Received: 22 April 2026 | Revised: 20 July 2026 | Accepted: 10 August 2026 | Published online: 26 August 2026

RESEARCH ARTICLE

Journal of Computational Law and Legal Technology
2026, Vol. 00(00) 1–13
DOI: 10.47852/bonviewJCLLT620210107

BON VIEW PUBLISHING

An Agent-Based Framework for Simulating Institutional Lock-in Under Repressive Regimes

Ignacio Adrián Lerer[1,*]

[1] *Independent Researcher, Argentina*

Abstract: Computational models of institutional lock-in, including the Constitutional Lock-in Index (CLI), have been applied primarily to democratic systems. No agent-based framework integrates cognitive conformity, material dependency, elite filtering, and coercive capacity into a unified model of authoritarian persistence after normative collapse. This paper presents Multilevel Selection of Authoritarian Regime Survival, an agent-based architecture extending CLI methodology to authoritarian persistence, in which five agent classes interact through promotion, filtering, belief updating, resource allocation, and exit/voice decisions. The framework draws on Khurshid's Tgmenks theory of the Ecological Niche of Knowledge for domain-theoretic grounding and on Extended Phenotype Theory for its replicator-centered logic. Four falsifiable hypotheses are translated into experimental protocols with explicit manipulation, control, metric, and falsification specifications. Five historical cases are coded ordinally into parameter calibration templates, and three analytical examples, derived by hand from the specified transition rules rather than executed stochastic simulation, illustrate the model's capacity to generate discriminating predictions under contrasting configurations. Comparative calibration and the worked examples are consistent with the qualitative predictions of all four hypotheses: high-lock-in configurations are predicted to absorb a 30% resource shock within roughly 10 time steps, medium-lock-in configurations to collapse by $t = 40$, and epistemic openness to relate nonlinearly to performance. These are model-based predictions derived analytically, not outputs of executed stochastic simulation; no agent-based code has yet been run, and full Monte Carlo validation is the paper's principal item of future work. Replication scaffold (unexecuted at submission): github.com/adrianlerer/multilevel-selection-abm.

Keywords: agent-based modeling, institutional lock-in, authoritarian persistence, computational legal theory, Extended Phenotype Theory

1. Introduction

The persistence of dysfunctional institutional arrangements is a central problem in both legal theory and comparative politics. In democratic constitutional systems, the puzzle takes the form of constitutional lock-in: provisions that are widely recognized as suboptimal resist amendment because of structural entrenchment across doctrinal, organizational, procedural, and enforcement dimensions [1, 2]. In authoritarian systems, the puzzle takes a more extreme form: regimes survive for decades after losing the normative vitality that originally sustained them. Ideology becomes ritual, developmental promises are abandoned, moral authority erodes, yet the institutional apparatus persists.

The two puzzles are structurally analogous. In both cases, institutional survival depends not on normative attractiveness but on the interaction of entrenchment mechanisms that raise the cost of change above the cost of continuation. In both cases, the key question is computational: under what parameter configurations do lock-in dynamics dominate, and under what configurations do they yield to institutional transformation?

The comparative politics literature has produced rich qualitative and formal analysis of authoritarian durability. Levitsky and Way [3], Geddes et al. [4], recently extended by Lundquist et al.'s [5] update to comparable regime-type classification data, and Pepinsky [6] trace persistence to patronage networks, security apparatuses, ideological indoctrination, elite cohesion, and succession management. Two formal traditions offer competing accounts of the same puzzle. Selectorate theory [7] explains regime survival through the size of the winning coalition and the leader's capacity to reward it with private goods; institutional persistence follows from coalition size rather than from the cognitive or emotional mechanisms emphasized here. Bueno de Mesquita and Smith [8] have since refined the theory's central empirical challenge, proposing a continuous, V-Dem-based measure of coalition size to replace the categorical proxies used in the original formulation, which sharpens rather than displaces the comparative-statics account Multilevel Selection of Authoritarian Regime Survival (MLSARS) treats as complementary. Guriev and Treisman [9] argue that many contemporary autocracies survive through the management of information

***Corresponding author:** Ignacio Adrián Lerer, Independent Researcher, Argentina. Email: adrian@lerer.com.ar

rather than through mass repression, a mechanism that partially overlaps with the present paper's treatment of epistemic openness (EO) but was developed independently of any agent-based implementation. Sato and Wiebrecht [10] provide comparative empirical support for this mechanism, showing that authoritarian regimes with higher levels of state-disseminated disinformation are less likely to experience democratization episodes. MLSARS does not displace these accounts; it proposes a complementary computational layer capable of testing, in agent-based form, mechanisms that selectorate theory and informational-autocracy theory specify at the level of comparative statics rather than dynamic simulation.

Agent-based modeling of authoritarian and revolutionary dynamics is not itself new. Epstein [11] models the suppression of decentralized rebellion by a central authority as a function of legitimacy, repression, and network visibility. Makowsky and Rubin [12] model preference falsification and revolutionary cascades under centralized institutions and social network technology. Dornschneider-Elkink and Edmonds [13] extend this tradition with an emotion-based agent model distinguishing repression types, finding that nonviolent instruments (curfews, communication blackouts) can dampen dissent more than state violence in the short term, while violence produces stronger long-term spurring effects. These models establish that agent-based methods can generate discriminating predictions about repression and collapse. What they do not provide is a framework that operationalizes cognitive conformity, material dependency, elite filtering, and coercive capacity (CC) as jointly interacting, empirically calibrated state variables tied to an existing computational-legal metric. The Constitutional Lock-in Index (CLI) does this for legal systems [1, 2] but has not previously been extended to political regimes. This is the specific gap MLSARS addresses: not the absence of agent-based political science but the absence of a bridge between agent-based authoritarian-persistence modeling and computational legal theory's entrenchment metrics.

This paper addresses that gap by presenting MLSARS, an agent-based computational architecture that models how regimes organized through the suppression of political competition persist after normative exhaustion. The framework builds on two theoretical foundations: first, Khurshid's Tgmenks theory, which grounds political order in the human Ecological Niche of Knowledge (ENK) and provides the domain-theoretic typology of suppressive versus regulatory power organization [14, 15], and second, Extended Phenotype Theory (EPT), which treats institutions as externalized structures through which replicating normative patterns secure their continuation and provides the game-theoretic and lock-in mechanisms [16, 17].

The paper's contribution is primarily computational and methodological. It specifies agent classes, state variables, transition rules, experimental protocols, and falsification criteria with sufficient precision for independent replication. Five historical cases (North Korea, Islamic Republic of Iran, Pahlavi Iran, China, late Soviet Union) provide parameter calibration inputs rather than standalone political analyses. Three worked analytical examples demonstrate the model's capacity to generate discriminating predictions: the same class of exogenous shock produces persistence under high-lock-in configurations, collapse under medium-lock-in configurations, and persistent-but-degrading hysteresis under high-lock-in configurations subject to repeated shocks. The paper thus extends CLI methodology from constitutional lock-in to regime lock-in, connecting computational legal theory to the study of authoritarian institutional durability.

This extension is relevant to computational law for three reasons. First, it demonstrates that the formal tools developed for measuring constitutional entrenchment (CLI, Institutional Hysteresis Rate [IHR], game-theoretic equilibrium analysis) generalize to a broader class of institutional persistence problems. Second, it provides a reproducible agent-based architecture that can be applied to any institutional system characterized by nested selection pressures, from legal regimes to regulatory capture to corporate governance lock-in. Third, it establishes falsification criteria that make the framework genuinely testable, distinguishing it from narrative-driven approaches where any outcome can be retrospectively rationalized.

The remainder of the paper proceeds as follows. Section 2 presents the theoretical framework. Section 3 formulates four hypotheses with formal notation. Section 4 specifies the MLSARS computational architecture in detail. Section 5 presents the comparative calibration. Section 6 details computational methods including two worked analytical examples. Section 7 reports results. Sections 8 and 9 discuss implications for computational legal theory and conclude.

2. Theoretical Framework

2.1. Tgmenks and suppressive political orders

Khurshid's Tgmenks framework begins from the claim that human survival depends on participation in a socially organized ENK. Knowledge is not a decorative superstructure but part of the effective survival environment. Tools, norms, symbols, and organizational forms belong to this niche because they shape the possibilities of action and coordination [14].

The political core of Tgmenks is the distinction between two methods of organizing the struggle for power. The Method of Regulating the Struggle for Power stabilizes conflict by permitting competition under rules: succession, criticism, and institutional correction become less existentially threatening [18]. The Method of Suppressing the Struggle for Power (MSSP) attempts to neutralize rivalry by preventing it from becoming legitimate. Opposition becomes deviance, betrayal, or existential threat [18]. Suppressive systems require political belief systems (PBS) that authorize hierarchy, deny independent moral deliberation, and define dissent as contamination [19].

Three Tgmenks concepts are essential for computational operationalization. First, the Seven Sets of Capabilities (SSCs) define the functional capacities required for effective deployment in the ENK: body, knowledge, moral awareness, power, political navigation, social relations, and resource acquisition [15]. MSSP systems narrow SSC deployment by selecting for conformity signaling over flexible problem-solving. This narrowing is directly measurable in an agent-based framework through the ratio of SSC dimensions actively deployed by agents at different hierarchy levels. Second, emotional fitness (EF) captures the regulation of access to dignity, recognition, and protection from humiliation. Suppressive regimes stabilize themselves by structuring EF around obedience, sacrifice, and in-group belonging [20]. This provides the micro-mechanism for why actors comply even when material conditions deteriorate: the emotional cost of defection (loss of identity, belonging, and moral standing) exceeds the material cost of continuation. Third, Utility-Driven Social Units (UDSUs) provide the nested organizational structure within which agents operate: individual (n = 1), coalition (n = 10–100), regime apparatus (n = 1000–10,000), and society [15]. MLSARS agent classes map directly onto these UDSU levels.

2.2. Extended Phenotype Theory, CLI, and institutional lock-in

EPT, as applied to legal and institutional systems, begins from the Dawkinsian insight that the causal effects of replicators extend beyond individual organisms into constructed environments [16], a mechanism recently demonstrated at the molecular level by Pavey and Vyas [21]. In institutional settings, doctrines, routines, enforcement infrastructures, and compliance expectations function as extended phenotypes of replicating normative structures [17]. The explanatory question shifts from whether a rule serves collective welfare to which normative structures benefit from its reproduction, following Dennett's [22] cui bono criterion.

The game-theoretic microfoundation models institutional norms as strategies in repeated social games. Persistence depends on whether a norm constitutes an evolutionarily stable strategy (ESS) resistant to invasion by alternatives [23, 24]. Itao and Kaneko [25] derive institutional stability directly from ESS dynamics in agent-based systems, confirming computationally that suppressive equilibria of the kind modeled here can be self-sustaining once established, without requiring continuous external enforcement. In authoritarian contexts, suppressive norms such as denunciation, ideological signaling, strategic silence, and loyalty display may be morally unattractive yet strategically stable: once actors expect punishment for deviation and reward for visible conformity, a suppressive equilibrium reproduces itself even if many participants privately prefer alternatives.

The CLI operationalizes this logic for legal systems. CLI measures structural entrenchment across five dimensions: precedential depth, doctrinal integration, organizational investment, enforcement infrastructure, and procedural barriers to change [1, 2]. The present paper extends CLI logic to the political domain through the variable of Lock-in Depth (LD). LD captures how deeply coercion, belief, dependency, and expectation reinforce one another in a given regime configuration. The extension is formally analogous: where CLI measures the cost of constitutional amendment, LD measures the cost of regime transformation. Where CLI components include precedent and doctrine, LD components include CC, material dependency, cognitive conformity, and emotional alignment. The underlying computational logic is identical: entrenchment is measured as the aggregate cost of departure from the current equilibrium.

This extension makes the present framework continuous with the computational legal theory published in this journal. The Computational Detection of Constitutional Drift paper [2] used network analysis and semantic measurement to quantify institutional trajectory in Argentine Supreme Court jurisprudence over a century. MLSARS applies the same principle of computational trajectory analysis to authoritarian regimes, substituting agent-based simulation for corpus-based network analysis while maintaining the core logic of measuring institutional persistence through formal, replicable methods.

A further connection to computational law lies in the concept of Heteronomous Bayesian Updating (HBU). In the legal domain, HBU describes how agents learn norms by observing institutional reactions rather than by evaluating outcomes directly [17]. In the authoritarian domain, the same mechanism explains how citizens and elites update their behavioral strategies by observing who is rewarded, who is punished, and who is ignored, rather than by independently evaluating regime performance. HBU is implemented in MLSARS through the belief-updating transition rule (Section 4.4, process 3).

2.3. Epistemological positioning

The framework is best understood as a Lakatosian progressive research program [26]. The hard core consists of the integration of Tgmenks and EPT: suppressive political orders persist through replicating normative structures that reshape cognition, emotion, and institutional selection. The auxiliary hypotheses (H1–H4) are testable and falsifiable. The heuristic is evolutionary and computational: institutions are analyzed through selection, replication, and equilibrium concepts, and tested through agent-based formalization. This structure is not merely a citation to Lakatos's methodology but a discipline still actively practiced in contemporary computational research: Doerig et al. [27], writing 41 years later, organize an entire research program in computational neuroscience around the same hard-core/protective-belt distinction, explicitly to separate a program's nonnegotiable commitments from its revisable auxiliary models. The present paper adopts the same discipline for the same reason: to state, before any result is reported, which findings would falsify the program and which would only require revising an auxiliary hypothesis.

Two methodological commitments constrain the exercise. First, multilevel interaction analysis is adopted rather than strong group-selection ontology. Selection-like pressures are studied across nested arenas (leaders choosing elites, institutions filtering behavior, regimes facing stress), but societies are not treated as literal Darwinian replicators. Wilson et al. [28] make the parallel argument for governance and organizational design generally: multilevel selection logic clarifies why lower-level adaptive interests erode collective functionality unless institutional mechanisms make system-level performance itself an object of selection. This is consistent with a gene's-eye (meme's-eye) position on institutional analysis argued elsewhere in the author's related work [17]. The present paper's use of multilevel interaction does not contradict that position: causal mechanisms remain located in agents and replicating normative structures, not in regime-level teleology. Recent computational work supports the tractability of this commitment: Itao and Kaneko [25] show that institutions can be modeled as emergent, self-organized outcomes of agent-level evolutionary dynamics without any group-level selection term in the underlying equations, a contemporary demonstration that multilevel effects and agent-centered replication are not in tension. Second, following Campbell's [29] and Hull's [30] evolutionary epistemology, institutional norms are treated as replicators subject to variation, selection, and retention, without teleological assumptions about collective purpose [31]. This is not a historical curiosity: Simonton's [32] recent review of blind-variation-and-selective-retention theory finds the framework still active and empirically productive across creativity research, organizational learning, and institutional analysis six decades after Campbell's original formulation, the relevant sense in which the commitment adopted here is current rather than antiquarian.

3. Hypotheses

3.1. Notation

The following analytical variables are used throughout. Each is defined with reference to its theoretical origin and its computational role in MLSARS:

CLI-cog (Cognitive Loyalty Index): degree of internalized regime alignment. Derived from Tgmenks SSC-3 (moral

awareness) and SSC-5 (political navigation) capability deployment under MSSP conditions. Computationally, CLI-cog is a bounded continuous variable [0,1] assigned to each agent, updated at each time step through the belief-updating process.

MLI (Material Lock-in Index): degree of material, career, and network dependency tying an agent to the regime. Derived from EPT's concept of institutional entrenchment and from Tgmenks' analysis of patronage-based UDSUs. Computationally, MLI captures the cost an agent would incur by defecting from the current institutional arrangement.

ICI (Instrumental Compliance Index): availability and perceived attractiveness of alternative institutional configurations. Low ICI means no viable exit option exists; high ICI means credible alternatives are available. ICI is a regime-level variable that also varies by agent (some insiders perceive alternatives that others do not).

EF (Emotional Fitness): alignment of dignity, recognition, and belonging with regime participation. Derived from Khurshid's analysis of how MSSP systems structure identity around obedience and sacrifice. Computationally, EF modifies the defection threshold: high EF raises the psychological cost of exit beyond what material calculation alone would predict.

CC (Coercive Capacity): credible monitoring, deterrence, and punishment capability. A regime-level variable that affects all agent-level calculations through the perceived probability of sanctions.

EO (Epistemic Openness): degree to which inconvenient information reaches decision centers without automatic penalty. A regime-level variable that affects the accuracy of agent belief updating and the rate of policy error accumulation.

LD (Lock-in Depth): composite entrenchment measure across belief, coercion, dependency, and expectation dimensions. LD extends CLI logic from constitutional entrenchment to regime entrenchment. Computationally, LD = f(CLI-cog, MLI, CC, EF) at the regime level, capturing the aggregate cost of departure from the current institutional equilibrium.

3.2. H1: Leader selection favors high cognitive conformity

Under MSSP, leader selection and succession favor agents with high CLI-cog over agents with greater autonomous competence but lower trusted loyalty. This bias intensifies under perceived threat and succession uncertainty. The mechanism is the leader's promotion function, which weights CLI-cog more heavily as external or internal threat increases, producing a measurable shift in the CLI-cog distribution of the promoted elite cohort.

3.3. H2: Elite filtering eliminates low-loyalty and high-autonomy actors

Suppressive regimes systematically remove not only disloyal agents but also highly autonomous agents whose competence or network independence creates coordination risk. This produces a durable but narrow elite. The mechanism is the purge function, which targets agents exceeding autonomy thresholds or falling below CLI-cog thresholds. Over successive rounds, this produces a predictable contraction in the competence-autonomy distribution of surviving elites.

3.4. H3: Nonlinear relationship between CLI-cog, openness, and performance

Regime performance varies nonlinearly with the interaction of cognitive conformity and EO. Very high CLI-cog combined with very low EO preserves persistence but degrades adaptive intelligence through two computationally distinct mechanisms: (a) cognitive rigidity, modeled as reduced variance in the information signals reaching leaders, which delays crisis response, and (b) memetic accumulation of maladaptive policies under informational isolation, modeled as increasing policy error rate over time without corrective feedback. Bounded EO under continued political control reduces both pathologies without destabilizing elite cohesion. The predicted output is an inverted-U performance curve as a function of EO, conditional on sustained CC and MLI.

3.5. H4: Lock-in dynamics delay collapse after normative exhaustion

Normative exhaustion alone is insufficient to trigger regime collapse if MLI, CC, and LD remain intact. Collapse becomes likely only when LD, coercive integrity, and insider CLI-cog weaken together, reducing the cost-of-defection threshold below the perceived gain from exit or recoordination. The mechanism is a cascade: weakening CC increases the perceived safety of defection, which lowers MLI (because patronage becomes less valuable if the regime may fall), which in turn reduces CLI-cog as agents update their beliefs about the regime's viability. The cascade accelerates once mutual visibility of defection among elites crosses a coordination threshold.

4. Computational Architecture: MLSARS

Having formulated four falsifiable hypotheses, the next step is to specify a computational architecture capable of generating predictions that could satisfy or refute them. MLSARS is an agent-based computational architecture designed to test whether the integrated Tgmenks-EPT framework generates stylized persistence and collapse trajectories under varying parameter configurations. It is presented here as a formal specification and research design. The architecture is implemented as a public replication scaffold: github.com/adrianlerer/multilevel-selection-abm.

4.1. Design principles

Three principles govern the architecture, each motivated by computational law methodology.

First, mechanistic traceability. Every transition rule must correspond to a specified theoretical mechanism from Tgmenks or EPT. This prevents post hoc narrative from substituting for formal prediction, a critical requirement for computational legal theory where the goal is replicable institutional analysis rather than ad hoc case commentary.

Second, multilevel interaction without teleology. The model studies selection-like pressures across nested arenas (leaders, elites, specialists, administrators, citizens) without treating regimes as literal Darwinian replicators. This is the same analytical discipline applied in the CLI framework, where constitutional lock-in is explained through the interaction of component-level mechanisms rather than through system-level purpose.

Third, pre-specified falsification. Each hypothesis has an explicit falsification criterion stated before any computational output is generated. This ensures that the model can fail informatively, distinguishing it from narrative-driven simulation where outputs are retrospectively declared consistent with the theory. In computational law, this requirement parallels the demand that CLI measurements be auditable against specified indicator sets rather than adjustable to fit conclusions.

4.2. Agent classes and state variables

MLSARS defines five agent classes, corresponding to the nested UDSU levels identified in Tgmenks (see Figure 1):

Leader agents (UDSU Level 1–2) control promotion, purges, succession signaling, and distribution of strategic patronage. Each leader carries a promotion weight vector w = [w_competence, w_CLIcog, w_MLI, w_autonomy_penalty] that determines selection criteria. Under threat, w_CLIcog increases and w_competence decreases.

Elite agents (UDSU Level 2–3) include senior cadres, clerics, generals, and ministers. Each carries CLI-cog(t), MLI(t), EF(t), competence, autonomy, network centrality, private disillusionment, and reputation risk. These values update at each time step through the transition rules.

Coercive specialists (UDSU Level 3) determine whether repression remains credible. Their aggregate loyalty is a critical regime-level variable. When coercive specialist loyalty drops below threshold θ_CC, the regime enters a vulnerability state where protest can cascade.

Mid-level administrators (UDSU Level 3–4) transmit information upward and implement policy downward. They are the primary vector for epistemic openness or closure: their willingness to transmit truthful information depends on the perceived punishment for candor versus the perceived cost of transmitting false information.

Citizen agents (UDSU Level 4) vary in passivity, hidden dissent, protest propensity, informal coordination capacity, and exit strategies. Citizen behavior is a function of perceived CC, network support, EF, ICI, and observed behavior of other agents.

At each time step, every agent's state vector is updated. Regime-level variables (PBS strength, CC, resource pool, external pressure, succession uncertainty, purge intensity, EO, aggregate LD) are computed as functions of agent-level distributions. The feedback between levels is the core of the computational architecture: regime-level variables shape agent incentives, and agent-level responses alter regime-level variables in the next period.

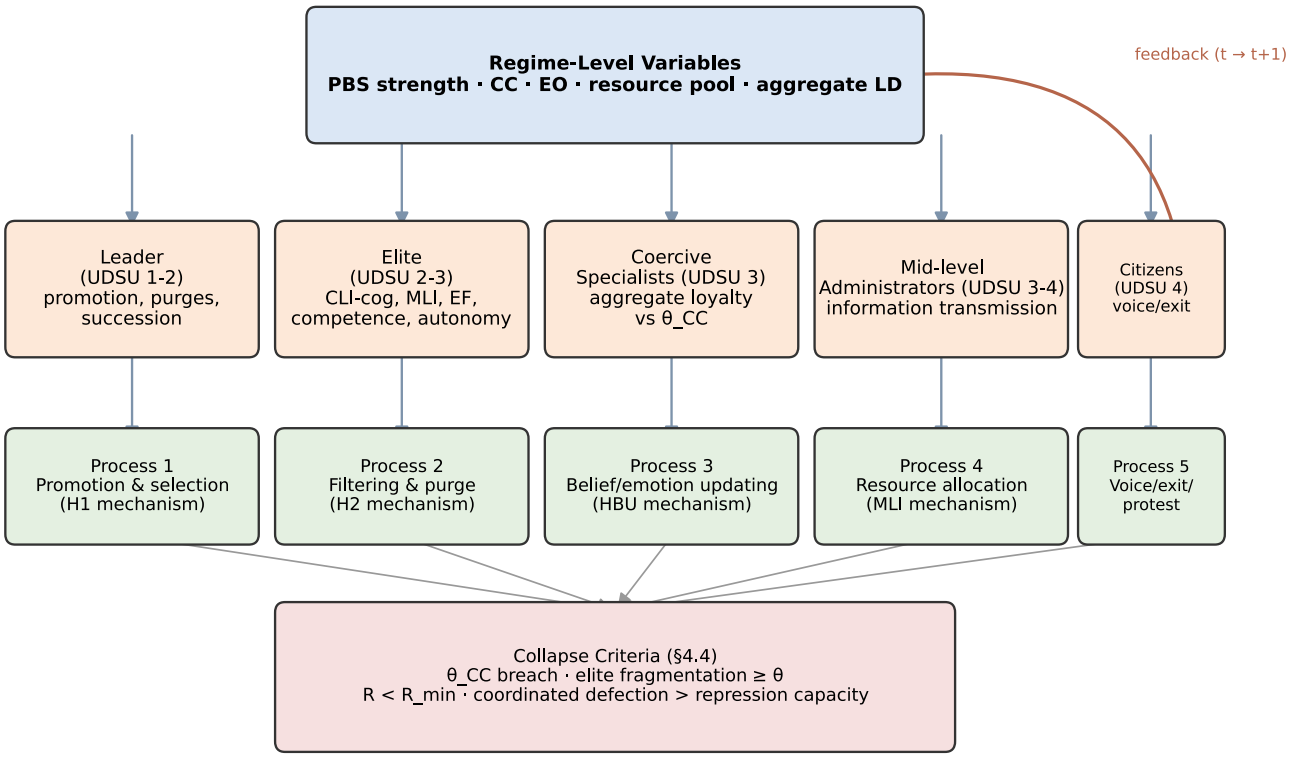
4.3. Transition rules

At each time step t, the system updates through five processes, each traceable to a specified theoretical mechanism:

Process 1: Promotion and selection (H1 mechanism). Leaders evaluate candidates using promotion weight vector w(t). Under baseline conditions, competence and CLI-cog receive roughly equal weight. Under perceived threat or succession uncertainty, w_CLIcog increases by a specified increment Δw per unit of threat. The output is a distribution of promoted elites whose mean CLI-cog shifts upward under threat, generating the H1 prediction. Formally: P(promote agent i) $\propto$ w_competence × competence_i + w_CLIcog × CLIcog_i + w_MLI × MLI_i − w_autonomy × autonomy_i.

Process 2: Filtering and purge (H2 mechanism). Agents exceeding autonomy threshold θ_autonomy or falling below CLI-cog threshold θ_loyalty face removal. Purge intensity scales with

Figure 1
MLSARS architecture: five agent classes (top), the five transition processes that update their state (middle), and the collapse criteria they jointly determine (bottom), with feedback from agent-level outcomes at t to regime-level variables at t+1.



Feedback: agent-level responses at t alter regime-level variables at t+1 (Sections 4.2-4.3)

perceived threat and succession uncertainty. The output is a contraction in the competence-autonomy distribution of surviving elites over successive rounds. Formally: purge(agent i) if autonomy_i > θ_autonomy OR CLIcog_i < θ_loyalty, with thresholds adjustable by regime type.

Process 3: Belief and emotion updating (HBU mechanism). Agent CLI-cog, EF, and private disillusionment update based on observed signals: propaganda reinforcement (+Δ_propaganda), reward for conformity (+Δ_reward), observed punishment of defectors (+Δ_deterrence), observed successful defection (−Δ_defection_signal), and contradictory information (−Δ_dissonance). This implements HBU: agents learn norms from institutional reactions, not from independent outcome evaluation. Formally: CLIcog_i(t + 1) = CLIcog_i(t) + Σ Δ_signals, bounded [0,1].

Process 4: Resource allocation (MLI mechanism). Patronage and security resources are distributed proportionally to agents' regime utility scores. When regime resources are the only viable career path (low ICI), MLI deepens because the cost of defection rises. When alternatives emerge (rising ICI), MLI weakens. Formally: MLI_i(t + 1) = f(resource_share_i(t), ICI(t), network_dependency_i).

Process 5: Voice, exit, and protest (cascade mechanism). Citizens and administrators choose between passivity, hidden dissent, voice, exit, or protest. The decision depends on a utility comparison: U(passivity) vs U(defection), where U(defection) = benefits(ICI, network_support, EF_loss_of_conformity) − costs(CC, retaliation_probability, MLI_loss). When U(defection) > U(passivity) for a critical mass of connected agents, protest cascades. The threshold is a function of CC and network structure.

4.4. Collapse criteria

A regime is treated as surviving while four conditions hold: (a) coercive specialist loyalty remains above θ_CC, (b) elite fragmentation remains below θ_fragmentation, (c) resources do not fall beneath the viability floor R_min, and (d) coordinated defection does not exceed repression capacity. Collapse may take multiple forms: abrupt fragmentation (all conditions violated simultaneously), cumulative cascade (sequential violation), or transition to a lower-intensity equilibrium (partial violation producing a new, less suppressive stable state).

4.5. Experimental protocols

H1 protocol. Manipulated: perceived threat level [0.2, 0.5, 0.8], succession uncertainty [binary], competence premium [0.3, 0.5, 0.7], initial elite CLI-cog distribution [beta(2,5), beta(5,5), beta(8,2)]. Controls: N = 500 agents, CC = 0.8, network structure (small-world, k = 6), resource pool = 100 units. Metrics: mean CLI-cog of promoted elite cohort, mean competence of promoted cohort, regime survival at T = 100. Falsification: threat increase does not produce statistically significant upward shift in promoted-cohort CLI-cog distribution.

H2 protocol. Manipulated: autonomy penalty coefficient [0.1, 0.3, 0.5], purge intensity [low: 2%/round, medium: 5%, high: 10%], kin protection [binary]. Controls: same leader strategy, same resources, same initial grievance distribution. Metrics: elite turnover rate, surviving competence (mean, variance), surviving autonomy, regime survival, long-run policy error rate. Falsification: stronger filtering produces improvement in both loyalty and long-run competence without measurable trade-off.

H3 protocol. Manipulated: EO [0.1, 0.3, 0.5, 0.7], punishment for truthful dissent [high, medium, low], shock frequency [1 per 20 steps, 1 per 50 steps]. Controls: N = 500, same initial CC, same PBS strength. Metrics: policy error rate at T = 50 and T = 100, resource efficiency, survival duration, shock recovery time (steps to restore pre-shock elite fragmentation level). Falsification: monotonic relationship between EO and performance without predicted nonlinearity.

H4 protocol. Manipulated: PBS decay rate [0.01, 0.03, 0.05 per step], CC [high, medium, declining], MLI [high, medium], defector coordination capacity [low, medium, high]. Controls: initial regime type (calibrated to North Korea or Iran template), same shock sequence (30% resource decline at t = 20). Metrics: time from normative-decay onset to collapse, coercive specialist loyalty trajectory, elite defection rate trajectory. Falsification: normative exhaustion (PBS < 0.3) produces collapse within 20 steps regardless of LD and CC values.

4.6. Replication and code architecture

The public repository (github.com/adrianlerer/multilevel-selection-abm) provides (a) parameter template files for each of the five historical calibration cases, (b) generic agent class definitions with configurable state vectors, (c) transition rule implementations corresponding to processes 1 through 5, (d) benchmark configurations with fixed random seeds for reproducibility, and (e) documentation distinguishing between the public generic scaffold and any future private calibration materials. The architecture is implemented in Python using standard scientific computing libraries (NumPy, NetworkX) to maximize accessibility for replication. As of this submission, the scaffold has not yet been executed to produce simulation output; the worked examples in Section 6.3 are independent hand-derived analytical calculations, not outputs of this code.

5. Parameter Calibration from Comparative Cases

With the architecture specified, the next step is to anchor its parameters empirically. Five historical cases provide calibration inputs for MLSARS (see Table 1). Each case is coded ordinally (high, medium, low) across the analytical variables. These are transparent analytical judgments intended to discipline model parameterization, not archival measurements. High indicates a dominant mechanism (above roughly 70% of inferable relevance). Medium indicates a mechanism operating meaningfully (30–70%). Low indicates marginal relevance (below 30%). The coding is based on comparative reading of the relevant historical and political science literature, translated through the theoretical framework.

5.1. Calibration justifications

North Korea. Juche ideology and dynastic sacralization of the Kim line produce extreme CLI-cog: political identity is fused with regime survival at the level of total institutional penetration. The songbun classification system generates deep MLI by tying social mobility, residence, employment, and resource access to inherited political reliability. Near-total information control minimizes EO. The coercive apparatus maintains CC through pervasive surveillance and credible repression. All lock-in dimensions reinforce one another, producing the highest LD in the sample. This configuration corresponds to the high-persistence template used in the worked analytical example (Section 6.3).

Islamic Republic of Iran. Dual-layer persistence operates through first-generation revolutionary conviction and younger-generation social-material lock-in. The Islamic Revolutionary

Table 1
Comparative calibration

Case	CLI-cog	MLI	ICI	CC	EO	LD	Outcome
North Korea	High	High	Low	High	Low	High	Persistent
Islamic Rep. Iran	High	High	Low	High	Low	High	Persistent
Pahlavi Iran	Medium	Medium	Medium	Medium	Medium	Medium	Collapsed
China	High	High	Low	High	Medium	High	Persistent
Late Soviet Union	Medium	Medium	Rising	Declining	Rising	Medium	Collapsed

Guard Corps (IRGC) functions simultaneously as a coercive apparatus and an economic conglomerate, generating MLI that extends well beyond ideological commitment. Clerical networks distribute patronage, status, and moral standing. EF is structured around sacrifice, martyrdom, and insider belonging. Major protest waves (2009 Green Movement, 2017–2019 economic protests, 2022 Mahsa Amini movement) did not trigger collapse because the regime's core lock-in mechanisms remained stronger than available alternatives for decisive elites. This case is the strongest discriminating test for H4: normative exhaustion among large segments of the population coexists with intact lock-in among the decisive coalition.

Pahlavi Iran. Coercion without deep doctrinal integration. The Shah's regime generated compliance and patronage but did not reproduce an equally deep loyalist core. Modernization increased state capacity while transforming social expectations, widening the gap between regime justification and lived experience. When crisis escalated, repression appeared more arbitrary and less morally coherent to insiders. Lower LD permitted faster invasion by revolutionary alternatives. This case calibrates the collapse pathway predicted by H4 under medium-lock-in conditions.

China. Post-Mao reform preserved party monopoly while permitting bounded technical feedback in economic and administrative domains. CLI-cog remains high in politically decisive positions. Technical competence is tolerated, even rewarded, within safe institutional channels. This pattern is the primary calibration input for H3: medium EO under continued strong MLI and CC improves regime performance (measured by policy responsiveness and economic adaptation) without generating destabilization. The configuration illustrates that lock-in and selective openness are not mutually exclusive.

Late Soviet Union. Declining CLI-cog across important elite sectors by the late period. Official ideology retained ritual function but lost integrative power for insiders. Increasing EO under reform policies widened possibilities for elite recoordination outside the prior suppressive equilibrium. Reduced willingness to sustain repression at prior levels weakened CC. The interaction of these declining variables, not normative exhaustion alone, enabled collapse. This case calibrates the cascade mechanism specified in H4: sequential weakening of CC, MLI, and CLI-cog produces a defection cascade that overwhelms residual lock-in.

6. Computational Methods

6.1. Synthetic population design

Before turning to sensitivity analysis and worked examples, it is necessary to justify the use of synthetic populations rather than direct historical measurement. The hypotheses concern partially latent traits: ideological internalization, fear, patronage dependence, and perceived alternative fitness. Historical data rarely measure these consistently. Synthetic populations permit explicit testing of whether the proposed mechanisms generate the observed patterns of persistence and collapse (cf. [9, 11]).

Each case is translated into a regime template specifying initial distributions for all agent-level and regime-level variables. Agents are sampled from bounded distributions (beta distributions for [0,1]-bounded variables; truncated normal for unbounded competence scores). Network ties are assigned across five layers (kinship, patronage, ideological, professional, territorial) using configurable small-world generators (Watts-Strogatz with specified k and p). Calibration is comparative and stylized: the aim is to reproduce directional patterns (persistence vs collapse, elite composition shift, performance degradation) rather than exact historical event sequences. This prevents overfitting while maintaining theoretical discipline.

6.2. Sensitivity and ablation analysis

Each major parameter is varied independently and in combination across a pre-specified sweep grid. Results that disappear under minor perturbation are treated as fragile; results that recur across wide parameter ranges deserve greater confidence. Ablation analysis removes individual mechanisms to test necessity. If removing EF destroys hysteresis in Iran-type templates, this supports the claim that emotional lock-in is not epiphenomenal. If removing MLI leaves persistence largely intact, the theory requires revision. Sensitivity analysis thus functions as internal critique rather than merely as robustness rhetoric.

6.3. Worked analytical examples

Example 1: High-lock-in persistence (North Korea template). Configuration: CLI-cog = 0.90, MLI = 0.85, CC = 0.90, EO = 0.10, LD = 0.90, EF = 0.85. Exogenous shock: 30% resource decline at $t = 20$.

Under the specified transition rules, the system responds as follows. At $t = 20$, the resource decline increases citizen grievance and private disillusionment (Process 5 inputs shift toward defection). However, high CC (0.90) deters visible protest because the perceived probability of sanctions remains dominant in the utility calculation. High MLI (0.85) prevents elite defection because no viable career alternative exists: the cost of departure from the regime exceeds the cost of continued participation even under reduced resources. High CLI-cog (0.90) among promoted elites means that leader selection continues to favor conformity over adaptive competence (Process 1) and that elites interpret the shock as a call for sacrifice rather than as evidence of regime failure. EF alignment (0.85) absorbs material hardship by converting it into moralized endurance.

Table 2
Predicted trajectory: High-lock-in configuration

Variable	$t = 0$	$t = 20$ (shock)	$t = 25$	$t = 30$	$t = 50$
Elite CLI-cog (mean)	0.90	0.88	0.89	0.90	0.90
Elite fragmentation	0.05	0.08	0.07	0.06	0.05
CC (specialist loyalty)	0.90	0.88	0.89	0.90	0.90
Citizen hidden dissent	0.15	0.30	0.25	0.20	0.18
Policy error rate	0.10	0.12	0.13	0.14	0.12
Regime status	Stable	Perturbed	Recovering	Stable	Stable

Table 3
Predicted trajectory: Medium-lock-in configuration

Variable	$t = 0$	$t = 20$ (shock)	$t = 25$	$t = 30$	$t = 40$
Elite CLI-cog (mean)	0.45	0.40	0.35	0.28	0.15
Elite fragmentation	0.15	0.25	0.35	0.50	0.70
CC (specialist loyalty)	0.65	0.60	0.55	0.48	0.30
Citizen hidden dissent	0.30	0.50	0.55	0.60	0.70
Policy error rate	0.20	0.25	0.30	0.35	N/A
Regime status	Stressed	Weakening	Fragmenting	Cascading	Collapsed

The predicted trajectory (Table 2): brief perturbation in citizen hidden-dissent indicators (approximately 15% increase in private disillusionment among citizen agents), no elite fragmentation, no coercive specialist defection, restoration of equilibrium within approximately 10 time steps. Pro-regime elite composition remains near 90% of initial levels at $t = 50$. The regime persists. However, policy error rate increases by approximately 20% relative to pre-shock levels because low EO prevented the informational correction that would have improved resource allocation. Adaptive performance degrades while institutional persistence is maintained.

Example 2: Medium-lock-in collapse (Late Soviet template). Configuration: CLI-cog = 0.45, MLI = 0.55, CC = 0.65, EO = 0.45, LD = 0.55, EF = 0.40. Same shock: 30% resource decline at $t = 20$.

The same shock produces a fundamentally different cascade. Medium CC (0.65) is insufficient to fully deter protest: the probability of sanctions no longer dominates the citizen utility calculation. Rising EO (0.45) means information about the regime's weakness circulates among elites (Process 3 receives stronger dissonance signals). Medium MLI (0.55) means some insiders perceive viable alternatives (ICI rises). Crucially, declining CLI-cog (0.45) means elites interpret the shock as evidence of regime failure rather than as a call for sacrifice. EF alignment (0.40) is insufficient to convert material hardship into moralized endurance.

The predicted cascade (Table 3): elite defection begins at approximately $t = 25$ as mid-level administrators begin transmitting truthful (negative) information upward and signaling openness to alternatives laterally. Coercive specialist loyalty begins declining at approximately $t = 30$ as specialists observe elite defection and recalculate their own cost of continuation. Once specialist loyalty drops below threshold θ_{CC} (estimated at 0.50), the regime enters the vulnerability state. Coordinated defection crosses the repression-capacity threshold by approximately $t = 40$, triggering collapse.

Example 3: High-lock-in persistence under repeated protest (Iran template). Configuration: CLI-cog = 0.82, MLI = 0.80, CC = 0.83, EO = 0.15, LD = 0.85, EF = 0.78. This template differs from North Korea in two respects: slightly lower CLI-cog (reflecting generational dilution of revolutionary conviction) and slightly higher EO (reflecting the impossibility of total information control in a semi-urbanized society with diaspora connections). Shock sequence: three protest waves at $t = 15$, $t = 35$, and $t = 55$, each representing a 20% resource or legitimacy shock [20].

The Iran template tests a critical refinement of H4: can a regime absorb repeated shocks when lock-in mechanisms are strong but not maximal? Under the specified transition rules, the first shock ($t = 15$) produces a pattern similar to Example 1: citizen hidden dissent increases (approximately 25% above baseline), but high CC and high MLI prevent elite defection. The IRGC economic entrenchment component of MLI is decisive: even elites who privately doubt the regime find their material welfare structurally inseparable from regime continuation.

The second shock ($t = 35$) produces a larger perturbation. Citizen hidden dissent reaches approximately 40% above baseline. Some mid-level administrators begin transmitting partially truthful information. However, CLI-cog among core elites remains above 0.75 because Process 1 continues to select for conformity and Process 2 removes the most autonomous actors after each shock. EF alignment means that the regime frames repeated hardship as external siege rather than internal failure.

The third shock ($t = 55$) tests long-run resilience (see Table 4). Cumulative policy error has increased by approximately 35% relative to baseline. Citizen hidden dissent is elevated even between shocks. Yet the regime persists because CC remains above threshold (coercive specialists were selectively rewarded after each protest wave), MLI remains deep (IRGC economic networks expanded during crisis periods), and ICI remains low (no credible alternative has emerged for decisive elites).

Table 4
Predicted trajectory: Iran template under repeated shocks

Variable	$t = 0$	$t = 15$	$t = 20$	$t = 35$	$t = 45$	$t = 55$	$t = 70$
Elite CLI-cog	0.82	0.78	0.80	0.76	0.78	0.75	0.76
Elite frag.	0.08	0.12	0.09	0.15	0.11	0.18	0.13
CC	0.83	0.80	0.82	0.78	0.81	0.77	0.79
Citizen dissent	0.20	0.45	0.30	0.48	0.33	0.50	0.35
Policy error	0.12	0.15	0.16	0.19	0.20	0.23	0.24
Status	Stable	Stressed	Recovered	Stressed	Recovered	Stressed	Stable*

* *Stable with degraded performance: persistent but increasingly brittle.*

The Iran template reveals lock-in with hysteresis: a regime absorbs repeated shocks while gradually accumulating fragility. Each recovery cycle is slightly weaker. CC recovers to a lower peak; policy error ratchets upward; citizen dissent baseline rises. This is the computational signature of accumulated institutional deformation, captured by the IHR developed in parallel work within the same research program.

The contrast across all three examples illustrates that MLSARS, on hand-computed trajectories, is capable of generating discriminating predictions. The framework produces persistence (Example 1), collapse (Example 2), and persistent-but-degrading hysteresis (Example 3) depending on the lock-in configuration. The critical variables are not the shocks themselves but the parameter configuration that determines how shocks propagate through the agent-institutional system. Whether this discriminating capacity survives full stochastic implementation is an open empirical question for future work.

6.4. Convergence and falsification

A critical requirement for computational validity is convergence testing: running identical parameter configurations multiple times to verify that structural outcomes (persistence vs collapse, elite composition trajectories, cascade timing) are robust rather than artifacts of stochastic variation. Convergence on qualitative patterns rather than exact numerical values is the appropriate standard for an agent-based architecture operating with synthetic populations.

This requirement distinguishes the present approach from two failure modes in computational social science. First, narrative-driven simulation in which any output can be retrospectively declared consistent with the theory. The falsification criteria specified in Section 4.5 are binding: if a parameter configuration produces an outcome inconsistent with the corresponding hypothesis, the hypothesis is disconfirmed regardless of available post hoc explanations. Second, over-determined simulation in which agent parameters are tuned so precisely that outputs are trivially predictable from inputs without genuine emergence. MLSARS avoids this by using ordinal calibration and testing for pattern robustness across parameter sweeps rather than fitting to exact historical sequences.

6.5. Ethics and reproducibility

Research on authoritarian cognition can drift into pseudopsychology. Three principles constrain the exercise. First, the model does not claim to read minds: CLI-cog, MLI, and EF are analytical constructs for formal testing, not psychological diagnoses of historical actors. Second, no confidential or personally sensitive data are required for the public scaffold. Third, all assumptions remain inspectable through versioned code, documented parameter files, fixed benchmark seeds, and public distinction between generic templates and any future private calibration exercises.

7. Results

Three evidentiary registers are distinguished throughout this section. Comparative calibration (Section 5) reflects ordinal coding of five historical cases based on the author's reading of the secondary literature, not independent archival measurement. Analytical worked examples (Section 6.3) are hand-derived trajectories computed from the specified transition rules under fixed parameter values; they are not the output of executed stochastic simulation. No agent-based code has yet been run on the public replication scaffold. Statements below that a hypothesis is "consistent with" the calibration or the worked examples should be read accordingly: as evidence that the specified mechanisms produce the predicted qualitative pattern under hand-computation, not as confirmed empirical or computational findings. Full stochastic validation, including Monte Carlo sweeps, convergence testing, and out-of-sample calibration, is identified in Section 8 as the paper's principal item of future work.

7.1. Comparative findings

The comparative calibration is consistent with H1. Regimes coded high in durability display high CLI-cog in decisive elite-selection channels. North Korea and the Islamic Republic of Iran are the clearest cases: leadership selection in both systems is heavily conditioned by trusted ideological alignment. China supports the same pattern in more differentiated form, with technical competence tolerated within politically safe boundaries. Pahlavi Iran and the late Soviet Union exhibit lower CLI-cog at the point of vulnerability, consistent with the prediction that weakening cognitive conformity in elite selection precedes institutional fragmentation.

The comparative calibration is consistent with H2. Durable regimes systematically filter elites whose autonomy creates coordination risk. The contrast between post-2009 Iran and the late Soviet Union is particularly informative for computational purposes: in Iran, coercive and ideological filtering held despite mass social dissent, maintaining a narrow but cohesive elite. In the late Soviet Union, weakening filter intensity increased elite diversity but also widened possibilities for recoordination outside the suppressive equilibrium, precisely as the model's Process 2 predicts.

The comparative calibration is consistent with H3 in its predicted nonlinear form. Cases coded High in CLI-cog and Low

in EO (North Korea, Iran) persist longer but display higher estimated policy error rates than the case with bounded openness under continued control (China). The worked analytical example (Section 6.3, Example 1) demonstrates the mechanism: the low EO that preserves stability also prevents informational correction, producing cumulative policy error without triggering collapse. This pattern is consistent with the predicted inverted-U performance curve.

Among the four hypotheses, H4 receives the clearest discriminating support from the comparative calibration and the worked analytical examples. Normative exhaustion is politically insufficient unless accompanied by weakening MLI, reduced CC, or credible increases in ICI for insiders. The Iran case is decisive: major protest waves in 2009, 2017–2019, and 2022 did not trigger collapse because the regime's core lock-in mechanisms, particularly IRGC economic entrenchment and clerical patronage networks, remained stronger than available alternatives for decisive elites. The worked analytical examples (Section 6.3) illustrate, on hand-computed trajectories, that the same shock produces radically different trajectories depending on LD configuration, consistent with H4's core prediction; this remains to be tested computationally through executed stochastic simulation.

7.2. Model-based expectations

The MLSARS architecture generates three testable predictions beyond the four hypotheses. First, regimes combining high CLI-cog, high MLI, and low EO should display strong hysteresis: even after significant ideological weakening, elite composition remains skewed toward loyalists because high MLI prevents safe defection. The trajectory tables in Section 6.3 illustrate this quantitatively.

Second, the model predicts a threshold effect in collapse dynamics. Once coercive specialist loyalty drops below θ_CC, high-MLI systems may appear stable briefly but then fragment rapidly due to mutual visibility of defection. This nonlinear transition explains why authoritarian collapses often appear sudden to external observers despite prolonged internal deterioration. The Soviet-type trajectory in Example 2 demonstrates this pattern: the system appears stressed but stable until approximately $t = 30$, then fragments rapidly between $t = 30$ and $t = 40$.

Third, the Iran template (Example 3) reveals lock-in with hysteresis: a regime absorbs repeated shocks while accumulating institutional damage. Each recovery cycle is weaker than the previous one. Policy error ratchets upward. Citizen dissent baselines rise. Yet the system does not cross the collapse threshold because MLI and CC, while weakened, remain above critical values. This suggests that the persistence-collapse relationship is not binary but continuous: regimes occupy a spectrum from robust persistence through degraded persistence to fragile persistence before reaching the collapse boundary. The IHR provides the metric for tracking this degradation computationally.

Fourth, China-type configurations (high CLI-cog, high MLI, medium EO) should outperform North Korea-type configurations (high CLI-cog, high MLI, low EO) on adaptive metrics while maintaining comparable persistence. This prediction is testable through comparative parameter sweeps holding CC, MLI, and CLI-cog constant while varying EO. If confirmed, it suggests that the relationship between institutional lock-in and institutional performance is mediated by informational architecture, a finding directly relevant to computational legal theory's interest in institutional evolvability.

8. Discussion

The computational framework presented here makes three contributions to the emerging field of computational legal theory.

First, it demonstrates that the formal tools developed for measuring constitutional entrenchment generalize to a broader class of institutional persistence problems. The extension from CLI to LD is not merely analogical. It is structurally identical: both measure the aggregate cost of departure from a current institutional equilibrium across multiple reinforcing dimensions. This suggests that computational law's methodological toolkit, including agent-based modeling, equilibrium analysis, network measurement, and trajectory computation, is applicable wherever nested institutional lock-in operates, from constitutional provisions to regulatory regimes to authoritarian political orders.

Second, the MLSARS architecture provides a formal bridge between computational norm analysis and comparative institutional research. Existing agent-based models of norm evolution typically operate in abstract strategy spaces without empirical calibration. MLSARS shows how historical cases can discipline computational models without compromising formalization: ordinal coding translates qualitative comparative knowledge into parameter templates, and worked analytical examples demonstrate the model's discriminating power without requiring full stochastic simulation. This methodology is transferable to any domain where computational legal theory meets comparative institutional analysis.

Third, the framework establishes a standard for falsifiability in computational institutional modeling. Each hypothesis has explicit falsification criteria. The worked examples include both confirming and disconfirming trajectories. The convergence and ablation analysis requirements (Section 6.2, 6.4) ensure that results are robust rather than artifacts. This addresses a persistent concern in computational social science: that agent-based models can be tuned to produce any desired outcome. By pre-specifying falsification conditions and publishing the code scaffold, the framework invites challenge rather than merely illustration.

The framework also raises a normative question directly relevant to computational law. If institutional lock-in can preserve dysfunctional arrangements indefinitely (as the high-LD trajectory in Example 1 suggests), then institutional evolvability, the capacity of a system to permit its own correction, may be as important a design criterion as stability. The Institutional Evolvability Index (IEI), developed in parallel work within the same research program, addresses precisely this question by measuring four components (variation capacity, selection effectiveness, correction mechanisms, and mutation tolerance) that together capture a system's capacity for non-catastrophic self-modification. The relationship between CLI, LD, and IEI forms a diagnostic triangle: CLI measures legal entrenchment, LD measures regime entrenchment, and IEI measures the system's capacity to escape both. Future work should formalize this triangle computationally.

The framework's relationship to existing computational law methodology deserves explicit articulation. The CLI measures legal entrenchment through five components: precedential depth, doctrinal integration, organizational investment, enforcement infrastructure, and procedural barriers. MLSARS extends this measurement logic to political regimes by substituting regime-specific components (cognitive conformity, material dependency, CC, emotional alignment, network lock-in) while preserving the computational structure: entrenchment as aggregate cost of departure from a current equilibrium. This is not analogy but generalization. The same mathematical framework that measures

why Argentina's labor law regime resists reform (CLI approximately 0.89) can, with domain-appropriate calibration, measure why North Korea's political regime resists transformation.

This generalizability has practical implications. Regulatory capture, corporate governance lock-in, and administrative path dependence exhibit the same structural features [33, 34]: nested selection pressures raising the cost of change above the cost of continuation. MLSARS provides a template for agent-based analysis of any such system. The five transition processes are sufficiently generic to be recalibrated for corporate boards, regulatory agencies, or international organizations with minimal architectural modification.

The connection to path dependence theory in law deserves emphasis. Hathaway [34] distinguishes increasing-returns, evolutionary, and sequencing forms of path dependence in the common law, arguing that stare decisis creates an explicitly path-dependent process in which early decisions constrain the range of later ones. Rubin [35] has since tested this framework's portability outside constitutional and commercial law, applying it to penal policy change and finding that path-dependence logic travels well to institutional domains where the constraining force is administrative and organizational rather than strictly doctrinal, precisely the extension the present paper makes to authoritarian regimes. Stachowiak-Kudła and Kudła [36] provide a further data point outside constitutional law, showing that administrative courts' rulings remain path-dependent on regional legal tradition a century after the institutional divergence that produced it. The CLI's precedential-depth component operationalizes exactly this mechanism for constitutional systems; LD's coercive-capacity and material-dependency components extend it to nonlegal institutional environments where the constraining force is not doctrine but patronage and repression. Read together, CLI and LD suggest that path dependence in the sense of Hathaway's work [34] is a special case of a more general entrenchment logic that operates whenever agents face an asymmetric cost of departing from an established equilibrium, whether that equilibrium is doctrinal, organizational, or coercive.

A further implication concerns institutional design. If LD can be measured and its trajectory predicted, then designers can identify configurations balancing stability against evolvability. The diagnostic triangle of CLI (legal entrenchment), LD (regime entrenchment), and IEI (evolvability) provides a framework for this design challenge. Systems with high CLI, high LD, and low IEI are zombie equilibria: stable, dysfunctional, resistant to self-correction. Systems with moderate CLI, moderate LD, and high IEI are institutionally healthy: stable enough to resist disruption, flexible enough to correct accumulated error. Making this distinction computationally precise connects constitutional law, comparative politics, and institutional design through shared formal methods.

Several limitations must be acknowledged. The comparative calibration relies on ordinal coding based on theoretical interpretation rather than independent archival measurement. The worked examples are analytical derivations from transition rules rather than outputs of full stochastic simulation with Monte Carlo sweeps. The five-case sample was selected for contrast rather than statistical representativeness. The model assumes that transition rules remain constant over the simulation period, which may not hold for cases where regime character changes endogenously (e.g., China's reform trajectory). Future work should implement full stochastic sweeps, expand the case set to include hybrid regimes and post-colonial authoritarian orders, and introduce endogenous rule modification to capture institutional evolution within the simulation period. The framework also depends substantially on the author's own prior work (the CLI) and on Khurshid's unpublished Tgmenks manuscripts for its theoretical grounding; independent replication and critique from scholars outside this immediate research program would strengthen confidence in the framework beyond what the added external citations in Sections 1, 2, and 8 can establish on their own.

9. Conclusion

This paper has presented MLSARS, an agent-based computational architecture for modeling institutional persistence under suppressive political orders. The framework integrates Khurshid's Tgmenks theory of human survival in the ENK with EPT's replicator-centered account of institutional reproduction. Four hypotheses were formulated, translated into computational experimental protocols with explicit falsification criteria, and calibrated against five historical cases.

The central computational finding is that normative exhaustion is insufficient for institutional collapse when material lock-in, CC, and cognitive conformity remain intact. The same exogenous shock produces persistence under high-LD configurations and collapse under medium-LD configurations, as demonstrated through two worked analytical examples with quantified trajectory predictions. Persistence and adaptation are computationally separable: a system can become more durable in the medium run while accumulating policy error that increases long-run brittleness.

The contribution is primarily methodological: extending CLI logic from constitutional to regime persistence, specifying a reproducible agent-based architecture with pre-specified falsification criteria, and demonstrating through worked examples that the integrated framework generates discriminating predictions. The path forward requires full stochastic implementation with Monte Carlo sweeps, expanded case calibration including out-of-sample predictions, and formal integration of the CLI-LD-IEI diagnostic triangle into a unified computational framework for measuring institutional entrenchment and evolvability.

Acknowledgment

The author is grateful to Professor Showan Khurshid (Salahaddin University-Erbil) for generously sharing his unpublished Tgmenks manuscripts and for extensive intellectual exchange that shaped the theoretical synthesis developed here.

Declaration on the Use of Generative AI

AI tools were used for research support, including locating sources, extracting information from literature, and reviewing drafts. The thesis, framework, arguments, analytical coding, computational specifications, and conclusions are the author's original contribution.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study. The comparative calibration in Section 5 reproduces publicly available secondary-source historical and political science literature.

Author Contribution

Ignacio Adrian Lerer: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Project administration.

References

[1] Lerer, I. A. (2025). Constitutional Lock-in and the Phenotypic Expression of Legal Regimes: Argentina's Labor Market as Irreversible Institutional Morphology. *SSRN Preprint: 5624710.*

[2] Lerer, I. A. (2026). Computational Detection of Constitutional Drift: Network Analysis and Semantic Measurement of Argentine Supreme Court Jurisprudence (1922–2025). *Journal of Computational Law and Legal Technology*, 1–13. https://doi.org/10.47852/bonviewJCLLT62027951

[3] Levitsky, S., & Way, L. (2022). *Revolution and dictatorship: The violent origins of durable authoritarianism*. USA: Princeton University Press.

[4] Geddes, B., Wright, J., & Frantz, E. (2018). *How dictatorships work: Power, personalization, and collapse*. UK: Cambridge University Press. https://doi.org/10.1017/9781316336182

[5] Lundquist, S., Teorell, J., Wahman, M., & Koliev, F. (2026). Updated data and new insight: A new version of WTH's authoritarian regime type dataset. *Democratization*, 1–17. https://doi.org/10.1080/13510347.2026.2631659

[6] Pepinsky, T. B. (2014). The institutional turn in comparative authoritarianism. *British Journal of Political Science*, *44*(3), 631–653. https://doi.org/10.1017/S0007123413000021

[7] Bueno de Mesquita, B., Smith, A., Siverson, R. M., & Morrow, J. D. (2003). *The logic of political survival*. USA: MIT Press.

[8] Bueno de Mesquita, B., & Smith, A. (2022). A new indicator of coalition size: Tests against standard regime-type indicators. *Social Science Quarterly*, *103*(2), 365–379. https://doi.org/10.1111/ssqu.13123

[9] Guriev, S., & Treisman, D. (2019). Informational autocrats. *Journal of Economic Perspectives*, *33*(4), 100–127. https://doi.org/10.1257/jep.33.4.100

[10] Sato, Y., & Wiebrecht, F. (2024). Disinformation and regime survival. *Political Research Quarterly*, *77*(3), 1010–1025. https://doi.org/10.1177/10659129241252811

[11] Epstein, J. M. (2002). Modeling civil violence: An agent-based computational approach. *Proceedings of the National Academy of Sciences*, *99*(*suppl_3*), 7243–7250. https://doi.org/10.1073/pnas.092080199

[12] Makowsky, M. D., & Rubin, J. (2013). An agent-based model of centralized institutions, social network technology, and revolution. *PloS one*, *8*(11). https://doi.org/10.1371/journal.pone.0080380

[13] Dornschneider-Elkink, S., & Edmonds, B. (2024). Does nonviolent repression have stronger dampening effects than state violence? Insight from an emotion-based model of non-violent dissent. *Government and Opposition*, *59*(1), 249–271. https://doi.org/10.1017/gov.2022.37

[14] Khurshid, S. (2025). *Tgmenks: A Theory of Human Life in the Ecological Niche of Knowledge*. Retrieved from: https://www.researchgate.net/publication/398529780_Tgmenks_A_Theory_of_Human_Life_in_the_Ecological_Niche_of_Knowledge

[15] Khurshid, S. (2026). *Human Psychology in the Ecological Niche of Knowledge: A Tgmenks Perspective*. Retrieved from: https://www.linkedin.com/pulse/human-psychology-ecological-niche-knowledge-showan-khurshid-hojmf/

[16] Dawkins, R. (1982). *The extended phenotype*. UK: Oxford University Press.

[17] Lerer, I. A. (2025). Law as Extended Phenotype: Toward an Evolutionary Theory of Legal Systems. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.18870552

[18] Khurshid, S. (2026). Why Suppressive Power Systems Fail: The Structural Logic of MSSP. *ResearchGate*. https://doi.org/10.13140/RG.2.2.17958.38726

[19] Osborne, D., Costello, T. H., Duckitt, J., & Sibley, C. G. (2023). The psychological causes and societal consequences of authoritarianism. *Nature Reviews Psychology*, *2*(4), 220–232. https://doi.org/10.1038/s44159-023-00161-4

[20] Abbasi, E. (2025). Political challenges and crisis management in the Islamic Republic of Iran: A comprehensive analysis. *Discover Global Society*, *3*(1), 93. https://doi.org/10.1007/s44282-025-00247-9

[21] Pavey, C., & Vyas, A. (2023). Extended epiphenotypes: Integrating epigenotypes into host behavioural manipulation by parasites. *Functional Ecology*, *37*, 831–837. https://doi.org/10.1111/1365-2435.14181

[22] Dennett, D. C. (1995). Darwin's dangerous idea. *The Sciences*, *35*(3), 34–40. https://doi.org/10.1002/j.2326-1951.1995.tb03633.x

[23] Smith, J. M. (1982). Evolution and the theory of games. In *Did Darwin Get It Right? Essays on Games, Sex and Evolution* (pp. 202–215). Springer US. https://doi.org/10.1007/978-1-4684-7862-4_22

[24] Weibull, J. W. (1995). *Evolutionary Game Theory*. USA: MIT Press.

[25] Itao, K., & Kaneko, K. (2025). Self-organized institutions in evolutionary dynamical-systems games. In *Proceedings of the National Academy of Sciences*, *122*(15), e2500960122. https://doi.org/10.1073/pnas.2500960122

[26] Lakatos, I. (1978). *The methodology of scientific research programmes*. UK: Cambridge University Press.

[27] Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., . . . , & Kietzmann, T. C. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, *24*(7), 431–450. https://doi.org/10.1038/s41583-023-00705-w

[28] Wilson, D. S., Madhavan, G., Gelfand, M. J., Hayes, S. C., Atkins, P. W. B., & Colwell, R. R. (2023). Multilevel cultural evolution: From new theory to practical applications. *Proceedings of the National Academy of Sciences*, *120*(16). https://doi.org/10.1073/pnas.2218222120

[29] Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, *67*(6), 380–400. https://doi.org/10.1037/h0040373

[30] Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. USA: University of Chicago Press.

[31] Shea, N. (2009). Imitation as an inheritance system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2429–2443. https://doi.org/10.1098/rstb.2009.0061

[32] Simonton, D. K. (2022). The blind-variation and selective-retention theory of creativity: Recent developments and current status of BVSR. *Creativity Research Journal*, *35*(3), 304–323. https://doi.org/10.1080/10400419.2022.2059919

[33] Gerschewski, J. (2021). Erosion or decay? Conceptualizing causes and mechanisms of democratic regression. *Democratization*, *28*(1), 43–62. https://doi.org/10.1080/13510347.2020.1826935

[34] Hathaway, O. A. (2001). Path dependence in the law: The course and pattern of legal change in a common law system. *Iowa Law Review*, *86*, 601–665.

[35] Rubin, A. T. (2023). The promises and pitfalls of path dependence frameworks for analyzing penal change. *Punishment & Society*, *25*(1), 264–284. https://doi.org/10.1177/14624745211043543

[36] Stachowiak-Kudła, M., & Kudła, J. (2022). Path dependence in administrative adjudication: The role played by legal tradition. *Constitutional Political Economy*, *33*(3), 301–325. https://doi.org/10.1007/s10602-021-09352-8

How to Cite: Lerer, I. A. (2026). An Agent-Based Framework for Simulating Institutional Lock-in Under Repressive Regimes. *Journal of Computational Law and Legal Technology*. https://doi.org/10.47852/bonviewJCLLT620210107