self-concealment, culminating in two synthesis theorems on the dynamical and ontological character of institutional change. The economic history serves as empirical evidence for a broader claim: that large-scale systems change principally by reconfiguring which futures remain reachable, and that the deepest transformations are those which eliminate not only the futures themselves but the capacity to recover them.

Contents

Introduction: The Problem of Large-Scale Change

Historical explanation oscillates between two unsatisfactory poles. The first attributes large-scale social transformations to impersonal structural forces — the logic of capital, the pressures of geopolitics, the imperatives of technology. The second attributes them to the intentions of powerful individuals or coordinated elites: the memo written, the rate set, the election won. Both approaches capture part of the truth while obscuring a mechanism that neither can describe adequately.

Systems do not merely change what they do. Over time, they change what they remember how to do.

The loss of reconstructive capacity is difficult to observe precisely because the processes that generate it simultaneously eliminate the vantage points from which it could be recognized. Reachability closure removes visible alternatives. Memory decay removes experienced practitioners. Opacity selection removes auditability. By the time the process completes, the observers best positioned to identify it have themselves been selected out of the institutional environment.

The purpose of the present work is therefore not primarily historical but diagnostic: to develop a formal language capable of describing how societies lose access to previously reachable futures, and why that loss tends to be recognized only retrospectively, after the procedural pathways have already vanished.

The choice of formalism requires justification. A skeptical reader might fairly ask whether five theorems and an opacity functional add anything that a careful narrative historian could not have said. The answer is yes, but only once we recognize why narrative alone is insufficient: narrative operates inside the same temporal embedding that the mechanism exploits. A historian situated after the transition will naturally describe the outcome as the product of intention, because the alternatives have disappeared and the actors who could have testified otherwise are gone. The formalism attempts — provisionally, fallibly — to construct a vantage point outside that temporal structure. It does not succeed perfectly. No formalism does. But the attempt itself has diagnostic value, because it forces explicit statement of what is being claimed about mechanism, timing, and irreversibility — claims that narrative can make only implicitly, and therefore cannot easily be examined or contested.

The paper proceeds as follows. Section 2 introduces the background concepts of agency, attractors, and distributed optimization. Section 3 describes the Bretton Woods configuration and its internal logic. Sections 4 and 5 trace the crisis, synchronization, and transition dynamics. Section 6 introduces the formal apparatus, opening with the definition of Operational Fiber Death and proceeding through Theorems 1–4 in logical dependency order. Section 7 develops the opacity and financialization analysis. Section 8 states Theorem 5 on self-concealment and the Emergent Agency synthesis. Section 9 addresses repair. Section 10 generalizes the framework beyond the economic case. The conclusion states the two synthesis theorems and returns to the essay's central proposition.

Agency Without an Agent

The human tendency to anthropomorphize complex systems is not merely a cognitive error. It is a reasonable inference under conditions of opacity. When a system produces coherent, persistent, directional behavior, the most parsimonious explanation available to an observer without access to the underlying mechanism is that the system has purposes. The inference is wrong about the mechanism but correct about the behavior.

A thermostat does not want the room to be warm. An ant colony does not decide to build chambers. A market does not choose to clear. A bureaucracy does not intend to expand. Yet all of these systems produce behavior that, observed from the outside, resembles intention. The source of that resemblance is not hidden consciousness but recursive self-correction: feedback loops that cause the system to return toward certain states and away from others, producing apparent goals without requiring any goal-bearing agent.

The concept of an *attractor* formalizes this observation. In dynamical systems theory, an attractor is a region of state space toward which trajectories converge under the system's evolution equations [2, 1, 6]. A system with a stable attractor will return to that region after perturbation. It will appear to *seek* the attractor. The appearance is real in the sense that the behavior is reliable; the seeking is illusory in the sense that no seeker exists.

Large institutional systems — economies, governments, legal orders, scientific communities — possess attractors in this sense [3, 4, 5]. The behavior of such a system over time is not random, and it is not fully controlled by any individual actor. It is shaped by the distribution of incentives, feedback mechanisms, selection pressures, and constraint fields that constitute the institution's effective priority structure [9, 12]. Understanding how that structure changes is the central problem this essay addresses.

The Bretton Woods Configuration

The postwar economic order established at Bretton Woods in 1944 represented a specific institutional configuration with a specific priority vector [22, 34]. Fixed exchange rates constrained competitive devaluation. Capital controls limited cross-border financial flows. Strong labor unions maintained wage bargaining power. Progressive taxation compressed income distributions. Public investment in infrastructure, housing, and education sustained broad-based growth.

The system's dominant objective function — to use the language we will formalize below — weighted employment, social cohesion, reconstruction, and political stability heavily [33]. Financial returns existed within this framework but were subordinate to it. Capital was a means; production and employment were ends.

This configuration was not the product of a unified plan. It emerged from the particular balance of political forces, international pressures, and institutional memories of the 1930s that shaped postwar settlements across the industrial democracies [22, 29]. Its coherence was real but distributed: different institutions in different countries pursued different local objectives, yet the aggregate behavior of the system was recognizably consistent [35].

The Bretton Woods configuration sustained three decades of relatively broad-based growth, falling inequality, and expanding public infrastructure across the industrial world [32]. It also accumulated the contradictions that would eventually destabilize it: dollar outflows from military spending and corporate investment abroad, inflationary pressures from full-employment commitments, rising commodity prices, and the increasing mobility of capital as communications technology reduced transaction costs [27, 31].

Crisis as Attractor Destabilization

The Formal Framework

The following formalism should be understood as a dynamical model of institutional evolution rather than a derivation from economic first principles. It supplies a precise language for mechanisms the historical narrative identifies but cannot itself formalize.

Let $x(t)$ denote the state of the economy at time t, encompassing wages, unemployment, inflation, asset prices, capital mobility, public spending, and labor power. Let $\mathbf{p}_i(t) \in \mathbb{R}^n$ denote the priority vector of institution I_i, where each coordinate represents a weighted institutional objective. The aggregate institutional field is:

$$P(t) = \sum_{i=1}^{N} w_i \mathbf{p}_i(t),$$

where w_i denotes the relative power and amplification capacity of institution I_i.

Define an institutional potential function:

$$V(x,t) = -\langle P(t),\ F(x)\rangle,$$

where $F(x)$ maps the economic state into institutionally visible features. The system evolves approximately by gradient descent:

$$\frac{dx}{dt} = -\nabla_x V(x,t) + \xi(t),$$

where $\xi(t)$ represents exogenous shocks: oil price movements, political conflict, technological change, and geopolitical disruption.

Instability

A crisis occurs when the existing institutional configuration can no longer minimize its own potential function. The old attractor x^* becomes unstable when:

$$\nabla_x V(x^*,t) = 0$$

but the Hessian loses positive definiteness:

$$\nabla_x^2 V(x^*,t) \not\succ 0.$$

Constraint accumulation in the years before a visible transition can be interpreted as the institutional equivalent of entropy-density growth: contradictory objectives, rising cost pressures, and failing legitimacy accumulate before any observable phase change occurs. The Hessian instability criterion supplies the local stability condition that marks when accumulated tension exceeds the restoring capacity of the existing configuration [2, 37].

The Empirical Transition

The destabilizing forces of the early 1970s are well documented [27, 33, 30]: the end of dollar-gold convertibility in 1971, the oil shocks of 1973 and 1979, the resulting stagflation that conventional Keynesian tools could not resolve, and the declining confidence in the postwar economic order that followed. These events do not collectively constitute a plan. They constitute a fitness landscape deformation: the institutional configurations that had reproduced successfully under Bretton Woods began to fail, and the space of viable alternatives temporarily broadened [37].

Synchronization and Transition

The Replicator Dynamics

The standard account of the neoliberal transition emphasizes ideology: the Powell Memo of 1971 [26], the Mont Pelerin Society, the Chicago School, the think tanks that proliferated through the 1970s [28]. This account is not wrong, but it is incomplete. It treats priority vector rotation as the primary mechanism:

$$\mathbf{p}_i(t) \to \mathbf{p}_i(t + \Delta t).$$

But institutions often do not change people. They change who survives.

A more precise formulation replaces individual vector rotation with population dynamics. Let $\rho(\mathbf{p}, t)$ denote the density of institutional actors possessing priority vector $\mathbf{p}$ at time t, and let $F(\mathbf{p}, t)$ denote the institutional fitness of that vector. The evolution of the population follows:

$$\frac{\partial \rho(\mathbf{p}, t)}{\partial t} = \rho(\mathbf{p}, t)\big(F(\mathbf{p}, t) - \bar{F}(t)\big),$$

where:

$$\bar{F}(t) = \int F(\mathbf{p}, t)\, \rho(\mathbf{p}, t)\, d\mathbf{p}.$$

This is a replicator equation [10, 13]. The equation is borrowed from evolutionary biology and applied here in the tradition established by Nelson and Winter [10], who explicitly adopt formal analogy as a methodological tool: the replicator dynamics are not claimed to be literally true of promotion committees but to capture, at the population level, the net effect of differential retention across a large number of local hiring and advancement decisions. Actors whose priority vectors match the emerging institutional fitness landscape are promoted, retained, and hired. Actors whose vectors do not match are marginalized, passed over, or retired. The institution therefore evolves without anyone necessarily changing their beliefs. The population composition changes; the belief distribution of any individual actor may remain constant throughout [11, 14].

The fitness functional itself decomposes into multiple components:

$$F(\mathbf{p}, t) = F_{\text{economic}} + F_{\text{institutional}} + F_{\text{cultural}} + F_{\text{network}}.$$

During the crisis period, F deforms: vectors that were previously adaptive become maladaptive, and vectors that were previously marginal become viable. This deformation is the window during which synchronization mechanisms matter — not because they persuade individuals, but because they arrive at the moment when the selection environment is briefly underdetermined.

Synchronization

Let institutional alignment be measured by:

$$A(t) = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{\langle \mathbf{p}_i(t), \mathbf{p}_j(t) \rangle}{|\mathbf{p}_i(t)||\mathbf{p}_j(t)|}.$$

When $A(t)$ is low, institutions pull in competing directions. When $A(t)$ rises, the system begins to behave as if it possessed a unified will.

The Powell Memo [26], think tanks, industry associations, and elite policy networks function as synchronization mechanisms [28, 27]. They do not create coherence from nothing. They align already-existing institutional vectors into a more coherent field at the moment when fitness landscape deformation has made alignment advantageous. Timing, not content, is what made such interventions effective.

The Volcker Shock as Control Input

Let $\pi(t)$ denote inflation, $u(t)$ unemployment, and $L(t)$ labor bargaining power. A simplified monetary control rule is:

$$r(t) = r_0 + \alpha(\pi(t) - \pi^*),$$

where $r(t)$ is the interest rate and π^* is the target inflation. The labor channel evolves as:

$$\frac{dL}{dt} = -\beta u(t) - \gamma C(t),$$

where $C(t)$ represents deindustrialization, plant closures, and strike fragility. Since high interest rates suppress investment and raise unemployment:

$$r(t) \uparrow \Rightarrow \ u(t) \uparrow \Rightarrow \ L(t) \downarrow .$$

The Federal Reserve's rate increases to near 20% in the early 1980s were therefore not merely an anti-inflation measure. They were a control input that restructured the fitness landscape for labor-aligned priority vectors.

The language of control inputs requires a clarification that applies throughout this paper. Describing Volcker's rate decisions, Thatcher's union legislation, or the Powell Memo as inputs to an attractor system does not deny that these actors had intentions. It distinguishes two levels of agency that the standard historical narrative conflates. *Local intentionality* refers to specific actors pursuing specific goals within a fitness landscape they did not design and largely did not understand as a system. *Global apparent intentionality* refers to the coherent long-run trajectory of the system as a whole — the sense in which the economy appears to have *decided* to financialize, or the state appears to have *chosen* to withdraw from labor protection. The paper's claim is about the second level, not the first. Volcker intended to break inflation; he did not intend, and could not have designed, the full reachability reconfiguration that followed from his decision interacting with the existing fitness landscape, the synchronization already underway, and the global debt mechanics that propagated the shock. Local intentions are real. Global apparent intention is emergent. The error — in conspiracy thinking and in ideological history alike — is to read the second as evidence of a scaled-up version of the first.

Global Propagation

Let a developing country hold dollar-denominated debt D. Its debt service burden is $B(t) = r_{\text{USD}}(t)\, D$, so:

$$\frac{dB}{dr_{\text{USD}}} = D > 0.$$

A rise in U.S. interest rates mechanically increases repayment pressure. When $B(t) > Y(t) - G_{\text{min}}(t)$, where $Y(t)$ is national income and $G_{\text{min}}(t)$ is the minimum politically sustainable public expenditure, the country enters debt crisis [36, 35]. The resulting IMF structural adjustment programs — privatization, austerity, deregulation, labor market liberalization — propagated the priority vector rotation globally through a financial mechanism that required no ideological conversion [27, 31]. The fitness landscape of recipient economies was simply restructured by the terms attached to necessary credit.

The Formal Apparatus

We now state the paper's primary formal contributions in their logical dependency order.

Definition: Operational Fiber Death

Definition 1 (Operational Fiber Death)**.** *Let $\pi : X \to M$ be a projection from institutional state space X onto the publicly observable manifold M. A state $m \in M$ exhibits operational fiber death when:*

$$\pi^{-1}(m) \neq \varnothing$$

but the measure of admissible reconstruction pathways $\mathcal{P}_m$ satisfies:

$$\mu(\mathcal{P}_m) \to 0.$$

That is: the preimage of m remains formally non-empty, but the procedural operators required to construct or reach elements of that preimage have decayed to measure zero.

Operational Fiber Death is strictly stronger than information loss, stronger than forgetting, and stronger than political infeasibility.

The possibility still exists in principle. The route has disappeared in practice.

To make this concrete in the institutional context: a society may retain complete declarative knowledge that capital controls existed, that industrial unions coordinated wage bargaining, that public housing regimes allocated residential space, and that investment banks were legally separated from commercial banks. It may possess historical archives, academic literature, and policy documents describing how each of these operated. Knowledge remains. The procedural capacity — the trained personnel, established protocols, regulatory expertise, and operational infrastructure — has decayed to the point where reconstruction would require building an entirely new institutional apparatus rather than restoring an existing one.

Theorem 1: Reconstructability Loss

Theorem 1 (Reconstructability Loss). *Let $\mathcal{A}(t)$ denote the admissible set of reachable social futures at time t. Reachability closure $\mathrm{Vol}(\mathcal{A}_{labor}) \to 0$ is not a temporary equilibrium displacement. Once the replicator dynamics have operated for a sufficient period, the closure becomes structural: the institutional population required to operate the closed configuration has been selected out of the system. Reachability closure is reversible in principle; reconstructability loss is irreversible without procedural memory. Formally, $\mathcal{A}_{old} \not\subset \mathcal{A}(t)$ establishes that the old trajectory is no longer reachable under current constraints, while $A_p(t) \to 0$ establishes that the system no longer possesses the procedural capacity to reconstruct the operators that once made that trajectory available. These are distinct conditions. The second is strictly stronger.*

The mechanism operates through two channels. First, actors with non-aligned priority vectors are not promoted and retire without successors trained in their orientation [13, 15]. Second, institutional memory decays at different rates for different knowledge types [10, 14]. Let $A(t)$ decompose into declarative memory A_d and procedural memory A_p:

$$\frac{dA_d}{dt} = -\lambda_d D(t), \qquad \frac{dA_p}{dt} = -\lambda_p D(t),$$

where $D(t)$ represents demographic turnover and $\lambda_p \gg \lambda_d$. Procedural memory decays faster because it requires continuous practice to maintain [20, 19]. After approximately one generational cycle, $A_p \to 0$ while A_d remains substantial: people know a configuration existed without retaining the capacity to instantiate it. This is operational fiber death at the institutional scale.

Theorem 2: Anticipatory Collapse

Theorem 2 (Anticipatory Collapse)**.** *The contraction of the admissible set begins before the institutional transition completes. Formally, there exists $t^* < t_{transition}$ such that:*

$$\frac{d}{dt}\operatorname{Vol}(\mathcal{A}_{labor})\Big|_{t=t^*} < 0,$$

that is, the admissible set is already shrinking during the crisis and synchronization phases, before the new attractor has stabilized.

This theorem has an important corollary for the interpretation of political resistance. At any moment during the transition, the actors most aware of what is being lost are also the actors whose procedural knowledge is actively decaying. Their knowledge of the prior configuration is real, but their capacity to articulate what specifically is being foreclosed diminishes precisely as the foreclosure accelerates. This temporal asymmetry explains why resistance to large-scale institutional transitions tends to be described, by historians writing afterward, as reactive and nostalgic rather than prescient — even when the resistors correctly identified the mechanism in real time.

Theorem 3: Approximate Repair

Theorem 3 (Approximate Repair)**.** *Under operational fiber death, exact institutional restoration is unavailable. The maximum achievable repair is functional recovery:*

$$R_\epsilon(S_{broken}) \approx_{\text{function}} S_{old}, \qquad R_\epsilon(S_{broken}) \ncong S_{old}.$$

A repaired institution may reproduce the lost function — broad bargaining power, capital discipline, housing access, public investment — without reproducing the historical machinery that once achieved it.

This theorem is a corollary of Theorem 1 rather than an independent result. It is stated separately because it has direct implications for political discourse. Most reform proposals implicitly assume that repair is restoration:

that the correct objective is to rebuild the institutions that existed prior to the transition [29, 28]. The theorem implies that this framing is not merely politically difficult but formally unavailable once procedural memory has decayed. What remains possible is reconstruction that achieves functional equivalence through new institutional forms [17, 16]. This is not a counsel of despair; it is a correction of the target.

The reachability framework formalizes the distinction between distributional and topological accounts of inequality [32, 23]. Most standard treatments operate on a distributional ontology:

$$\text{Inequality} = \text{Unequal allocation.}$$

The present framework implies a different ontology:

$$\text{Inequality} = \text{Unequal reachability.}$$

Suppose each citizen i possesses a future opportunity set $\mathcal{A}_i$. Define aggregate social freedom as:

$$\mathcal{F} = \sum_i \mathrm{Vol}(\mathcal{A}_i).$$

A society can increase aggregate GDP while reducing $\mathcal{F}$. A housing market can appreciate in value while reducing $\mathrm{Vol}(\mathcal{A}_{\text{young}})$. An educational system can expand expenditure while reducing $\mathrm{Vol}(\mathcal{A}_{\text{student}})$. A labor market can increase productivity while reducing $\mathrm{Vol}(\mathcal{A}_{\text{worker}})$. This is a substantially deeper claim than conventional inequality metrics capture.

Proposition 1 (Reachability Principle). *The effective freedom of an actor is proportional not to the resources possessed but to the volume of admissible futures accessible from their current state [24]:*

$$\mathit{Freedom}_i \;\propto\; \mathrm{Vol}(\mathcal{A}_i).$$

Aggregate social freedom is therefore:

$$\mathcal{F} = \sum_i \mathrm{Vol}(\mathcal{A}_i),$$

and inequality is best measured not by distributional distance from the mean but by the variance of $\mathrm{Vol}(\mathcal{A}_i)$ *across the population.*

Opacity, Financialization, and Self-Concealment

The Opacity Operator

Define a visibility operator:

$$\pi : X \to M,$$

mapping the true institutional state space X onto the publicly observable manifold M. Opacity is information loss under this projection:

$$\Omega = H(X) - H(\pi(X)).$$

When Ω is small, participants can trace causal chains and understand local mechanisms. When Ω becomes large, the system appears to possess unified motives because the underlying interactions are inaccessible.

Opacity decomposes into two qualitatively distinct components:

$$\Omega = \Omega_S + \Omega_M,$$

where Ω_S is *structural opacity* arising from genuine system depth, scale, and recursion, and Ω_M is *manufactured opacity* arising from strategic concealment, regulatory arbitrage, and deliberate complexity. Only Ω_M enters directly into institutional optimization:

$$F(\mathbf{p}, t) = F_0(\mathbf{p}, t) + \kappa\, \Omega_M.$$

This implies that some systems actively select for actors capable of increasing opacity — not through explicit coordination, but because increased opacity raises the fitness of the actors who generate it.

Theorem 4: Opacity Selection

Theorem 4 (Opacity Selection). *Institutional measurement degrades as the mechanism operates. Specifically, the component Ω_M of institutional opacity is itself a fitness-enhancing trait under conditions of reachability closure and procedural memory decay. Systems will therefore evolve toward increasing manufactured opacity through selection, independently of any deliberate strategy to obstruct accountability.*

The empirical prediction is not that opacity will increase without limit.

The fitness functional is concave in Ω_M:

$$F(\Omega_M) = F_0 + \kappa\,\Omega_M - \eta\,\Omega_M^2.$$

The optimum is:

$$\Omega_M^* = \frac{\kappa}{2\eta}.$$

Proposition 2 (Optimal Opacity)**.** *Institutions selected for opacity will converge toward a locally optimal opacity level Ω_M^* rather than maximal opacity.*

Corollary 1 (Systemic Fragility)**.** *When $\Omega_M > \Omega_M^*$, internal coordination costs dominate external accountability benefits. The institution becomes fragile: it is protected from external scrutiny but unable to understand its own risk structure.*

The financial crisis of 2008 illustrates this corollary [33, 8, 30]. The same opacity mechanisms — securitization chains, off-balance-sheet vehicles, derivatives complexity, algorithmic trading — that protected financial actors from regulatory oversight eventually prevented the institutions themselves from accurately modeling their own exposures. Opacity ceased to be protective and became destabilizing. This is exactly what a concave fitness function with interior optimum predicts.

A note on empirical scope is warranted here. The parameters κ, η, and Ω_M are not directly measurable in any existing institutional dataset, and the paper does not claim otherwise. The fitness functional is a modeling schema rather than an empirically calibrated law. What it offers is a qualitative, directional prediction: institutions under opacity selection will exhibit internal coordination failure before reaching maximal opacity, and the signature of that failure — risk blindness coexisting with external impermeability — is observable without measuring the parameters themselves. Calibration is a research program that the present framework makes possible to specify; it is not a precondition for the framework's validity. The emergence threshold $\Omega > \Omega_{\text{threshold}}$ in the Emergent Agency theorem operates differently: it is a conceptual marker identifying the condition under which alignment produces apparent intentionality, not an independently measurable criterion. These two uses of opacity should be kept distinct.

Financialization can now be described formally as a signal reweighting [30, 31]. Let economic output decompose into productive income Y_P and financial income Y_F:

$$Y(t) = Y_P(t) + Y_F(t).$$

Financialization occurs when institutional reward functions increasingly privilege Y_F over Y_P:

$$U_{\text{firm}} = a\,Y_P + b\,Y_F, \quad b \gg a.$$

Under this regime, firms optimize asset appreciation, shareholder returns, buybacks, leverage, and speculative instruments rather than wages, durability, or productive capacity. The economy becomes, in effect, a machine for processing financial information — one in which the signals that drive institutional behavior have become increasingly abstracted from the material processes those institutions nominally govern.

Apparent Agency

Theorem 5: Self-Concealment

Theorem 5 (Self-Concealment)**.** *Let $V(t)$ denote the volume of accessible diagnostic vantage points from which the transition mechanism can be identified. Throughout the process of reachability closure, procedural memory decay, and opacity selection:*

$$\frac{dV}{dt} < 0.$$

The mechanism progressively eliminates the vantage points from which it could be diagnosed. $V(t) \to 0$ is not a contingent outcome but a structural feature of the dependency chain:

$$\begin{aligned} &\textit{Closure} \;\to\; \textit{Memory Decay} \\ &\quad \to\; \textit{Opacity Selection} \;\to\; \textit{Apparent Agency.} \end{aligned}$$

The logic is sequential. Reachability closure removes the visible alternatives that would make the closure recognizable as a loss. Procedural memory decay removes the experienced practitioners who could articulate what was being foreclosed. Opacity selection removes the audit capacity that would make the selection mechanism legible. At each stage, the set of available diagnostic positions shrinks. By the time the process is complete, the observers best positioned to identify it have been removed from the institutional environment.

This theorem explains why institutional transformations are almost always recognized retrospectively. The recognition feels like historical understanding; it is in fact confirmation of operational fiber death. It also explains why the present formalism is necessary: narrative history operates inside the temporal embedding that the mechanism exploits, and formalization — however provisional — creates explicit statements that can be examined independently of the narrative momentum that carries historical description toward intentional attributions.

Emergent Agency (Synthesis)

Theorem 6 (Emergent Agency)**.** *A distributed institutional system exhibits apparent agency when the following conditions hold simultaneously:*

$$\textit{Persistence} + \textit{Feedback} + \textit{Alignment} + \textit{Opacity} \implies \textit{Apparent Agency.}$$

More precisely: if $A(t) > \theta$ for some alignment threshold θ, if the system recursively updates according to:

$$\mathbf{p}_i(t+1) = \mathbf{p}_i(t) + \eta\,\nabla_{\mathbf{p}_i} R_i(x(t)),$$

and if $\Omega > \Omega_{threshold}$, then the aggregate field $P(t)$ generates coherent macroscopic trajectories even when no individual actor controls the whole.

The role of opacity in this theorem distinguishes it from standard accounts of emergent coordination. Without opacity, aligned institutions are merely visible coordination: observers can see the mechanism and attribute behavior to distributed optimization. With opacity, the same alignment appears as unified intention. The phenomenology of conspiracy thinking is precisely the observer's correct detection of alignment through an opacity screen that prevents them from seeing the distributed mechanism beneath. The conspiracy theorist is often right about the behavior and wrong about the architecture.

The dependency chain can now be stated in its complete form:

$$\text{Reachability Closure} \rightarrow \text{Procedural Memory Decay}$$
$$\rightarrow \text{Opacity Selection} \rightarrow \text{Apparent Agency.}$$

Each stage removes explanatory alternatives. First, alternative trajectories become inaccessible. Second, the operational knowledge required to enact them disappears. Third, opacity becomes adaptive because comparison classes have vanished. Fourth, observers attribute unified intention to what remains.

Repair and Reconstruction

The foregoing framework implies a specific account of institutional repair that differs substantially from both conservative restoration and progressive replacement.

Repair, as a formal category, is distinct from replacement [12, 16]. Replacement destroys the existing configuration and substitutes a new one, losing whatever institutional memory the existing configuration embodies. Repair attempts to recover function while preserving structural continuity. The distinction matters because institutional memory — including memory of how to navigate regulatory structures, negotiate with counterparties, and maintain operational coherence under stress — is itself a productive resource [25, 10] that is destroyed by replacement and potentially recoverable through repair.

However, Theorem 1 establishes that exact restoration is unavailable once operational fiber death has occurred: $A_p(t) \rightarrow 0$ means the procedural operators for the old configuration have decayed beyond recovery through ordinary institutional continuity. The repair that remains possible is therefore approximate: functional recovery without formal identity. Institutions designed to recover functions lost through procedural decay cannot be designed by analogy to the historical institutions that originally served those functions, because the trained personnel, regulatory environment, and counterpart institutions those historical forms required no longer exist. A functional equivalent must be designed for the current environment.

This observation has a second practical implication. The policy discourse that frames institutional reform as a choice between nostalgic restoration and radical replacement is falsely dichotomous [17, 7]. Theorem 3 implies a third category: functional reconstruction under memory loss. This is the appropriate target for institutional design once the extent of procedural decay has been honestly assessed.

Generalization Beyond Political Economy

The framework developed in the preceding sections applies to any adaptive system that undergoes reachability closure, exhibits differential memory decay rates, generates opacity through selection, and thereby produces apparent agency. The economic case is illustrative, not definitional.

Scientific paradigms undergo analogous transitions [21]. Experimental techniques fall out of use. Instrument-building skills decay. Theoretical frameworks that were once pursued as live options become effectively unrecoverable because the training pipelines that produced practitioners have been discontinued. A paradigm shift in Kuhn's sense [21] is partly a reachability reconfiguration: after sufficient time, the prior paradigm is not merely unfashionable but operationally inaccessible — a case of operational fiber death at the level of scientific practice rather than economic institution.

Software ecosystems exhibit the same dynamics. Programming languages, architectural patterns, and system designs fall out of production use. Organizations that once possessed the capacity to modify or extend certain systems eventually retain only the declarative knowledge that those systems exist; $A_p \to 0$ while A_d persists in documentation that no one can execute.

Religious traditions, legal systems, and educational institutions all exhibit operational fiber death over sufficiently long time horizons. The capacity to perform certain liturgical functions, argue certain legal forms, or teach certain subjects requires continuous practice maintained across generations. When the chain breaks — through persecution, institutional collapse, or sustained disuse — the knowledge of what was done may survive in texts while the knowledge of how to do it disappears.

The generalization to artificial systems is particularly relevant. A large-scale machine learning system optimizing an implicit objective function through feedback on outcomes is a distributed institutional system in exactly the sense formalized here. It exhibits persistence, feedback, and alignment. If its internal operations are opaque to external observers — and if the selection mechanisms that shaped its optimization are not legible from its outputs — it will exhibit apparent agency. The question this essay has been asking throughout — *what is the system actually optimizing, and who can see inside it?* — is precisely the question the alignment research community is attempting to answer for artificial systems.

Toward a Theory of Institutional Attractors

The economic history examined in this essay demonstrates the mechanism. The formal apparatus developed in response to it claims a much wider jurisdiction. The preceding theorem chain established the dependency relations among its components:

$$\text{OFD} \to \text{Reconstructability Loss} \to \text{Anticipatory Collapse} \\ \to \text{Opacity Selection} \to \text{Self-Concealment.}$$

We are now in a position to state the two synthesis theorems of which this chain is the derivation.

Theorem (Dynamical). *Large systems change principally by reconfiguring which futures remain reachable.*

$$P_{\text{Bretton}} \to P_{\text{Crisis}} \to P_{\text{Monetarist}} \to P_{\text{Financialized}}.$$

Each transition is not merely a redistribution of resources within a fixed possibility space. It is a transformation of the space itself:

$$\mathcal{A}(t+1) = \mathcal{T}_{\text{policy}}(\mathcal{A}(t)).$$

Theorem (Ontological). *System-level intention can emerge from recursive alignment long before anything resembling a mind appears.*

The apparent agency of large institutions — the sense in which markets appear to want things, governments appear to decide things, and corporations appear to believe things — is not an illusion in the sense of being false, nor is it a reality in the sense of implying a subject. It is a structural feature of distributed systems that are sufficiently aligned, sufficiently persistent, sufficiently recursive, and sufficiently opaque.

The first theorem is demonstrated by the historical case. The second is implied by the first: once we see how priority vector alignment produces coherent macroscopic trajectories without centralized control, we see why those trajectories appear intentional to observers who lack access to the distributed mechanism.

The task of political repair is therefore not merely to replace leaders or redistribute resources. It is to alter the priority vectors, fitness landscapes, feedback channels, and admissibility constraints through which institutions determine what futures remain reachable. This is a substantially harder problem than electoral politics addresses, and a substantially more tractable problem than revolutionary replacement proposes. It is the problem of institutional redesign under conditions of partial memory loss — reconstruction that recovers function without recovering form.

Conclusion

The central question is not whether a hidden group controls society. The more important question is: what objective functions are our institutions currently optimizing, and what futures does that optimization make unreachable?

The present essay has argued that this question cannot be answered through historical narrative alone, because narrative operates inside the same temporal embedding that the mechanism exploits. It has proposed a formal apparatus — priority vectors, replicator dynamics, operational fiber death, opacity selection, approximate repair — as an attempt to construct, provisionally, a vantage point from which the mechanism can be described without being subject to its self-concealment.

The remaining gaps in that apparatus identify areas where the framework has not yet been formally specified and therefore indicate directions for future work: empirical calibration of the opacity functional, micro-level derivation of the replicator dynamics from promotion and hiring mechanisms, and extension of the reachability formalism to domains beyond political economy.

The practical implication follows directly from Theorem 3. The appropriate political target is not restoration of prior institutional forms — Theorem 1 establishes that this is formally unavailable once procedural memory has decayed — but functional reconstruction: the design of new institutional forms capable of recovering the functions that were lost. Stable employment, compressed inequality, capital discipline, and broad-based investment are functions, not forms. They were once achieved by one set of institutional arrangements; they could in principle be achieved by others that have never yet existed. The question is what those arrangements would need to look like in an environment where the old procedural knowledge is gone and the old counterpart institutions no longer exist. That question has not been seriously asked, because the dominant reform discourse has been organized around restoration rather than reconstruction. This paper offers a reason to ask it instead.

The behavior of large-scale systems will continue to appear mysterious, intentional, and conspiratorial as long as the underlying mechanism remains invisible. The underlying mechanism remains invisible in part because it progressively eliminates the vantage points from which it could be seen. The appropriate response is not to conclude that no mechanism exists, but to build the formal language required to describe one that systematically resists description.

Institutions do not merely constrain what societies can do. Over time, they alter what societies remember how to do. The deepest political transformations therefore occur not when resources are redistributed, but when reconstructive pathways disappear.

A civilization loses a future twice: first when it ceases to pursue it, and later when it forgets how it was ever reached.

Appendices

A. Reachability Geometry

The central object of the paper is the admissible future set $\mathcal{A}(t)$. We may define a reachability metric on institutional state space by

$$d_R(x, y) = \inf_{\gamma} \int_0^1 c(\gamma(s))\, ds,$$

where γ ranges over admissible trajectories connecting x and y, and c denotes the institutional constraint density.

The reachable future volume from state x is

$$\mathcal{V}(x) = \int_{\mathcal{A}(x)} d\mu.$$

The local contraction rate is therefore

$$\Lambda(x, t) = -\frac{d}{dt} \log \mathcal{V}(x, t).$$

Large positive values of Λ correspond to rapid closure of institutional possibilities.

Proposition 3. *If $\Lambda(x, t) > 0$ over an interval $[t_0, t_1]$, then the admissible future volume satisfies*

$$\mathcal{V}(t_1) = \mathcal{V}(t_0) \exp\left(- \int_{t_0}^{t_1} \Lambda(\tau)\, d\tau \right).$$

Hence persistent closure produces exponential loss of reachable futures.

B. Procedural Memory Dynamics

Let procedural memory be represented by a vector

$$\mathbf{m}(t) = (m_1, \ldots, m_n).$$

Each component corresponds to an institutional capability: industrial policy, union negotiation, capital control management, public housing administration, and so forth.

Assume

$$\frac{dm_i}{dt} = -\lambda_i m_i + u_i(t),$$

where $u_i(t)$ denotes active practice and maintenance.
The solution is

$$m_i(t) = m_i(0)e^{-\lambda_i t} + \int_0^t e^{-\lambda_i (t-\tau)} u_i(\tau)\, d\tau.$$

When maintenance activity vanishes,

$$u_i(t) = 0,$$

capability decays exponentially:

$$m_i(t) = m_i(0)e^{-\lambda_i t}.$$

Operational Fiber Death corresponds to

$$m_i(t) < m_{\text{crit}}$$

for all capabilities required to reconstruct a historical institutional configuration.

C. Attractor Transition Criterion

Consider the institutional potential

$$V(x,t) = -\langle P(t), F(x)\rangle.$$

An attractor state x^* satisfies

$$\nabla V(x^*, t) = 0.$$

Local stability requires

$$H(x^*, t) = \nabla^2 V(x^*, t) \succ 0.$$

Define the instability index

$$I(t) = -\lambda_{\text{min}}\big(H(x^*, t)\big),$$

where λ_{min} is the smallest eigenvalue.

Then

$$I(t) > 0$$

marks loss of local stability.

Institutional crises can therefore be interpreted as bifurcation events in which

$$\lambda_{\text{min}} \to 0.$$

The transition period corresponds to a temporary expansion of accessible trajectories before a new attractor basin forms.

D. Opacity as Information Compression

Let

$$X$$

denote the full institutional state and

$$M = \pi(X)$$

the publicly observable projection.

Opacity was defined as

$$\Omega = H(X) - H(M).$$

We may normalize this quantity by defining

$$\omega = \frac{\Omega}{H(X)} = 1 - \frac{H(M)}{H(X)}.$$

Thus

$$0 \leq \omega \leq 1.$$

A value near zero corresponds to institutional transparency.

A value near one corresponds to near-complete information loss under projection.

The apparent agency threshold may therefore be written

$$A(t)\,\omega(t) > \Theta,$$

where Θ is the minimum alignment-opacity product required for observers to attribute unified intention to the system.

E. Functional Reconstruction Under Memory Loss

Let

$$S_{\text{old}}$$

denote a historical institution and

$$S_{\text{new}}$$

a proposed replacement.
Define a functional map

$$\Phi(S) = (f_1, f_2, \ldots, f_k),$$

where each coordinate measures an institutional function.
Examples include

$$f_1 = \text{employment stability}, \qquad f_2 = \text{housing accessibility}, \qquad f_3 = \text{capital discipline}.$$

Exact restoration requires

$$S_{\text{new}} \cong S_{\text{old}}.$$

Approximate repair requires only

$$\|\Phi(S_{\text{new}}) - \Phi(S_{\text{old}})\| < \epsilon.$$

Theorem 3 follows immediately:

$$S_{\text{new}} \ncong S_{\text{old}} \quad \text{but} \quad \Phi(S_{\text{new}}) \approx \Phi(S_{\text{old}}).$$

The object of institutional repair is therefore preservation of function rather than preservation of form.

References

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

[2] Strogatz, S. H. (2015). *Nonlinear Dynamics and Chaos*. Westview Press.

[3] Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.

[4] Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems*. MIT Press.

[5] Holland, J. H. (1995). *Hidden Order: How Adaptation Builds Complexity*. Addison-Wesley.

[6] Meadows, D. H. (2008). *Thinking in Systems: A Primer*. Chelsea Green.

[7] Page, S. E. (2018). *The Model Thinker*. Basic Books.

[8] Taleb, N. N. (2012). *Antifragile: Things That Gain from Disorder*. Random House.

[9] North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

[10] Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Harvard University Press.

[11] Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Prentice Hall.

[12] March, J. G., & Olsen, J. P. (1989). *Rediscovering Institutions: The Organizational Basis of Politics*. Free Press.

[13] March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1), 71–87.

[14] Weick, K. E. (1995). *Sensemaking in Organizations*. Sage.

[15] Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18(7), 509–533.

[16] Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press.

[17] Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

[18] Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *Economic Journal*, 99(394), 116–131.

[19] Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press.

[20] Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies*, 29(3), 155–173.

[21] Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

[22] Polanyi, K. (1944). *The Great Transformation: The Political and Economic Origins of Our Time*. Beacon Press.

[23] Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91(3), 481–510.

[24] Sen, A. (1999). *Development as Freedom*. Oxford University Press.

[25] Hayek, F. A. (1945). The Use of Knowledge in Society. *American Economic Review*, 35(4), 519–530.

[26] Powell, L. F. (1971). *Attack on the American Free Enterprise System*. Confidential Memorandum to the U.S. Chamber of Commerce.

[27] Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford University Press.

[28] Mirowski, P. (2013). *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown*. Verso.

[29] Streeck, W. (2016). *How Will Capitalism End? Essays on a Failing System*. Verso.

[30] Krippner, G. R. (2011). *Capitalizing on Crisis: The Political Origins of the Rise of Finance*. Harvard University Press.

[31] Duménil, G., & Lévy, D. (2004). *Capital Resurgent: Roots of the Neoliberal Revolution*. Harvard University Press.

[32] Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

[33] Minsky, H. P. (1986). *Stabilizing an Unstable Economy*. Yale University Press.

[34] Galbraith, J. K. (1967). *The New Industrial State*. Houghton Mifflin.

[35] Rodrik, D. (2011). *The Globalization Paradox: Democracy and the Future of the World Economy*. W. W. Norton.

[36] Stiglitz, J. E. (2002). *Globalization and Its Discontents*. W. W. Norton.

[37] Turchin, P. (2016). *Ages of Discord: A Structural-Demographic Analysis of American History*. Beresta Books.