

Reckoning with Systems and Complexity Sciences in the Age of AGI

Emergence, Cross-Scale Feedback, and the Inferential Boundaries of Civilizational Governance

Reckoning with Systems and Complexity Sciences in the Age of AGI

Emergence, Cross-Scale Feedback, and the Inferential Boundaries of Civilizational Governance

Author: Zijun Fu Produced by: Civilization Leap Research Group

Independent Edition V1.3 Revision on dimensional transition and interdisciplinary causal relations

For joint review by researchers in systems science, complexity science, network science, complex adaptive systems, social-ecological systems, system safety, and interdisciplinary governance. This paper independently analyzes boundaries, scales, feedback, emergence, cascades, reflexivity, local optimization, and system retreat, and specifies the handoff boundaries when systems evidence enters engineering, political, legal, ethical, economic, and social judgments. It provides neither a safety guarantee for a particular engineering system, proof of regulatory compliance, nor a manual of governance parameters.

Text revised October 10, 2026 Author: Zijun Fu Produced by: Civilization Leap Research Group

Abstract

As advanced artificial intelligence enters cross-domain prediction, resource allocation, public infrastructure, and multi-agent collaboration, the central question for systems science is no longer merely how to see a larger system more completely. Models, indicators, and interventions themselves begin to change their objects: boundary choices affect who is counted; predictions alter behavior and data generation; local optimization produces cross-scale externalities through shared dependencies; and system capability may be misrepresented as subjecthood, legitimacy, or sovereignty. Systems science must therefore preserve its disciplinary independence in explaining how boundaries, scales, feedback, cascades, and emergence arise, while identifying where its evidence ends and where engineering, political, legal, ethical, and other disciplines must reconsider the judgment. Local judgments may each hold while the overall consequences of aggregation, coupling, and feedback still fail. This paper proposes a mechanism-scale-counterfactual audit requiring any emergence claim carrying policy weight to specify its object, baseline, spatial and temporal scales, generative mechanism, competing explanations, process evidence, and discriminating interventions or counterfactuals. It also establishes a cross-level non-conversion principle: macro-patterns do not automatically generate new entities; functional integration does not automatically prove experience; self-organized order does not automatically acquire authorization; resilience does not automatically equal goodness; critical signals do not automatically trigger political emergency powers; and cross-scale coordination needs do not automatically generate a single central sovereignty. The paper further proposes a fourfold boundary analysis, a double rejection of optimization, reflexive evidence updating, conditional model pluralism, a handoff from distributed causation to responsibility, and system retreat as an auditable institutional proposal for exit, decomposition, and transformation. A minimal shared-dependency example explains how local optimization can generate system damage even when every component is locally reasonable. Drawing on debates over small-world and scale-free networks, complexity-stability research, whole-lake early-warning experiments and their false-positive limits, and the 2003 North American blackout, the paper distinguishes relatively well-supported findings, competing frontier explanations, speculative interdisciplinary connections, and explicit normative commitments. Civilization Leap provides only a minimal philosophical interface: intelligence, consciousness, subjects, responsibility, rights, and bearers of meaning can no longer be inferred from one another through historical co-occurrence; disciplinary facts may in turn revise that interface. The central conclusion is that complexity science can explain how overall patterns form, when they change, and within which boundaries models fail, but cannot alone determine whether a whole experiences, whether an order is legitimate, who should bear risk, or to whom power should belong. The paper further moves disciplinary responsibility upstream to the generation of goals and purposes: shifting from self-interest first to self-interest with others first brings actual stakeholders into judgment before purposes are generated and locked in. Disciplinary facts, direction, methods, and outcomes are stated separately; local success or failure cannot substitute for shared survival and the three long-term goals. The paper accordingly advances research from distinguishing disciplinary objects to investigating interdisciplinary relations, tracing how judgments in this field change reality through other fields and how feedback returns to alter original conditions and the next round of purposes. Stakes across multiple subjects, stages, and settings must be presented separately; correlations, common causes, mediation, and causal feedback must be tested separately; and added relational dimensions must admit specific refutation.

Keywords: advanced artificial intelligence; systems science; complexity science; emergence; multiscale feedback; reflexivity; critical transitions; resilience; system retreat; civilizational governance

Abstract

As advanced artificial intelligence enters cross-domain prediction, resource allocation, public infrastructure, and multi-agent coordination, the central systems question is no longer merely how to observe a larger whole.

Models, metrics, and interventions increasingly alter the systems they describe: boundary choices determine who is counted; predictions reshape behavior and data generation; local optimization propagates externalities through shared dependencies; and system-level capability may be mistaken for subjecthood, legitimacy, or sovereignty. Systems and complexity sciences must therefore remain professionally independent in explaining boundaries, scales, feedback, cascades, and emergence, while also marking where their evidence ends and where engineering, political, legal, ethical, and other judgments must take over. Locally valid conclusions can still combine into system-level failure. This paper advances a mechanism-scale-counterfactual audit for emergence claims that carry policy weight. Such claims must specify the object, baseline, spatial and temporal scales, generative mechanism, competing explanations, process evidence, and discriminating interventions or counterfactuals. It also proposes a cross-level non-conversion principle: macro-patterns do not automatically establish new entities; functional integration does not establish experience; self-organized order does not authorize itself; resilience is not goodness; dynamical warning signals do not create political emergency powers; and cross-scale coordination needs do not generate unitary sovereignty. The paper develops a fourfold boundary analysis, a double rejection of both local and single-global optimization, reflexive evidence updating, conditional model pluralism, a handoff from distributed causation to responsibility analysis, and system retreat as an auditable proposal for exit, decomposition, and transformation. A minimal shared-dependency example shows how locally rational optimization can generate system damage even when no component is obviously wrong. Evidence from network science, complexity-stability research, whole-lake early-warning experiments and their false-positive limits, and the 2003 North American blackout is separated into established support, competing frontier interpretations, speculative interdisciplinary connections, and explicit normative commitments. Civilization Leap supplies only a minimal philosophical interface: intelligence, consciousness, subjecthood, responsibility, rights, and meaning-bearing can no longer be inferred from their historical co-instantiation in human beings. Scientific evidence may in turn revise that interface. Complexity science can explain how system-level patterns form, change, and fail within declared boundaries; it cannot by itself determine whether a whole experiences, whether an order is legitimate, who ought to bear risk, or where public power belongs. The paper further locates disciplinary responsibility before goal and purpose generation: a shift from self-interest first to self-interest with others first brings subjects with genuine stakes into judgment before goals are formed and locked. Facts, direction, methods, and outcomes remain distinct. The analysis is further extended to cross-disciplinary causal relations. It traces how judgments in this field change real-world conditions through other domains, and how feedback modifies the original conditions and subsequent purposes. Effects across actors, stages, and settings must be distinguished from correlations, common causes, and merely asserted connections.

Keywords: advanced AI; systems science; complexity science; emergence; cross-scale feedback; reflexivity; critical transitions; resilience; system retreat; civilizational governance

Research Positioning, Scope of Materials, and Evidentiary Discipline

The starting point of this study is a transition in theory and dimensions. It retains disciplinary precision while deepening causal relations with neighboring fields. Disciplinary boundaries mark burdens of proof; they do not end the tracing of shared consequences. Existing scholarship participates in competing explanations and counterexample testing; the absence of a place in an inherited classification cannot require new questions to retreat into old frameworks.

This paper is a study of conceptual systems analysis, complexity-governance methodology, and interdisciplinary normative interfaces. It is not a history of general systems theory, a systematic review, a specific system-dynamics model, engineering safety certification, or an opinion on current law. Its materials primarily comprise verifiable original papers, publishers' texts, official accident reports, and peer-reviewed research available by August 24, 2026. Classical theories are used to recover concepts' original conditions of application; cases identify mechanisms and limits of extrapolation. Neither any tradition nor this paper itself is exempt from substantiation.

The paper divides claims into four argumentative and evidentiary categories, abbreviated A/B/C/F for verification. Category A covers structures, mechanisms, or qualified facts with relatively stable support. B covers frontier judgments involving model dependence, domain differences, or competing explanations. C covers interdisciplinary connections and AGI scenarios not yet adequately tested against actual objects. F covers openly attributed philosophical/normative commitments, such as preserving plural sources of meaning, denying that system complexity automatically generates rights or sovereignty, and prohibiting aggregate indicators from silently erasing actual subjects and affected parties. A conclusions must still specify objects, scales, time windows, and model assumptions. B must present counterexamples and external validity together. C serves only as stress testing and cannot be presented as scientific facts that have already occurred. F does not masquerade as empirical truth but accepts scrutiny through value premises, logical consistency, counterexamples, and real consequences. Empirical findings, explanations, normative judgments, and institutional proposals require separate substantiation.

An Entry Point for New Readers: Civilization Leap's Minimal Interface

Readers need neither prior knowledge nor endorsement of Civilization Leap to review this paper independently. It adopts only three revisable premises. First, meaning is a plural dimension of value formed by actual subjects through life, relationships, practice, choices, and commitments. It is neither a system state, a persistence indicator, nor a unified objective function; nor is meaning the sole value of public order. Second, finitude includes at least limits to knowledge, time for action, material resources, life courses, and actual capacities for repair. It gives weight to delays, distribution, and irreversible consequences, but complexity science cannot thereby prove the sole source of meaning or any specific institutional obligation. Third, a system is an analytical unit delineated for a problem, scale, and time window, whereas a subject requires independent evidence of experience, continuity, and a relation to its own consequences. The two are not coextensive.

Within this interface, the age of AGI asserts neither a particular capability threshold nor an already-realized machine subject. Rather, advanced artificial systems mean that intelligence, consciousness, subjects, responsibility, rights, political standing, and meaning-bearing can no longer be inferred from one another merely because they historically coincided in humans. Current public evidence remains insufficient to establish subjective experience, non-transferable subject continuity, or awareness of one's own consequences in current advanced AI. In specific deployments with traceable responsibility, explicit authorization, auditability, correctability, and exit, current advanced AI can play functional roles in understanding, protecting, and transmitting meaning and extending capabilities. These describe roles in accountable deployments, not a system's intrinsic nature, and constitute no evidence of subjecthood, standing, or power, while leaving future subject evidence open. Systems science supplies disciplinary facts and may require upstream philosophical premises to be reopened. Philosophy only specifies boundaries against automatic inference; it does not adjudicate whether emergence, feedback, critical transitions, or resilience exist.

The Paper's Minimal Commitments

This paper does not require acceptance of a closed theory of civilization. Provided readers accept that system boundaries require explanation, cross-level inference requires bridging arguments, and disciplinary facts may revise philosophical interfaces, all remaining propositions can be tested independently within systems and complexity sciences.

Contents

Abstract.....	2
Abstract.....	2
Research Positioning, Scope of Materials, and Evidentiary Discipline.....	3
Multiple Boundaries in Systems Research and the Endogenization of Purposes.....	6
From Risk Feedback to Purpose Recursion.....	6
1 Introduction: From Stronger Systems to Governable Complexity.....	7
2 Theoretical Legacies and Unsettled Debts: Which Tools Remain Effective.....	8
3 Objects, Boundaries, Scales, and Observers: System Definitions Must Be Accountable.....	10
4 A Mechanism-Scale-Counterfactual Audit of Emergence.....	12
5 Cross-Scale Feedback, Time Delays, and Reflexive Evidence Updating.....	13
6 A Double Rejection of Optimization: Shared Limits of Local Rationality and a Single Aggregate Objective.....	14
7 Self-Organization, Order, and Legitimacy: Formation Mechanisms Do Not Answer Questions of Justification.....	16
8 Path Dependence, Critical Transitions, Resilience, and System Retreat.....	17
9 Multiple Agents, Systems of Systems, and Distributed Causation.....	18
10 Empirical Materials and Constructed Stress Tests: Case Names Cannot Substitute for Mechanisms.....	19
11 The Strongest Objections, Defeat Conditions, and an Open Research Agenda.....	21
Conclusion: Greater Overall Capability Requires Clearer Inferential Boundaries.....	23
Further Reading and Entry Points for Joint Research.....	24
Appendix A Register of Source-Claim Boundaries.....	27
Appendix B Checklist for Evidence, Extrapolation, and Cross-Level Auditing.....	29
References.....	30

Multiple Boundaries in Systems Research and the Endogenization of Purposes

Understanding the Whole Must Allow Boundaries to Be Reopened

Systems science has extensively investigated wholes, feedback, scales, and emergence. Under AGI conditions, the new task is to incorporate observers, models, institutions, and purpose generation into the same revisable research process. A boundary is not the edge of the world fixed once and for all: when optimization in one field produces costs beyond it, research must reopen that boundary without expanding indefinitely until mechanisms become unidentifiable.

From Outcome Feedback to Purpose Recursion

Suppose a scheduling system improves one network's efficiency but increases other networks' shared loads and dependencies. Pressure changes organizations' safety needs; those needs drive redundancy or concentration; structural changes then alter the next round of scheduling objectives. Engineering, economic, political, and social research jointly identify this chain. Systems research explains delays, thresholds, cascades, and reflexivity rather than using these concepts as proof of normative conclusions. Purposes change within the chain and cannot always be treated as constants external to the model.

Results at Different Scales Cannot Cancel One Another Out

Instances, organizations, infrastructure, regions, and generations may exhibit different outcomes. Rapid concentration during initiation may resolve imminent risks but also leave dependencies difficult to exit during stable operation. An improved average cannot offset irreversible closure at a critical node. Models of multiple subjects should retain conflicts of interest, refusal, and conditions of non-cooperation rather than assuming perfect coordination and then claiming the whole is established.

How to Test the Theoretical Contribution

Test whether adding purpose updating, cross-domain dependencies, and slow variables yields different and more verifiable explanations or risk warnings. Feedback mechanisms require independent process evidence and cannot be proclaimed merely from covarying curves. If a simpler model explains the results equally well, or added relations introduce only untestable parameters, simplify the model. If omitted consequences repeatedly return and defeat predictions, revise the system boundaries and the original local conclusions.

From Risk Feedback to Purpose Recursion

Systems science should also treat goal generation as an internal system change. New capabilities alter relationships; new relationships alter safety needs and interests; and these needs enter the next round of purposes. Fixing an external objective and optimizing action alone omits the most important variable in civilizational recursion. Self-interest first and self-interest with others first produce different default paths here.

The old path can be expressed as strategic interaction, cooperation controlling strategic interaction, new interests, and renewed strategic interaction. The new path instead admits actual stakes when purposes form, allowing cooperation, new needs, and new cooperation to generate one another. Conflict remains, but whether cooperation becomes a generative logic should be judged by actual changes in the next round of purposes rather than a declaration of goodwill in a single action.

Recursion among humans and between humans and AGI must separately identify actors, capability changes, control relations, and irreversible consequences. The first chain does not require machine consciousness; the

second must specify the conditions under which stronger capabilities and autonomous purposes apply. When the chains couple, trace who strengthens their own power through the other relationship rather than substituting an abstract whole for concrete subjects.

Formal models can demonstrate certain mechanisms; parameter settings cannot provide proof for all foundational judgments. Comparisons should disclose boundaries, updating rules, observable variables, and competing models, and examine the absence of external rewards, single-point failures, malicious actors, and rent-seeking over relational weights. If cooperation in a model comes from assumed perfect goodwill, its generative mechanism has not been tested.

Local outcomes and global consequences must appear together. Average benefits cannot conceal safety margins or room for correction. Refutation should identify how a particular recursion is persistently blocked or which stronger explanation implies a different causal trajectory. Winning one simulation, reducing a local risk, or achieving a period of peace cannot alone settle the shared future.

1 Introduction: From Stronger Systems to Governable Complexity

Systems science has long addressed structures linking wholes, relations, hierarchies, control, and environments; complexity science investigates how nonlinear interactions generate macro-patterns that cannot readily be inferred from local intuitions. AGI did not originate these questions but changes their institutional density. A cross-domain model may not only observe a system but classify objects, allocate resources, modify interfaces, invoke other agents, and act. Multiple systems also share cloud, identity, payment, energy, and communications infrastructure. Predictions can therefore change those predicted; local success can transfer losses through dependency networks; and deployment itself can rewrite the next round of evidence.

The question consequently shifts from how to build a more accurate model of the whole to how to keep complexity observable, contestable, stoppable, and revisable when models are incomplete, subjects heterogeneous, feedback reflexive, losses unevenly distributed, and some consequences irreparable. This paper invents no overarching science of the whole above existing disciplines. It examines only the conditions for governing complex systems at a civilizational scale: where technologies, institutions, ecologies, and subjects continually coevolve, observation and intervention enter causal chains, consequences cross organizations, jurisdictions, and generations, and system changes may affect survival, rights, or irreplaceable life histories, engineering performance no longer exhausts the problem.

Civilizational scale here neither anthropomorphizes civilization nor assumes a unified will. It marks only the scale, persistence, and cross-domain coupling of effects. Governability likewise does not mean controllability by a single center; it means boundaries, assumptions, disagreements, power, responsibility, and exit conditions can enter public scrutiny. Civilization may develop collective capability without thereby becoming a super-subject possessing first-person experience, representation, or sovereignty.

1.1 Three Core Contributions

First, the paper restores emergence from generalized rhetoric to an auditable claim, proposing seven audit elements: object, baseline, scale, mechanism, competing explanations, process evidence, and counterfactuals. It neither demands that macro-explanations be fully deduced from micro-equations nor permits apparent novelty, difficulty of prediction, or a simulation image to support conclusions about entities, subjects, or policy directly.

Second, the paper proposes a cross-level non-conversion principle. Systems-science conclusions can provide factual premises for normative and institutional research but cannot bypass independent argument: self-organization is not self-authorization; stability and resilience are not goodness; causal centrality is not legal responsibility; predictive accuracy is not authority to set goals; coordination needs are not sovereignty; and

functional unity is not sufficient evidence of subjecthood. The reverse also holds: plural meaning, finitude, and openness concerning subjects cannot establish that a system exhibits emergence, approaches a critical point, or must adopt redundancy, polycentricity, and retreat.

Third, the paper extends boundaries, scales, and observers beyond mere model choices into objects of accountability, proposing a double rejection of optimization: local actors' rationality can produce aggregate externalities, while a single aggregate optimum can erase distribution, rights, local knowledge, and incommensurable losses. The corresponding institutional contribution is not a total complexity score but reopenable object boundaries, comparable model assumptions, sustained observation of slow variables, responsibility handoffs, and systems capable of exit or transformation.

1.2 Research Questions and Boundaries

Table 1 Five Research Questions and Tasks Not Undertaken

Research question	Undertaken here	Not undertaken here
How can emergence become testable?	Provide mechanism, scale, competing-explanation, and counterfactual audits.	Resolve every strong-emergence or ontological debate.
How does feedback change evidence?	Distinguish pre- and post-deployment objects, data generation, and reflexive effects.	Claim that all social predictions necessarily fulfill themselves.
How can local and overall judgments be reconciled?	Present totals, distributions, externalities, and non-negotiable floors separately.	Provide a unique civilizational objective function.
How does system capability enter governance?	Block automatic inference from capability to subjects, authorization, and sovereignty.	Determine specific sources of authority for political science and law.
How should action proceed under high uncertainty?	Propose risk-proportionate model pluralism, reversible trials, and exit design.	Use uncertainty to prevent all deployment or quantification.

1.3 What Systems Science Can Answer Independently and Where It Must Hand Off

Systems and complexity sciences can independently answer the following: once objects, environments, scales, and time windows are specified, how boundaries change conclusions; how feedback, delays, coupling, network structure, and local interactions generate overall patterns; why system-level failures can occur even when components or subsystems function normally; and which interventions and counterfactuals distinguish competing mechanisms. They can also compare the conditional consequences of different structures for cascades, recovery, lock-in, redundancy, adaptation, and exit.

They must stop at interdisciplinary handoff points. Moving from the existence of an overall system pattern to who experiences, who holds rights, which losses cannot be sacrificed, who may authorize, or who should bear responsibility requires renewed involvement by epistemology, studies of life and mind, ethics, law, political science, economics, and social research. The systems-level whole is not a higher adjudicator. Disciplinary boundaries protect precision in mechanism judgments while marking where evidence stops and who takes up the next judgment. Even individually valid local judgments cannot automatically establish that the overall direction after aggregation, feedback, and conflict is right.

2 Theoretical Legacies and Unsettled Debts: Which Tools Remain Effective

General systems theory introduced open systems, environments, hierarchies, and relations into interdisciplinary analysis; cybernetics established a common language of feedback, regulation, and communication; Simon's near-

decomposability explained why hierarchies and modules facilitate the formation and analysis of complex structures; Anderson's More Is Different emphasized that changes of scale may require new organizational concepts; and the Conant-Ashby theorem connected successful regulation with models of the regulated object under specific assumptions.[1–5,22] These remain effective tools, but they did not answer later questions about who sets goals, how boundaries are selected, or how system outputs enter public power. Having a better system model does not mean possessing complete reality or standing to rule.

Nonequilibrium self-organization, self-organized criticality, complex adaptive systems, and network science revealed mechanisms including ordered structures without stepwise central commands, local rules and macro-patterns, small-world structures, growth, and preferential attachment.[6–10] Cross-domain borrowing, however, often compresses mechanisms into metaphors: an apparently power-law distribution becomes a declaration of self-organized criticality; network hubs become proof that concentration is inevitable; spontaneous order is presumed freer or more legitimate. Broido and Clauset's tests of nearly a thousand networks found that strongly scale-free structures are not ubiquitous. This does not reject power laws; it reminds us that model forms must compete with alternative distributions.[11]

Complexity-stability research likewise resists slogans. May showed in random interaction models that size, connectance, and interaction strength may alter stability conditions; later research showed that interaction signs and network structure also change the conclusions.[12–13] Neither greater complexity means greater stability nor greater complexity means greater fragility is a universal law independent of object structure. Diversity, redundancy, modularity, and concentration can each increase or decrease resilience under different coupling conditions.

Research on ecological resilience, social-ecological systems, critical transitions, and early warnings provides important frameworks for multistability, adaptation, transformation, and cross-scale governance.[14–21,33–34] These literatures already recognize boundaries: resilience asks what maintains which functions against which disturbances; warning signals require mechanisms, controls, and false-positive analysis; and polycentric arrangements hold only under particular rules, capabilities, and externality conditions. This paper does not correct these mature literatures. Its contribution is to prevent their compression, when bridged into AGI institutions, into claims that more resilience is always better, a signal warrants expanded powers, polycentricity is inherently democratic, or a global model is inherently superior.

The system-safety legacy further shows that whole-system failures cannot be reduced to single-component errors. Perrow's normal-accident tradition emphasizes unexpected combinations under complex interactions and tight coupling; Leveson extends safety analysis to control structures, constraints, and organizational processes.[30,49] Relative to these legacies, this paper's contribution is not to prove again that complex systems have accidents but to integrate two AGI-era conditions into one audit. First, models, indicators, and deployments change the studied objects and the next round of evidence. Second, once system capability enters chains of public resources and power, system-assurance evidence must be explicitly handed off at authorization, rights, and legitimacy. The first requires evidence versioning; the second prevents the systems-level whole from subsuming independent political, legal, and ethical judgments.

2.1 Separate Scientific Propositions, Interface Propositions, and Institutional Proposals

Table 2 Evidentiary Ownership of Four Types of Proposition

Proposition level	Typical questions	Usable evidence	Must not masquerade as
Complexity science/method	Boundary sensitivity, emergence mechanisms, feedback, delays, cascades, model failure.	Data, models, process tracing, interventions, counterfactuals, external validation.	Value rankings, legal force, or public authorization.

Proposition level	Typical questions	Usable evidence	Must not masquerade as
Normative and political interfaces	Does formation imply justification? Does coordination require concentration? How do losses enter decisions?	Factual premises plus explicit arguments about rights, non-domination, distribution, and authorization.	Laws of complexity science.
Ontological handoff	Does functional unity establish a subject candidate? Do system boundaries equal subject boundaries?	Independent combined evidence from mind studies, life sciences, epistemology, and institutional procedures.	Direct confirmation or exclusion of subjects from network patterns.
Institutional proposals	How should exit, decomposition, monitoring, model competition, and appeals be designed?	Comparative implementation, failure rates, behavioral consequences, costs, rights, and legal sources.	A unique solution directly deduced from mechanisms.

2.2 Evidentiary and Argumentative Status A/B/C/F

Category A includes the following: system conclusions are sensitive to objects and boundaries; feedback, delays, network dependencies, and local interactions can generate counterintuitive aggregate consequences; candidate warnings can be observed in particular controlled environments; indicator use changes behavior in some social settings; and actual grid accidents can exhibit joint organizational-technological cascades.[9–10,12–25,31–34,37–41] B includes whether macro-scales have stronger causal efficacy, whether real networks are strictly scale-free, whether some social systems can be reliably identified as undergoing critical transitions, the relative advantages of governance structures in cross-border cascades, and new collective behaviors that multi-agent ecologies may produce.[11,26–30,39–46] C includes future machine collective subjecthood, a single civilizational critical point, and AGI automatically generating public will or new sovereignty. F contains the paper's explicit philosophical/normative premises and prohibitions on bridging inferences: complexity does not automatically generate rights, authorization, personhood, or sovereignty, and unified indicators cannot silently settle meaning and affected subjects. The paper presents neither B nor C as established science, nor F as an empirical law of complexity science.

3 Objects, Boundaries, Scales, and Observers: System Definitions Must Be Accountable

A system is not the uniquely given partition of the natural world. Researchers choose objects, environments, variables, scales, and observation windows in relation to questions. These choices may be reasonable but must be visible. Urban transport optimization bounded by an administrative district may miss commuting, energy, and housing consequences; platform recommendations bounded by a click window may miss long-term trust and public discussion; public infrastructure counting redundancy by vendor brands may omit shared cloud, identity, or energy failure domains. Incomplete boundaries do not justify stopping research, but do justify recording omissions, conducting sensitivity analysis, and specifying reopening conditions.

3.1 Fourfold Boundary Analysis

Table 3 Four Types of Boundary Cannot Substitute for One Another

Boundary type	Core question	Typical evidence	Common overreach
Causal boundary	Which variables and relations significantly affect the target outcome?	Domain data, mechanisms, interventions, and temporal order.	Treating correlation or proximity as the entire causal domain.
Model/observation boundary	Which objects, variables, time windows, and scales are represented?	Measurement protocols, sampling, model assumptions, omission analysis.	Treating what lies outside the model as outside reality.
Responsibility boundary	Who designs, authorizes, controls, benefits, foresees, monitors, stops, and repairs?	Organizational roles, permissions, logs, capabilities, and conduct.	Allocating fault directly from causal centrality.
Legal/jurisdictional boundary	Which legal sources, jurisdictions, rights, and remedies apply?	Legal texts, institutional powers, and judicial judgments.	Automatically generating jurisdiction or authorization from a system diagram.

Inclusion in the same modeled system does not establish a single subject, and exclusion from a model boundary does not cancel affected status. Analytical boundaries, moral-community boundaries, legal-standing boundaries, and political jurisdictions must be stated separately. A watershed may be an appropriate ecological unit of analysis without automatically becoming a political subject with unified preferences or representation. A platform ecology may form a functional network without thereby acquiring its users' common will.

3.2 Five Scales and Cross-Scale Non-Equivalence

This paper uses five analytical scales: individuals or nodes; relations or networks; organizations or institutions; systems of systems; and civilization and generations. Macro-patterns may be real and explanatory, but cannot absorb micro-level subjects' experiences, rights, and losses into averages. Nor can micro-causation exhaust constraints imposed by macro-structures. Cross-scale analysis should ask together: under which aggregation rules new variables emerge; how downward or upward interventions change outcomes; whether conclusions are robust across scales and time windows; and whether responsibility can be independently handed off from causal description.[48]

A change of scale may change the question, not merely add detail. Short-term recovery speed and long-term substitutability, node reliability and network cascades, personal-agent fidelity and public-resource congestion, urban efficiency and watershed ecology are not simply the same indicator at different resolutions. A better whole must specify whose whole, how long a time window, which losses are averaged, and which subjects cannot exit.

3.3 When the Observer Enters the System

The observer is part of the system is not a universal law of systems research. More precise conditions are that models become public, classifications change resources, indicators determine rewards or sanctions, observed parties can respond strategically, or predictions trigger interventions. Observation and deployment may then enter the causal chain. Goodhart, Campbell, and research on ranking reflexivity show that once an indicator becomes a high-consequence target, relationships may destabilize and behavior may be restructured. This does not mean all quantification fails.[23–25] Researchers must compare pre- and post-deployment data-generating processes and treat behavioral changes induced by the model as a new object.

4 A Mechanism-Scale-Counterfactual Audit of Emergence

This paper defines emergence as nontrivial macro-patterns, capabilities, or explanatory variables arising from interactions among elements relative to an explicit baseline and scale. Stability is a property to be tested, not a defining prerequisite; transient patterns may also be explanatory. Normative emergence is not used as a scientific category, because rules actually forming and rules acquiring justification are different questions.

4.1 Seven Minimum Audit Elements

Table 4 Seven Questions for Emergence Claims Carrying Policy Weight

Audit element	Required account	Failure signal
Object	What are the elements, relations, environment, and state variables?	Invoking the system as a whole without specifying an object.
Baseline	What is new relative to simple aggregation, common external causes, or random patterns?	Novelty arises only from presentation or threshold choice.
Scale	What are the spatial, organizational, temporal, and observational scales?	Conclusions disappear at another scale without explanation.
Generative mechanism	How do local rules, network structures, constraints, resources, and feedback produce the pattern?	Circular explanation: it emerges because it is complex.
Competing explanations	What common drivers, measurement artifacts, selection biases, and alternative models exist?	Only a simulation image or a single fit.
Process evidence	Are temporal order, propagation, media, thresholds, and intermediate states observable?	Only starting and ending states are observed.
Counterfactual/intervention	Which changed condition would produce distinguishable outcomes?	Every outcome is interpreted as emergence.

This audit does not deny the relative autonomy of macro-explanations. Bedau's weak emergence, research on macro-level causal efficacy, and its critics all show that whether the macro surpasses the micro depends on computation, intervention distributions, and measurement choices.[26–28] This paper requires only that macro-properties used to allocate public resources, restrict rights, or expand authority be connected to generative conditions and distinguishable consequences. A metaphysical claim rejecting all counterfactual distinctions in principle may remain discussable, but cannot directly serve as empirical policy evidence.

4.2 Criticality, Cascades, and Irreversibility Require Separate Definitions

Criticality may refer at least to physical phase transitions, self-organized criticality, critical slowing down, network thresholds, or general policy turning points. Every use should specify the dynamical-system class, state variables, candidate mechanisms, stability or attraction domains, control parameters, and network conditions. Sudden change does not automatically mean a critical transition. A cascade here is a sequence of secondary changes generated by an initial disturbance through dependencies, propagation, or load transfers. It requires topology, propagation rules, temporal order, and counterfactuals. Sequential occurrence is not causal cascading.

Irreversibility is not a single property either. Irreversible state mappings, dynamical hysteresis or non-returning paths, difficulty of recovery within real time and costs, and uncompensable losses in life, relationships, and history involve different evidence and disciplinary ownership. Systems science can test state transitions, hysteresis, and recovery paths. Whether a loss is uncompensable requires separate argument from life sciences, ethics, law, and

affected subjects. Dynamical irreversibility must never directly imply moral absoluteness or political emergency powers.

4.3 Hard Stops on Inference

Hard Stops for Emergence Claims

A claim cannot continue to support high-consequence institutional inferences if it relies solely on one simulation or image, omits scale and baseline, fails to compare alternatives, retains old evidence despite deployment changing the object, or directly infers consciousness, personhood, legitimacy, or sovereignty from functional patterns.

5 Cross-Scale Feedback, Time Delays, and Reflexive Evidence Updating

Feedback is more than the textbook division into positive and negative feedback. Civilizational systems simultaneously contain reinforcing, balancing, cross-domain, reflexive, and coevolutionary feedback: resource allocation changes behavior; behavior changes data; data retrain models; models alter the next allocation. Infrastructure efficiency increases usage, which expands shared dependencies. Risk labels bring support but may also produce stigma and data avoidance. Dynamic decision experiments also show that delays and accumulation can cause actors to misjudge the returning effects of their own actions.[47] Evidence for these objects is not permanently valid merely because it was collected once before deployment.

5.1 Fast Variables, Slow Variables, and Observation Windows

Transaction success, response time, and local failures can be observed by the minute; skill loss, institutional capability, social trust, ecological recovery, and local knowledge may change over years or generations. Evaluating systems only through fast variables misrepresents chronic lock-in as immediate success. Conversely, waiting until every slow variable is fully visible paralyzes action. A reasonable approach is dual-window registration: specify short-term performance, candidate slow variables, observable proxies, review intervals, and stopping conditions. Accuracy in one window must not become assurance for the system's entire lifecycle.

5.2 Delay, Hysteresis, and Repair Are Different

Delay is the interval between a changed input and an observable output. Hysteresis means the return path differs, so stopping the input does not immediately restore the original state. System stoppage, configuration rollback, functional recovery, and actual repair require separate verification. A model version can be rolled back, while altered resource opportunities, relational trust, local capabilities, and life events may not return to their starting points. Technical reversibility therefore cannot proxy actual reparability.

5.3 Evidence Versioning

Table 5 Evidence-Updating Procedures in Reflexive Systems

Stage	Minimum questions	Permitted conclusion	Reopening trigger
Before deployment	What are the object, boundaries, baseline, data generation, and competing models?	Candidate mechanisms or performance in the declared environment.	Changes in object, interface, scale, or use.
During pilot	Does observation or intervention alter behavior, choices, and data?	Bounded causal and side-effect judgments.	Strategic adaptation, distribution shift, permission propagation.

Stage	Minimum questions	Permitted conclusion	Reopening trigger
During diffusion	Do new regions, groups, supply chains, and jurisdictions remain the same object?	New objects require revalidation and do not inherit all pilot qualifications.	Concentrated dependencies, cross-domain externalities, fewer alternatives.
Long-term operation	How do slow variables, delays, lock-in, exit, and post-exit responsibility change?	Continuing but revisable institutional judgments.	Warnings, appeals, shared failures, or changes in subject evidence.

Model pluralism is not a permanent ritual for every problem. For low-consequence, reversible problems with stable boundaries and a principal model proven superior over time across heterogeneous environments, one main model may be used with failure monitoring retained. Requirements for competing models, scenarios, and sensitivity analysis should rise where structural uncertainty is substantial, model choice significantly alters distributions, consequences are irreversible, or cross-scale externalities are large. Model pluralism cannot postpone decisions indefinitely; single-model speed cannot purchase blind spots either.

6 A Double Rejection of Optimization: Shared Limits of Local Rationality and a Single Aggregate Objective

The first kind of complex-system failure arises from aggregated local rationality. Actors individually optimize prices, attention, speed, appointments, returns, or safety within existing rules, while the whole may suffer resource depletion, congestion, polarization, arms races, and cascades. Schelling's segregation model shows that mild local preferences can generate strong macro-separation, without proving a single cause for actual segregation. Hardin demonstrates commons dilemmas under specific open-access conditions, while Ostrom's long-term institutional research rejects the panacea that privatization or central coercion is the only solution.[31–34] Macro-patterns cannot directly reveal every actor's motives, and local rule compliance cannot establish that overall consequences are acceptable.

The second kind of failure comes from a single aggregate optimum. A global model may reduce a cost while deleting distribution, fundamental rights, local knowledge, relational continuity, and uncompensable life histories from its objective function. Improved averages do not show who bears tail risk; aggregate stability does not show that the maintained state deserves preservation. AGI expands search and coordination capabilities but does not remove the politics of goal-setting or convert incommensurable values into one unit.

6.1 Aggregate Outcomes and Distribution Require Parallel Accounts

Any high-consequence optimization must at least display aggregate performance, distributions across groups and regions, those unable to exit, tail risks, intertemporal transfers, key rights-based side constraints, uncompensable losses, and alternatives together. Distributional review does not reject aggregate indicators; it prevents them from acquiring default permission to sacrifice specific subjects. An option may be better in aggregate yet unacceptable because it concentrates serious harm on people unable to appeal. Nor may a locality invoke its knowledge and autonomy to transfer harm to external groups.

6.2 Incommensurability Does Not Make Decision Impossible

Incommensurable values do not require institutional paralysis. Non-negotiable floors can be set before multicriteria comparison; phased trials can address insufficient evidence with reversible consequences; irreversible consequences require higher burdens of proof and authorization; dissent, alternatives, and review can be retained; and authority to decide within specific time windows can be clarified. Rejecting a total civilization score does not reject comparison. It makes the dimensions, tradeoffs, and bearers of loss visible.

6.3 Indicator Reflexivity and Cumulative Disadvantage

When indicators of contribution, risk, reputation, or productivity enter allocation, they change actors' access to resources in the next round. Easily quantified, short-cycle, and public behaviors may be overvalued, while care, long-term foundational work, and invisible contributions are undervalued. Those with sparse records receive fewer resources and struggle to produce visible records; the system then interprets existing distributions as differences in ability. This is a constructed risk mechanism, not an established universal fact. The minimum boundary separates basic guarantees, personhood, rights, and political voice from a single performance score, records attribution uncertainty, and permits objections and corrections.

6.4 A Minimal Mechanism Example: Local Optimization Generates System Harm through Shared Dependencies

To keep the double rejection of optimization from remaining purely conceptual, consider an explicitly labeled constructed example, not a claim about actual frequency or a universal law. Public AI services A and B both call the same identity/cloud service S. Each optimizes response time and its own success rate; its controller automatically increases concurrency and retries during local congestion. While S's aggregate load remains in the safe range, each local strategy may improve its own indicators. When A and B expand simultaneously and retries also consume S, total load approaches shared capacity, queueing and failure probabilities may rise nonlinearly, retries further amplify load, and both services eventually deteriorate together. No local controller need make an obvious mistake: system harm comes from shared dependencies, timing, and externalities omitted from local objectives.

Table 6A A Minimal Constructed Example of Local Optimization under Shared Dependencies

Local action	Locally visible outcome and shared effect	Discriminating test
A increases concurrency	A's waiting time falls and throughput rises; shared capacity, identity, and network resources of S are also consumed.	Hold B fixed and change only A's concurrency; observe S's spare capacity and A/B tail latency.
B expands simultaneously	B's local indicators improve; combined with A, load approaches the shared capacity boundary.	Reverse timing or isolate resources; compare whether joint failure disappears.
Retries after failure	Individual-request success may rise briefly; positive retry feedback continues increasing S's load.	Limit retries or add backoff; observe changes in cascading.
System-level outcome	Both local strategies may be reasonable; shared capacity, delays, and feedback increase overall loss.	Include shared risk in local objectives or separate dependencies; downgrade the mechanism explanation if outcomes do not change.

A minimal formal expression captures this conditional mechanism: local controllers optimize $U_A(x_A)$ and $U_B(x_B)$, respectively, while system loss also includes shared risk $R(x_A+x_B)$. If local objectives omit R , and R rises steeply as aggregate load approaches capacity C , the combination of local optima x_A^* and x_B^* may significantly increase system loss. What matters is not this simple expression itself but its discriminating predictions: separating shared dependencies, limiting retries, incorporating public-capacity costs into local objectives, or changing load timing should observably change cascade risk. If these interventions do not alter outcomes, the mechanism explanation should be downgraded or withdrawn.

7 Self-Organization, Order, and Legitimacy: Formation Mechanisms Do Not Answer Questions of Justification

Self-organization here means organization formed under external constraints, resources, interfaces, and rules without stepwise instructions from a single center. It is not synonymous with no design, no center, no external inputs, or natural freedom. Protocols, property rights, platform interfaces, resource distributions, and legal order may profoundly shape the feasible space for self-organization. An order without a single commander may still produce de facto domination.

Formation, function, and justification must be separated. Generative order concerns how patterns arise; functional order concerns whether they coordinate, persist, or solve problems; normative and political judgment concerns consent, rights, rule changes, minority protection, distribution, remedies, and alternatives. Self-organized order may acquire legitimacy through independent procedures, but cannot authorize itself. Self-organization does not self-authorize is a logical-political interface, not a new empirical law of complexity science.

7.1 Design and Emergence Are Not Mutually Exclusive

Actual institutions often combine designed boundaries with emergent practices. Public design can establish inviolable rights, interoperability, responsibility, appeals, and exit conditions while allowing local rules, professional conventions, and collaborative forms to emerge within them. Complexity evidence sometimes supports decentralization and diversity, and sometimes centralized sensing or unified protocols. Structure should be chosen jointly according to cascade speed, knowledge distribution, externalities, reversibility, and accountability. Even empirically necessary centralized functions require separate justification of the source, scope, duration, and supervision of coercive authority.

7.2 A Firewall for Cross-Level Inference

Table 6 Conclusions That Existing Evidence Cannot Automatically Establish

Existing evidence	Cannot automatically establish
Macro-patterns or new variables at a scale	New entities, ontological independence, subjecthood, or unified will.
Functional integration, memory, self-repair, goal maintenance	Experience, a unified first person, personhood, representation, or public sovereignty.
Self-maintenance, adaptation, reproduction analogies	Life, sentience, moral standing, or rights.
Self-organized order, network consensus, or protocol agreement	Consent, freedom, legitimacy, self-authorization, or common will.
Stability, persistence, high resilience, or recovery speed	Goodness, justice, safety, worthiness of preservation, or success.
Critical signals, cascade risks, or higher-scale coordination needs	Political emergency, unlimited expansion of powers, a single center, or world sovereignty.
Causal centrality, control capability, or predictive accuracy	Legal fault, authority to set goals, standing to rule, or sovereignty.
Model boundaries, aggregate optima, or improved averages	Moral communities, jurisdictional boundaries, or permission to sacrifice specific subjects.

Existing evidence	Cannot automatically establish
Technical rollback, system survival, or complexity	Actual consequences already repaired, civilizational meaning, or ownership of meaning.

The firewall also works in reverse. Philosophical openness concerning subjects, plural meaning, and finitude do not imply emergence, proximity to a critical point, self-organizing capability, or a mandatory architecture. Political commitments to polycentricity and non-domination do not prove decentralized systems more effective in a particular cascade. Systems science's ability to study wholes does not give the whole superior adjudicative authority over law, ethics, politics, life, or subjects. No level can use its strongest language to complete the next level's independent argument.

8 Path Dependence, Critical Transitions, Resilience, and System Retreat

Historical sequences change current options, but path dependence is not absolute irreversibility. Arthur's models of increasing returns and technological competition show that small events may influence technological lock-in under specific feedback conditions. Establishing actual platform lock-in still requires showing that switching costs, network dependencies, interface control, data-migration obstacles, skill loss, or institutional commitments significantly restrict feasible alternatives.[35] Successful AGI pilots should record accumulated dependencies, alternative capabilities, and exit costs alongside performance. Expansion becomes a new object when scale, use, jurisdiction, or supply chains change.

8.1 Four Kinds of Criticality Must Not Be Conflated

Table 7 Four Non-Interchangeable Critical or Self-Organizing Mechanisms

Concept	Original mechanism	Minimum evidence	Prohibited inference
Dissipative structures	Far-from-equilibrium ordered structures maintained by fluxes.	Open driven systems, nonlinearity, and energy/matter exchange.	Social norms, SOC, or a free order.
Self-organized criticality	Specific slow-driving/dissipation/threshold models spontaneously approach critical states.	Model rules, scale distributions, and competing mechanisms.	Diagnosing SOC from power laws or 1/f noise alone.
Edge of chaos	The ordered-chaotic boundary in constructed models such as random Boolean networks.	Explicit networks, parameters, and performance definitions.	A universal optimum for life, organizations, or governance.
Critical transitions	Critical slowing down and changes in state domains near certain local bifurcations.	State variables, candidate bifurcations, controls, false positives and negatives.	All sudden changes share this mechanism or have warnings.

Scheffer and colleagues reviewed early-warning mechanisms such as critical slowing down; Carpenter and colleagues observed some indicators preceding a food-web transition in a controlled whole-lake experiment; Boettiger and Hastings showed that detection depends on data, mechanisms, and selection processes, and that common indicators can produce false positives and false negatives.[18–21] Warnings can therefore trigger investigation, independent review, buffering, and reversible preparation, but cannot alone declare a political emergency or automatically expand authority. Once sensitivity, specificity, and decision utility receive external validation in a domain, warnings may become bounded operational triggers, while the source of authority still requires separate justification.

8.2 Separate Resilience, Adaptability, and Transformability

This paper distinguishes resilience, adaptability, and transformability in responses to disturbance. Resilience is a system's capacity to retain basic functions, structure, identity, and feedback amid change; adaptability is actors' capacity to influence that resilience; transformability is the capacity to create a fundamentally different system when the existing one is unacceptable. The three may interact but must not be compressed into a total score.[14–17] Every use of resilience must specify its object, functions, disturbance set, time window, and performance boundaries.

Fast recovery does not mean greater long-term safety, nor does long persistence mean a state deserves preservation. A highly resilient oppressive institution, a fragile single-vendor entry point, or a rapidly recovering platform that continually erodes local capabilities may receive opposing assessments along different dimensions. Complexity science can explain which structures maintain what, but cannot alone determine whether the maintained state is good, who pays for recovery, or who may demand transformation.

8.3 System Retreat: An Institutional Proposal Rather Than a Complexity Law

The paper proposes system retreat: when a mission ends, authority expires, dependency risks rise, or alternatives mature, a system can leave its central position, decompose, migrate, or transform without turning existing dependencies into automatic grounds for renewal. This is not an established law of complexity science, but an institutional proposal subject to comparative implementation and failure tests. Minimum dimensions include portable data, replaceable interfaces, disableable permissions, recoverable human and public capabilities, preservable local knowledge, repair of actual harm, and continuing responsibility after exit.

Retreat is not simple shutdown. For life support and public infrastructure, abrupt exit may cause more harm than continued operation. Emergency stoppage, degraded operation, replacement takeover, phased migration, and long-term transformation must therefore be distinguished. If retreat design persistently causes greater irreversible harm without reducing lock-in or arbitrary domination, the proposal must be reconstructed.

9 Multiple Agents, Systems of Systems, and Distributed Causation

Multiple AI agents faithfully executing their users' objectives do not guarantee acceptable collective outcomes. Public appointments, transactions, advertising, computing, and attention may experience high-speed preemption and congestion; shared signals may produce exclusion or collusion; permissions may spread through delegation and tool chains; and where each designer is responsible only for a local module, aggregate consequences may be attributed to ecological behavior. These are category C prospective stress tests. Actual frequency and mechanisms require joint examination through machine-behavior research, network experiments, organizational studies, and field data.[29–30]

9.1 Systems of Systems Do Not Automatically Possess a Common Will

Subsystems in systems of systems have different missions, owners, responsibilities, and jurisdictions. Protocols coordinate states but do not generate public purposes; data fusion expands situational awareness but does not dissolve subject or jurisdictional boundaries; aggregate optimization reduces some costs but does not establish legitimate representation of collective interests. Network consensus, collective memory, goal maintenance, and self-repair establish only the corresponding functional unity. They neither prove nor rule out future subjecthood. If independent combined evidence of experience, continuity, and awareness of one's own consequences appears, it should be referred to studies of mind and life, epistemology, and institutional procedures, rather than affirmed or denied a priori by systems science.

9.2 Distributed Causation Does Not Dissolve Responsibility

The absence of a single sufficient cause does not imply the absence of identifiable responsibility relations. Complexity analysis identifies causal contributions, control positions, available information, shared dependencies, and stopping capabilities. Ethics, politics, and law then allocate responsibility according to roles, authorization, foreseeability, benefits, professional duties, fault, and remedy rules. Causal position does not automatically generate, cancel, or distribute legal and moral obligations.

Responsibility handoffs must at least ask who designed goals and interfaces; approved deployment and expansion; controls key data and credentials; can see warnings; can stop or isolate; benefits from scale; bears externalities; and must preserve evidence and complete repairs. Complex systems may lack an omnipotent controller while still having actors with asymmetric shaping power. No one fully controls must not mean no one is responsible.

9.3 Coordination Needs Do Not Generate Complexity-Based Sovereignty

Cross-border cascades, climate risks, and shared infrastructure may indeed require higher-scale coordination, and even centralized sensing and action within specific time windows. What requires proof is the particular function, necessity, inadequacy of alternatives, and exit conditions, not an unquestionable unified center inferred directly from the scale of risk. Nested arrangements may secure minimum irreversible boundaries, joint verification, and crisis communication at higher scales; retain legal responsibility and public capabilities nationally and regionally; preserve situated knowledge and appeals locally; and protect information, refusal, participation, and fundamental rights individually. Their empirical merits require comparison rather than prior determination by either polycentricity or centralization.

10 Empirical Materials and Constructed Stress Tests: Case Names Cannot Substitute for Mechanisms

This section separates empirical materials from constructed scenarios. Actual cases support only limited conclusions within their objects and evidentiary windows. Constructed scenarios expose questions the framework must ask; they do not prove mechanisms have the same frequency, strength, or causal structure in reality. Multiple causes do not make a case unanalyzable, and reproducing an experimental pattern does not mean a model exhausts an actual accident.

10.1 The 2003 US-Canada Blackout: An Actual Cascade Is Not a Single Network Story

The North American blackout of August 14, 2003 affected about 50 million people and 61,800 megawatts of load, providing empirical material on cross-regional power cascades. The joint US-Canada investigation did not attribute it to one topological indicator. It identified interactions among vegetation contact, line trips, software and alarm failures, inadequate situational awareness, operational coordination, and protective actions.[37] The case reliably supports escalation of local failures through combined dependencies, monitoring failures, and organizational failures. It does not establish that Buldyrev's interdependent random-network model explains the entire accident, or that centralization is necessarily the only remedy.

The case also illustrates the handoff between responsibility and causation. A cascade cannot be reduced to a single root cause, yet vegetation management, software assurance, operational monitoring, coordination protocols, and institutional responsibilities can be identified. Complexity does not dissolve responsibility; it requires more precise mapping of control positions, information availability, and stopping capacity. By contrast, the 2021 cold wave in Texas and the south-central United States primarily involved supply-demand crises and controlled load shedding from extreme cold affecting generating equipment, fuel supplies, and gas-electric dependencies. It cannot be classified as the same accident type as the 2003 transmission-grid cascade.[38]

10.2 Complexity-Stability and Financial Networks: Mechanisms Can Reverse Conditionally

May's random-matrix result suggests that complexity, connectivity, and interaction strength may constrain stability; Allesina and Tang further show that interaction signs and structure alter stability conditions.[12–13] In financial-network models, Acemoglu and colleagues found that denser connections can distribute risk under small shocks yet amplify propagation under sufficiently large ones, producing conditional robustness-fragility. Iyer and Peydró's natural experiment in Indian banking provides narrow-domain evidence of propagation through interbank exposures.[39–40] These materials support neither more connections means more danger nor more decentralization means more safety. They show that the same structure can reverse its effects when shock magnitude, load transfer, and recovery mechanisms change.

10.3 Early Warnings: Controlled Success and Detection Limits Both Hold

In a controlled whole-lake experiment, Carpenter and colleagues gradually increased apex predators to destabilize a food web, using a neighboring lake as a reference. Several statistical indicators changed before the state transition.[19] This is ecological evidence with a mechanism and a control, not a civilization-warning device. Boettiger and Hastings show that even favorable conditions leave early-warning detection constrained by data length, noise, model selection, and conditional sampling; stochastic transitions can produce false positives resembling the prosecutor's fallacy.[20–21] The appropriate institutional implication is therefore greater investigation and preparedness, not converting statistical signals into a definite countdown.

10.4 Network Models: Separate Constructed Results from Empirical Identification

The Watts-Strogatz model shows that high clustering and short average paths can coexist; the Barabási-Albert model shows that growth and linear preferential attachment can generate heavy-tailed degree distributions.[9–10] Observing short paths or heavy tails does not alone identify the generative mechanism. Broido and Clauset find strongly scale-free networks rare under their strict definitions and testing thresholds, while Voitalov and colleagues propose different criteria using regular variation.[11] The durable lesson is not to pick a camp but first to define scale-free, disclose tail estimation and sample selection, and compare alternatives such as lognormal distributions.

10.5 Eight Constructed Stress Tests

Table 8 Mechanism Questions and Inferential Boundaries in Constructed Scenarios

Scenario	Mechanism requiring testing	Must not be presented as
Watershed-city-energy-food coupling	Do administrative boundaries transfer ecological and downstream risks? Are seasonal and intergenerational slow variables omitted?	Urban efficiency establishes acceptability for the whole watershed.
Public health-prediction-behavior-trust	Does classification change care-seeking, migration, disclosure, and data quality?	Predictive accuracy automatically authorizes coercion or permanent classification.
Large numbers of personal AI agents	Does local fidelity produce congestion, preemption, collusion, and exclusion of those without agents?	Individual authorizations aggregate into public consent.
Recommendation-discussion-polarization	How do homophilous clustering, reward structures, events, and political mobilization compete as explanations?	One algorithm causes all polarization.
Cloud-payment-energy-communications	Does nominal redundancy share identity, network, energy, and supply-chain failure domains?	The number of brands equals the number of independent paths.

Scenario	Mechanism requiring testing	Must not be presented as
Diffusion of high-capability system pilots	Have objects, regions, interfaces, procurement, and talent dependencies changed?	Pilot success automatically generates universal qualification.
Recovery through a single platform after disaster	Does short-term recovery erode long-term public, local, and alternative capabilities?	Recovery speed equals comprehensive resilience.
Allocation through contribution or risk indicators	How do indicators alter resources, behavior, and the next round of visible records?	Records equal personhood, meaning, rights, or political standing.

The eight constructed scenarios jointly expose omitted boundaries, rewritten feedback, shared dependencies, path lock-in, and cross-level responsibility questions that the framework must address. They neither prove mechanisms ubiquitous in reality nor substitute for data. Their value is in generating falsifiable research designs: specify objects, compare mechanisms, seek process evidence, observe conditional reversals, and record which results would downgrade or withdraw the original hypothesis.

11 The Strongest Objections, Defeat Conditions, and an Open Research Agenda

An interdisciplinary complexity framework can readily immunize itself against criticism by endlessly adding qualifications. This paper therefore distinguishes logical non-entailment, empirical failure, revision of scope, and reopening of normative interfaces. Logical non-entailment is tested through counterexamples and bridging structures; empirical mechanisms through data, interventions, and external validity; institutional proposals through comparative implementation, costs, rights, and behavioral consequences.

11.1 Eight Strongest Objections

Table 9 Objections, Responses, and Genuine Defeat Conditions

Strongest objection	Response	Conditions requiring narrowing or abandonment
This packages political philosophy as complexity science.	The scientific level reports only mechanisms and failure; the normative level makes conditional inferences from explicit premises.	If the interface creates no new observational questions or mechanism distinctions, remove it from the disciplinary core.
Emergence is irreducible; counterfactual auditing dissolves it.	Complete micro-level deduction is not required; macro-claims need only connect to generative conditions and distinguishable consequences.	The audit persistently excludes explanatory macro-variables in the target domain while adding nothing.
Macro-subjecthood may itself emerge.	System functions neither prove nor exclude subjects; future combined evidence triggers interdisciplinary review.	If the framework permanently excludes systems as subject candidates, its openness claim fails.
Complex risks sometimes require a single center.	Conditional centralized functions are recognized; authority, duration, supervision, and exit require separate justification.	If the framework favors dispersion regardless of evidence, abandon that preference.
Model pluralism and reopening boundaries cause paralysis.	Requirements are graded by consequences, reversibility, and structural uncertainty; low-risk cases may use a principal model.	Downgrade requirements that persistently cause major delays without improving judgment.

Strongest objection	Response	Conditions requiring narrowing or abandonment
Resilience is not goodness is disciplinary common knowledge.	Agreed: it is not a new law; the contribution is an auditable interface for AGI infrastructure.	If the interface changes no case diagnosis or design question, merge it back into existing frameworks.
Distributed causation makes accountability impossible.	Causal analysis hands off control, information, and contributions; independent normative and legal rules allocate responsibility.	Withdraw responsibility maps if they add no auditable relations beyond ordinary organization charts.
Open-system boundaries are always incomplete.	Only purpose-relevant accounts, omission registers, sensitivity tests, and reopening conditions are required, not omniscience.	Stop expanding review if added boundaries do not change explanation, prediction, or distribution over time.

11.2 Empirical Failure, Scope Revision, and Interface Reopening

- If a claimed emergence mechanism cannot distinguish common external causes, simple aggregation, and measurement artifacts through interventions or counterfactuals, downgrade it to a descriptive pattern.
- If changes in boundaries or scales do not alter explanations, predictions, or risk distributions over time in the target domain, reduce the corresponding review requirements.
- If publishing models, classification, and deployment do not alter behavior or data generation, reflexivity requirements should not be applied mechanically.
- Early warnings with externally validated sensitivity, specificity, and decision utility in a particular domain may become bounded operational triggers, but still do not generate authority.
- If one model consistently outperforms competing models across heterogeneous environments, drift, and adversarial conditions, model pluralism may be downgraded to a failure contingency.
- If migration experiments and longitudinal tracking show negligible path effects, lock-in should no longer be claimed.
- If system retreat persistently causes greater irreversible harm than retaining the system without reducing lock-in or domination, the institutional proposal must be reconstructed.
- If shared vocabulary across physical, ecological, network, and social domains produces only metaphors and cannot specify object isomorphism and conditions for mechanism transfer, delete the cross-domain claims.
- If combined subject evidence reproducible across theories and systems emerges and cannot be explained by scripts, controllers, or observer projection, reopen the system-subject interface.

11.3 Reverse Calibration of Civilization Leap

Systems and complexity sciences require upstream philosophy to replace systems are not subjects with systemness and overall function are insufficient to establish subjects. They prohibit using entropy, finite resources, or dynamical irreversibility to prove that meaning can arise only from finitude. Research on mind and life must not use self-organization, self-maintenance, memory, adaptation, information integration, or resilience alone as tests of life, experience, or subject continuity. Epistemology must distinguish model completeness, mechanism maturity, and evidentiary strength; evidence grades do not mechanically trigger institutional powers.

Political governance must separate dynamical criticality from political emergency, higher-scale coordination needs from central sovereignty, and causal position from responsibility attribution. Interdisciplinary scenario analysis can construct competing causal hypotheses, but cannot treat feedback diagrams as mind-reading. First-person statements, observable behavior, interest positions, and possible motives must be presented separately, with refusal to answer permitted for unobserved states. Advanced AI system assurance must specify objects, environments, and scale; boundary changes, collective behavior, shared failure domains, and permission

propagation trigger revalidation. Indicator-based allocation must never become a total account of personhood, meaning, fundamental rights, or political standing.

11.4 An Open Research Agenda

- Develop a common representation of cross-scale causal maps and responsibility-handoff maps, comparing their added value against ordinary fault trees and organization charts.
- Test congestion, collusion, permission propagation, treatment of exiters, and exclusion of people without agents in controlled multi-agent environments.
- Study self-fulfillment, self-negation, data contamination, and dissent after predictions enter institutions, and identify when reflexivity is negligible.
- Construct heterogeneous backup and migration stress tests for shared cloud, identity, payment, energy, and communications dependencies.
- Develop long-term observation designs for slow variables such as trust, professional capability, ecological recovery, and local knowledge.
- Compare the conditional performance of centralized, polycentric, and nested governance across different cascade speeds, knowledge distributions, and rights structures.
- Measure recovery, maintenance of state domains, adaptation, and transformation separately; identify conditions under which one structure increases one kind of resilience while reducing another.
- Study how pilot diffusion, procurement standards, data-network effects, and talent migration create or release institutional lock-in.
- Establish review procedures for cross-domain emergence claims, clarifying model isomorphism, mechanism transfer, and failed extrapolation.
- Compare actual costs and irreversible residual consequences of system exit, decomposition, migration, degradation, and long-term transformation.
- Develop combined-evidence procedures for future candidates for machine collective subjecthood that neither substitute functional performance for evidence of experience nor permanently exclude candidates because of artificial origins.

Conclusion: Greater Overall Capability Requires Clearer Inferential Boundaries

Disciplinary reckoning must both clarify objects and deepen relations. Disciplines may retain boundaries, but cannot stop real causal chains at those boundaries. Research on multidimensional relations, feedback, and the next round of purposes determines whether specialist knowledge can jointly inform civilization's overall judgment.

What the AGI age requires systems science to reconsider is not whether holistic thinking still matters, but who defines systems and who is excluded once collective capabilities enter reality; how models change objects; how local optimization returns across scales; and why functional unity cannot silently inherit subjecthood and public power. Complexity licenses neither mystification nor abandonment of judgment. It requires conclusions tied to objects, boundaries, scales, mechanisms, evidence, and failure conditions, while acknowledging a more difficult fact: local judgments may each be correct, yet the overall consequences of aggregation, coupling, and feedback may still fail.

The paper accordingly closes with seven points: emergence requires mechanism-scale-counterfactual auditing; causal, model, responsibility, and legal boundaries require separate justification; both local rationality and a single aggregate optimum are constrained; self-organization does not self-authorize, resilience does not establish goodness, and critical signals do not generate emergency powers; reflexive systems require versioned evidence;

distributed causation hands off to responsibility research rather than deciding for it; and systems should possess real paths for exit, decomposition, and transformation.

Complexity science can explain how overall patterns form, when they change, and within which boundaries models fail. It cannot alone determine whether a whole experiences, whether an order is legitimate, who should bear risk, or to whom power belongs. Systems science's boundary is therefore neither a wall nor the endpoint of practical problems. It protects independent mechanism judgments while marking where evidence stops and responsibility and authorization require handoffs. Philosophical premises disclose value concerns and non-entailment boundaries; disciplinary facts constrain and may revise those premises; politics and law then address authorization, rights, responsibility, and institutional choices. No level may substitute complexity, meaning, or systemic consistency for the next level's independent argument.

Settled Conclusion

Systems may develop collective capability without thereby acquiring collective subject status or public sovereignty. Coordination may cross scales without silently inheriting authorization. Greater complexity requires more visible boundaries, revisable evidence, contestable power, reparable consequences, and systems capable of retreat.

Further Reading and Entry Points for Joint Research

Within systems and complexity sciences, this paper answers only how objects and boundaries are delineated; how feedback, delays, emergence, cascades, reflexivity, and shared dependencies alter overall consequences; and why individually reasonable components can nevertheless fail together. Further reading follows questions rather than document numbers. When questions enter technical assurance, rights and authorization, resources and allocation, social relations, evidence of subjects, and practical experiments, other disciplines, institutions, and practices must re-enter.

Upstream: Why Do Systems Questions Ultimately Become Questions of Civilization

Continue with [Reckoning with the Disciplines in the Age of AGI: Disciplinary Boundaries, Interactions, and a Shared Research Agenda](#), V2.4. It explains why specialization remains essential to precise knowledge but cannot end actual causal chains: outcomes arising from the joint operation of technical, institutional, economic, social, power, and meaning conditions determine whether a direction or proposal holds overall. To ask why civilization needs a leap, what must ultimately be protected, and where it should arrive, turn to *Civilization Leap: A Path for Civilization in the Age of AGI* (All Five Volumes) and the main entry point to the theoretical framework.

Civilization Leap, Five Volumes: <https://www.civitas.top/en/documents/p02-five-001/>

Civilization Leap Public Reading and Communication Hub: <https://www.civitas.top/en/documents/p01-entry-006/>

Complete Civilization Leap Library: <https://www.civitas.top/en/library/>

Downstream: How Do System Mechanisms Become Verifiable Engineering, Scenario Analysis, and Experiments

Continue with the P11 specialist paper on artificial intelligence and computer science, and current texts on Human-COS, AGI-COS, risk scenario analysis, the Public Window, and national-level comprehensive experiments. Examine how joint configurations, shared dependencies, fault injection, circuit breakers, retreat, responsibility handoffs, and slow-variable monitoring are implemented in real systems. Downstream does not mean translating complexity theory directly into governance parameters. Engineering tests, institutional prototypes, and real outcomes require independent evidence, authorization, and failure conditions, and may revise this paper in return. The Civilization Leap Library provides the current versions.

Across Disciplines: Who Must Re-Enter Where Systems Evidence Ends

- **Artificial intelligence and computer science:** investigate how models, agents, deployed services, and infrastructure are constructed, and whether system assurance, permissions, stopping, rollback, and retirement are verifiable.
- **Political science and ethics/law:** address de facto power, emergency powers, legitimacy, rights, responsibility, remedies, and sovereignty, preventing higher-scale coordination needs from silently becoming sources of authority.
- **Economics, political economy, and value theory:** investigate local optimization, shared resources, computing and platform concentration, externalities, distribution, and exit costs, testing how aggregate optimization changes real incentives.
- **Sociology, anthropology, and civilizational studies:** observe how indicators, classifications, platforms, and AI agents change behavior, relationships, local knowledge, trust, and institutional practices, preventing the modeled system from subsuming actual lifeworlds.
- **Epistemology, life sciences, and cognitive/mind studies:** when overall functions are extended into judgments about subjects, experience, continuity, or meaning-bearing, reconsider what evidence can adequately support those propositions.

In Reverse: Which Competing Routes Most Deserve Further Testing

Opposing possibilities must remain open. A strongly centralized approach asks whether rapid cascades and shared failures require more concentrated and unified sensing and control. Strict optimization asks whether, with sufficiently well-defined goals, a single global model is more effective than multiple models, reopened boundaries, and local discretion. Resilience engineering asks whether maintenance and recovery should take priority over exit, decomposition, and transformation in critical infrastructure. Decentralization asks whether concentration itself is a primary source of common failures and domination. This paper does not preselect concentration or dispersion, unity or plurality. If competing structures consistently deliver better safety, recovery, distribution, and accountability under the same consequence levels, rights constraints, and timescales, this paper's claims about model pluralism, retreat, nested governance, or boundary reopening must be narrowed, rewritten, or abandoned.

Forward: From Reading to Joint Research and Practical Verification

- Develop a common representation for cross-scale causal maps, shared-dependency maps, and responsibility-handoff maps, testing whether they genuinely add explanation and prediction beyond ordinary fault trees, organization charts, and single-layer network models.
- Conduct reproducible fault injection in multi-agent and public-resource environments. Test how local fidelity, retries, shared credentials, and shared cloud/energy dependencies produce congestion, cascades, collusion, or exclusion of those without agents, and publish negative results.
- Build joint fast-variable/slow-variable datasets tracking trust, professional capabilities, ecological recovery, local knowledge, vendor dependency, and exit capacity over time, verifying when short-term success becomes long-term lock-in.
- Compare centralized, polycentric, nested, and hybrid governance in parallel for the same cross-domain problem, preregistering which cascade speeds, knowledge distributions, externalities, and rights conditions support which structures.
- Test actual costs of system retreat: compare service losses, recovery times, reductions in lock-in, and residual responsibilities under stopping, degradation, decomposition, migration, replacement takeover, and long-term transformation.

- Make indicator reflexivity testable: compare behavior and data generation before and after indicator publication, entry into resource allocation, and incentive changes, identifying when reflexivity matters and when it can be ignored.
- If new mechanisms, data, or counterexamples show that particular boundary and scale reviews add no value over time, remove them. If new evidence of subjects appears, return the system-subject interface to neighboring disciplines for renewed judgment.

The most valuable participation does not treat complex systems as a new explanation of everything. It uses mechanism models, cross-scale data, failure cases, negative results, and actual engineering validation to force this paper to narrow, correct, reconstruct, or even fail.

Further reading is not intended to keep readers inside a framework, but to let different specialties advance from their own positions. The aim is not for more people to finish reading Civilization Leap, but for more people to accomplish Civilization Leap together.

The Civilizational Future Chain: If Local Correctness Can Still Produce Overall Failure, What Civilization Can Truly Embody Feedback, Exit, and Long-Term Consequences

Vision is not evidence. The entry points below do not prove this paper established. They place its disciplinary judgments within a future civilization capable of operating, being lived in, and forming, for further testing. Readers may derive direction and aspiration from them, or reject the current picture and change it through stronger disciplinary facts, institutional proposals, or real outcomes.

Further question	Suggested reading	What becomes visible	Boundary to preserve
A complete picture of the future	Living Civilization: The Complete Public Trilogy in One Volume	Return complex-system boundaries, feedback, and externalities to ordinary life and civilization formation, asking what collective capability actually serves.	Overall coordination in a narrative does not prove a unique system optimum.
How systems capable of retreat operate	Living Civilization, Volume I: Principles for the Operation of a Future Civilization	Examine how resources, governance, AI, and public infrastructure form shared capability under conditions of stoppability, substitutability, and exit.	Operating order cannot replace engineering tests of resilience, reversibility, or safety.
How systems grow from experiments into civilization	Living Civilization, Volume III: The Formation of a New Civilization	Observe how local experiments, communities, institutions, and infrastructure spread, lock in, or are corrected.	Successful diffusion is not overall correctness; historical formation does not automatically generate normative justification.
System support for a complex civilization	Nebula: The Full Overview	Further test how making the whole visible, cross-scale coordination, slow variables, and polycentric structures enter a stable civilization.	Nebula has neither unified will nor executive sovereignty; the panorama is a system vision requiring verification.
Machine-side retreat and boundaries	AGI-COS Series Overview	Place system assurance, permissions, rollback, and retirement within machine-side engineering constraints.	Engineering safety does not automatically establish the goodness of society as a whole.

Return Links: [Complete Civilization Leap Library](#) | [Public Reading and Communication Hub](#)

For entry points such as Human-COS that do not yet have unambiguous, stable individual pages in the unified public mapping, this paper describes their functions only and does not guess URLs. Consult current public versions through the complete library when needed.

Appendix A Register of Source-Claim Boundaries

The table below is neither an authority ranking nor a total evidence score. Each source group supports limited conclusions and records judgments it cannot establish alone. Classical status, elegant models, large cases, and citation counts cannot replace objects, assumptions, and external validity. A/B/C distinguish empirical and frontier evidence; F discloses philosophical/normative commitments and is not weighted as a scientific evidence grade. Sources were verified through August 24, 2026.

Table A Supported Scope of 26 Source Groups, Original Claims, and Normative Commitments

Number	Source/object	Evidentiary status	Supported limited conclusion	Insufficient alone to establish
[1–5]	Open systems, control, hierarchy, and scales	A: classical theory/method	Conditional tools for openness, feedback, near-decomposability, and scale differences.	Possessing a model means possessing complete reality, goal-setting authority, or sovereignty.
[6]	Nonequilibrium self-organization	A: physicochemical theory	Open driven nonlinear systems can form dissipative structures.	Social order, freedom, legitimacy, or civilizational meaning.
[7]	Self-organized criticality	A: constructed models	SOC and scale-dependent events in specific slow-driving/dissipation/threshold models.	Diagnosing SOC from power laws or $1/f$ noise alone.
[8]	Complex adaptive systems	A: research framework	Adaptive actors, aggregation, nonlinearity, flows, and diversity.	Adaptation equals progress, justification, or subject standing.
[9–10]	Small worlds and preferential attachment	A: constructed models	The generative possibility of high clustering with short paths, and growth plus preferential attachment.	Real networks necessarily arise through that mechanism or entail particular governance conclusions.
[11]	Debate on identifying scale-free networks	B: competing definitions/tests	Tail definitions, fitting thresholds, and alternative distributions must be disclosed.	Settling in one sentence whether real networks are scale-free.
[12–13]	Complexity-stability	A models/B extrapolation	Random models and interaction structures can alter local stability conditions.	Complex real ecologies, finance, or civilizations must be stable or unstable.
[14–17]	Resilience, adaptation, transformation, and institutional robustness	A: concepts/frameworks	Disturbance responses must specify objects, functions, time windows, and institutional relations.	High resilience is inherently good, or transformation costs have already been allocated.
[18]	Review of critical slowing down	A mechanism/B extrapolation	Candidate early-warning statistical features may appear near specific bifurcations.	A universal civilization-criticality alarm without false positives.
[19]	Whole-lake warning experiment	A: controlled ecological evidence	Some indicators preceded a transition in one controlled whole-lake experiment.	All social or civilizational abrupt changes have the same signals.
[20–21]	Limits of warning detection	A: methodological studies	Data, models, and conditional selection can produce false positives and negatives.	Warning research is worthless, or no precaution is needed.

Number	Source/object	Evidentiary status	Supported limited conclusion	Insufficient alone to establish
[23–25]	Reflexivity of indicators and rankings	A narrow domain/B extrapolation	High-consequence indicators and public rankings may change measured processes.	All quantification, evaluation, or auditing necessarily fails.
[26–28]	Debates on weak and causal emergence	B: philosophy/formal models	Macro-explanations or measures may add value within particular models.	Macro-entities are universally ontologically independent or naturally entitled to governance priority.
[29–30,49]	Machine behavior, system safety, and normal accidents	A frameworks/B extrapolation	Machines' social effects require interdisciplinary research; complex interactions, tight coupling, control, and organizational interactions may all enter system failures.	Machines already possess consciousness; accidents are inevitable; safety equals ethics and legitimacy.
[31]	Dynamic segregation model	A: constructed model	Local preferences can generate macro-separation; motives cannot be inferred directly from macro-patterns.	Actual segregation involves only preferences, without power, institutions, or history.
[32–34]	Commons and polycentric governance	A cases/B institutional extrapolation	Open-access risks and conditional self-organization and polycentric arrangements.	Universal superiority of privatization, centralization, or decentralization.
[35]	Path dependence and technological lock-in	A: formal model/B empirical identification	Historical events can alter technological paths under increasing returns.	Any existing investment proves irreversible lock-in in reality.
[36]	NK landscapes and the edge of chaos	A construction/B extrapolation	Performance structures in particular random Boolean and adaptive-landscape models.	Life, organizations, or AGI universally ought to operate at the edge of chaos.
[37]	2003 US-Canada blackout	A: official accident investigation	Vegetation, software alarms, situational awareness, coordination, and protective actions jointly produced cascading.	A single topological model explains the entire accident.
[38]	2021 Texas cold wave	A: official accident investigation	Extreme cold, generation and fuel supply, and gas-electric dependencies caused failures and load shedding.	It is the same classical transmission-cascade type as 2003.
[39–40]	Propagation in financial networks	A models/A narrow-domain evidence/B extrapolation	Connectivity may reverse from robustness to fragility as shock magnitude changes; interbank exposures can propagate shocks.	Connectivity density alone determines all financial crises.
[41]	Epidemic-behavior feedback	A model/B policy extrapolation	Risk information can alter behavior and feed back into transmission.	The model has quantified real trust, stigma, or coercive-policy effects.
[42–46]	Climate tipping and the AMOC debate	A candidate mechanisms/B thresholds and timing	Multiple nonlinear mechanisms and model disagreements warrant precaution and model comparison.	A definite countdown, definite collapse, or definite non-collapse.
Original to this paper	Mechanism auditing, double rejection of optimization, cross-level firewall	C: methodological / normative bridging	Generate observable questions, prohibited inferences, and revisable interfaces.	Complexity science has proved all their institutional conclusions.

Number	Source/object	Evidentiary status	Supported limited conclusion	Insufficient alone to establish
Original to this paper	System retreat and critique of complexity-based sovereignty	C: institutional proposal	Exit, decomposition, migration, authorization, and sovereignty require independent review.	A unique governance solution, or systems can never become subjects.
Explicit commitments of this paper	Plural meaning, openness concerning subjects, and cross-level non-conversion	F: philosophical / normative commitments	Explain why complexity research must not silently generate rights, personhood, authorization, sovereignty, or a total account of meaning.	Disguising normative commitments as complexity-science facts, a unified value score, or fixed governance solutions.

Appendix B Checklist for Evidence, Extrapolation, and Cross-Level Auditing

Table B-1 A/B/C/F Argumentative Status and Permitted Uses

Evidentiary status	Composition	Permitted uses	Prohibited overreach
A Relatively stable support	Theoretical results within explicit objects, assumptions, and boundaries; controlled experiments; multisource facts; official accident investigations.	Explanation, validation, or targeted monitoring design within the declared scope.	Automatic extrapolation to civilizational scale, normative justification, or political authorization.
B Competing explanations	Model dependence, disputed definitions, domain differences, limited external validity, or competing mechanisms.	Model comparison, sensitivity analysis, scenarios, bounded trials, and continuing updates.	Presenting them as consensus, definite thresholds, universal laws, or sole causes.
C Speculative connections	Future AGI scenarios, normative bridging, institutional constructions, or subject hypotheses without actual objects yet.	Conceptual stress testing, research design, minimum precaution, and evidence preservation.	Presenting them as already occurring or confirmed, or as automatically producing rights and sovereignty.
F Philosophical / normative commitments	Openly attributed value premises, non-entailment boundaries, and commitments to public justification.	Explicit premises, strongest counterexamples, competing norms, real consequences, and conditions for revision.	Presenting them as empirical facts, complexity laws, unified value scores, or unquestionable sources of institutional authority.

B.1 Ten Checks before Extrapolation

- Are objects, environments, boundaries, state variables, spatial scales, and time windows explicit?
- Does the original source provide mathematical proof, simulation, experiment, observation, accident investigation, or an explanatory framework?
- Are the actual object and original model isomorphic in mechanism, networks, adaptability, power, and reflexivity?
- Have key concepts been substituted, for example by conflating SOC, the edge of chaos, critical slowing down, and general sudden changes?
- Have common external causes, simple aggregation, measurement artifacts, selection bias, and alternative distributions been compared?
- Does deployment or model publication alter behavior, resources, and the next round of data generation?
- Are distributions, those unable to exit, slow variables, and uncompensable consequences displayed alongside overall outcomes?

- Are empirical mechanisms silently converted into value judgments, legal responsibility, authorization, rights, or sovereignty?
- Are there discriminating counterfactuals, stopping conditions, revision triggers, and exit paths?
- Can readers independently refute this paper without accepting the rest of Civilization Leap?

B.2 Reviewing the System-Subject Interface

System-level memory, coordination, self-repair, goal maintenance, and collective action establish only the corresponding functional unity. Individually or jointly, these properties are insufficient to establish first-person experience, an ontological subject, legal personhood, representation, or public sovereignty. Systemness itself neither proves nor excludes subjecthood. When new experience reports, continuity evidence, or relations to one's own consequences appear, preserve states, compare manipulation and script explanations, replicate across systems, and refer them separately to studies of mind and life, epistemology, ethics, law, and political procedures.

Complexity research can observe which system conditions support or undermine subjects' lived practices, relational continuity, and spaces for meaning formation. It cannot interpret persistence, goal consistency, information integration, successful adaptation, or aggregate performance as meaning itself. A system maintaining goals does not mean it possesses meaning; improving subjects' welfare does not give it authority to interpret meaning.

References

- [1] von Bertalanffy L. The theory of open systems in physics and biology. *Science*. 1950;111(2872):23–29. <https://doi.org/10.1126/science.111.2872.23>
- [2] Wiener N. *Cybernetics: Or Control and Communication in the Animal and the Machine*. 2nd ed. Cambridge, MA: MIT Press; 1961. First edition 1948. <https://mitpress.mit.edu/9780262537841/>
- [3] Ashby WR. *An Introduction to Cybernetics*. London: Chapman & Hall; 1956. Digital edition: <https://ashby.info/>
- [4] Simon HA. The architecture of complexity. *Proceedings of the American Philosophical Society*. 1962;106(6):467–482. <https://doi.org/10.2307/985254>
- [5] Anderson PW. More is different. *Science*. 1972;177(4047):393–396. <https://doi.org/10.1126/science.177.4047.393>
- [6] Nicolis G, Prigogine I. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. New York: Wiley; 1977.
- [7] Bak P, Tang C, Wiesenfeld K. Self-organized criticality: an explanation of 1/f noise. *Physical Review Letters*. 1987;59:381–384. <https://doi.org/10.1103/PhysRevLett.59.381>; Bak P, Tang C, Wiesenfeld K. Self-organized criticality. *Physical Review A*. 1988;38:364–374. <https://doi.org/10.1103/PhysRevA.38.364>
- [8] Holland JH. Complex adaptive systems. *Daedalus*. 1992;121(1):17–30. <https://www.jstor.org/stable/20025416>
- [9] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;393:440–442. <https://doi.org/10.1038/30918>
- [10] Barabási A-L, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509–512. <https://doi.org/10.1126/science.286.5439.509>
- [11] Broido AD, Clauset A. Scale-free networks are rare. *Nature Communications*. 2019;10:1017. <https://doi.org/10.1038/s41467-019-08746-5>; Voitalov I, van der Hoorn P, van der Hofstad R, Krioukov D. Scale-free networks well done. *Physical Review Research*. 2019;1:033034. <https://doi.org/10.1103/PhysRevResearch.1.033034>
- [12] May RM. Will a large complex system be stable? *Nature*. 1972;238:413–414. <https://doi.org/10.1038/238413a0>
- [13] Allesina S, Tang S. Stability criteria for complex ecosystems. *Nature*. 2012;483:205–208. <https://doi.org/10.1038/nature10832>
- [14] Holling CS. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*. 1973;4:1–23. <https://doi.org/10.1146/annurev.es.04.110173.000245>
- [15] Walker B, Holling CS, Carpenter SR, Kinzig A. Resilience, adaptability and transformability in social–ecological systems. *Ecology and Society*. 2004;9(2):5. <https://doi.org/10.5751/ES-00650-090205>
- [16] Anderies JM, Janssen MA, Ostrom E. A framework to analyze the robustness of social-ecological systems from an institutional perspective. *Ecology and Society*. 2004;9(1):18. <https://doi.org/10.5751/ES-00610-090118>

- [17] Ostrom E. A diagnostic approach for going beyond panaceas. *Proceedings of the National Academy of Sciences*. 2007;104(39):15181–15187. <https://doi.org/10.1073/pnas.0702288104>
- [18] Scheffer M, Bascompte J, Brock WA, et al. Early-warning signals for critical transitions. *Nature*. 2009;461:53–59. <https://doi.org/10.1038/nature08227>
- [19] Carpenter SR, Cole JJ, Pace ML, et al. Early warnings of regime shifts: a whole-ecosystem experiment. *Science*. 2011;332:1079–1082. <https://doi.org/10.1126/science.1203672>
- [20] Boettiger C, Hastings A. Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface*. 2012;9:2527–2539. <https://doi.org/10.1098/rsif.2012.0125>
- [21] Boettiger C, Hastings A. Early warning signals and the prosecutor's fallacy. *Proceedings of the Royal Society B*. 2013;280:20131372. <https://doi.org/10.1098/rspb.2013.1372>
- [22] Conant RC, Ashby WR. Every good regulator of a system must be a model of that system. *International Journal of Systems Science*. 1970;1(2):89–97. <https://doi.org/10.1080/00207727008920220>
- [23] Goodhart CAE. Problems of monetary management: the U.K. experience. In: *Papers in Monetary Economics*. Reserve Bank of Australia Conference Volume; 1975/1976. Bibliographic record: <https://www.rba.gov.au/publications/rdp/1990/9013/conference-volumes.html>
- [24] Campbell DT. Assessing the impact of planned social change. 1976; reprinted in *Journal of MultiDisciplinary Evaluation*. 2011;7(15):3–43. <https://doi.org/10.56645/jmde.v7i15.297>
- [25] Espeland WN, Sauder M. Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*. 2007;113(1):1–40. <https://doi.org/10.1086/517897>
- [26] Bedau MA. Weak emergence. *Philosophical Perspectives*. 1997;11:375–399. <https://doi.org/10.1111/0029-4624.31.s11.17>
- [27] Hoel EP, Albantakis L, Tononi G. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*. 2013;110(49):19790–19795. <https://doi.org/10.1073/pnas.1314922110>
- [28] Eberhardt F, Lee L-L. Causal emergence: when distortions in a map obscure the territory. *Philosophies*. 2022;7(2):30. <https://doi.org/10.3390/philosophies7020030>
- [29] Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. *Nature*. 2019;568:477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- [30] Leveson NG. *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA: MIT Press; 2012. <https://doi.org/10.7551/mitpress/8179.001.0001>
- [31] Schelling TC. Dynamic models of segregation. *Journal of Mathematical Sociology*. 1971;1(2):143–186. <https://doi.org/10.1080/0022250X.1971.9989794>
- [32] Hardin G. The tragedy of the commons. *Science*. 1968;162:1243–1248. <https://doi.org/10.1126/science.162.3859.1243>
- [33] Ostrom E. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press; 1990. <https://doi.org/10.1017/CBO9780511807763>
- [34] Ostrom E. Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change*. 2010;20(4):550–557. <https://doi.org/10.1016/j.gloenvcha.2010.07.004>
- [35] Arthur WB. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*. 1989;99(394):116–131. <https://doi.org/10.2307/2234208>
- [36] Kauffman SA. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press; 1993. <https://doi.org/10.1093/oso/9780195079517.001.0001>
- [37] U.S.–Canada Power System Outage Task Force. *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. April 2004. <https://www.ferc.gov/industries-data/electric/overview/electric-reliability/power-blackout>
- [38] Federal Energy Regulatory Commission, North American Electric Reliability Corporation, and Regional Entities. *The February 2021 Cold Weather Outages in Texas and the South Central United States*. November 2021. <https://www.ferc.gov/media/february-2021-cold-weather-outages-texas-and-south-central-united-states-ferc-nerc-and>
- [39] Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. Systemic risk and stability in financial networks. *American Economic Review*. 2015;105(2):564–608. <https://doi.org/10.1257/aer.20130456>
- [40] Iyer R, Peydró J-L. Interbank contagion at work: evidence from a natural experiment. *Review of Financial Studies*. 2011;24(4):1337–1377. <https://doi.org/10.1093/rfs/hhp105>
- [41] Funk S, Gilad E, Watkins C, Jansen VAA. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*. 2009;106(16):6872–6877. <https://doi.org/10.1073/pnas.0810762106>
- [42] Lenton TM, Held H, Kriegler E, et al. Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences*. 2008;105(6):1786–1793. <https://doi.org/10.1073/pnas.0705414105>

- [43] Intergovernmental Panel on Climate Change. Climate Change 2021: The Physical Science Basis. Working Group I Contribution to the Sixth Assessment Report. Cambridge University Press; 2021. <https://www.ipcc.ch/report/ar6/wg1/>
- [44] Armstrong McKay DI, Staal A, Abrams JF, et al. Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science*. 2022;377:eabn7950. <https://doi.org/10.1126/science.abn7950>
- [45] Ditlevsen P, Ditlevsen S. Warning of a forthcoming collapse of the Atlantic meridional overturning circulation. *Nature Communications*. 2023;14:4254. <https://doi.org/10.1038/s41467-023-39810-w>; Author Correction. 2025. <https://doi.org/10.1038/s41467-025-63201-y>
- [46] Baker JA, et al. Continued Atlantic overturning circulation even under climate extremes. *Nature*. 2025. <https://doi.org/10.1038/s41586-024-08544-0>
- [47] Sterman JD. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science*. 1989;35(3):321–339. <https://doi.org/10.1287/mnsc.35.3.321>
- [48] Bar-Yam Y. Multiscale variety in complex systems. *Complexity*. 2004;9(4):37–45. <https://doi.org/10.1002/cplx.20014>
- [49] Perrow C. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books; 1984.

Declaration of Generative AI Use: Generative AI tools were used to organize materials and literature leads, design the structure, draft and polish portions of the text, search for consistency across manuscripts, check citation leads, and inspect layout. Generative AI outputs were not treated as empirical data, authoritative sources, or independent evidence. The author reviewed and approved the research questions, theoretical propositions, normative judgments, selection of materials, adoption of citations, and final manuscript, and bears responsibility for factual accuracy, citation authenticity, arguments, and all conclusions. AI is not listed as an author and does not bear academic responsibility. For formal submission, tool names, versions, dates of use, and specific scope will be added according to the target journal's requirements.