Civilizations as Hypotheses

Simulation, Repair, and the Architecture of Admissible Futures

Flyxion

Independent Researcher

June 2026

Abstract

A civilization is not a belief to be held. It is a hypothesis to be tested: a model with assumptions, predictions, failure modes, and repair requirements. Contemporary institutions do not treat it this way. They coordinate collective action through narrative—assertions about what policies will produce, what values are at stake, and which futures are available. This coordination mechanism has serious structural limits. Narratives do not expose internal assumptions to testing. They do not allow consequence tracing. They do not distinguish between disagreements about facts, disagreements about values, and disagreements about mechanisms. The result is that most civic deliberation operates in a regime where no argument can definitively fail.

We propose an alternative substrate: the *civic simulator*. A civic simulator is a persistent, multi-model computational environment in which competing theories of social, economic, political, and technological organization are instantiated as executable models, subjected to perturbation, and compared by their ability to maintain navigable futures. The simulator functions simultaneously as an archive, a laboratory, a teaching environment, a repair engine, and a governance substrate.

This essay describes the architecture of such a system: its physical, institutional, preference, and narrative layers; the formal stack connecting resource geometry to admissibility, repair, persistence, and collective choice; its model comparison and disagreement localization mechanism; its reachability objective; its repair operators; and the conditions under which it could function as civic infrastructure rather than entertainment. The goal is not a new simulation game. It is a new epistemological institution.

The Failure Mode of Narrative Coordination

Every institution performs one of three functions with respect to models of the world: storing models by preserving descriptions of how things work; comparing models by testing which descriptions predict better; and repairing models by revising descriptions when they fail. Universities mostly store models. Media mostly broadcasts them. Politics mostly negotiates between them. Science compares models. Engineering repairs them. The problem is not that any of these functions is wrong. The problem is that they are fragmented, and the fragmentation is producing a structural pathology.

Consider what happens when two people disagree about housing policy. One asserts that restricting density causes price increases. Another asserts that increasing density drives out existing residents. Both claims might be true in different regimes. Both might be false. But the disagreement is conducted at the level of assertion, which means no mechanism exists to localize where the disagreement actually lives. Is it about empirical facts regarding construction costs? About assumptions regarding migration behavior? About different values regarding who the policy is supposed to serve? About different beliefs regarding how quickly markets clear?

Narrative coordination cannot answer these questions because it has no machinery for distinguishing them. The same rhetorical form—a sentence asserting causal structure—can encode a factual claim, a mechanistic assumption, a normative priority, or a definitional dispute. Audiences cannot tell these apart. Speakers often cannot either.

This has a compounding effect. When disagreements cannot be localized, they cannot be resolved. When they cannot be resolved, institutions optimize for winning arguments rather than discovering which arguments are correct. The production of persuasive narrative displaces the production of testable understanding. The ecology of civic discourse selects for rhetoric over mechanism, because rhetoric is what works.

The limit of narrative coordination is therefore not a problem of misinformation, polarization, or bad-faith actors, though these exist. It is a structural problem: the substrate itself is the wrong tool for the epistemic task being asked of it.

The Simulator as Epistemological Institution

The proposal here is not to eliminate narrative. Narrative remains essential for motivation, meaning, and mobilization. The proposal is to add a missing layer beneath narrative: a layer where models are executable, consequences are traceable, and disagreements can be localized.

Definition 1 (Civic Simulator)**.** *A civic simulator is a persistent multi-model computational environment that instantiates competing theories of social organization as executable models; subjects those models to shared perturbations and intervention proposals; traces the consequences of each model across physical, institutional, preference, and narrative layers; identifies the specific assumptions responsible for divergent predictions; maintains a historical record of model performance over time; and provides repair operators that revise models when their predictions fail.*

This is not a forecasting tool. Forecasting requires confidence that one model is approximately correct. The civic simulator explicitly does not require this. Its function is not to predict the future but to reveal the structure of disagreements about it.

It is also not a game in the entertainment sense. Games reward players. The civic simulator rewards theories. The objective is not to win but to remain navigable—to maintain conditions under which further deliberation, adaptation, and action remain possible.

The closest analogs are wind tunnels, flight simulators, and clinical trial registries. A wind tunnel does not tell an engineer what shape the airplane should be. It tells the engineer what happens to each candidate shape under load. A flight simulator does not determine where to fly. It determines whether a pilot has the skills to get there. A trial registry does not determine which drug to approve. It determines which predictions were registered in advance and whether they turned out to be true.

The civic simulator occupies this role for theories of collective organization.

Before describing its architecture, it is worth making the central epistemological claim precise.

Definition 2 (Civilization Model)**.** *A civilization model is a tuple*

$$C = (A,\, R,\, P,\, N)$$

consisting of: a set of assumptions A about human behavior, incentives, and institutions; a rule system R governing state transitions; preference dynamics P describing how collective wants evolve; and normative commitments N encoding what the civilization treats as inviolable. Each civilization model induces a family of trajectories

$$\tau(C, s_0) = \{s_0, s_1, s_2, \ldots\}$$

from an initial state s_0 under the rules and dynamics encoded in C.

Definition 3 (Civilization Hypothesis)**.** *A civilization hypothesis is a predictive function*

$$H_C : (s_0, \iota) \mapsto p(\tau \mid C, s_0, \iota)$$

mapping an initial state s_0 and an intervention ι to a distribution *over predicted trajectories τ. Two civilization models differ precisely when they generate different trajectory distributions under equivalent initial conditions and interventions. The deterministic case $H_C : (s_0, \iota) \mapsto \tau$ is a degenerate limit; in practice, uncertainty about mechanism parameters, agent behavior, and external shocks means civilization hypotheses are inherently probabilistic objects.*

A civilization is therefore not a doctrine but a predictive object. A tax policy, a zoning regime, a constitutional framework, an educational system—each is a parameterized hypothesis about what trajectory distributions will result from its implementation. The civic simulator is the instrument for running those hypotheses under controlled conditions and comparing their outputs.

One further extension is required. Because the simulator's own outputs enter agents' reflective states R_i and reshape preference fields, the simulator Σ is not external to the civilization model. It is a component of it. A complete specification must therefore write:

$$C = (A,\ R,\ P,\ N,\ \Sigma)$$

where Σ encodes the informational effects of the simulator on agents operating within it. This makes the recursion explicit: Σ shapes P, which shapes Φ_C, which shapes the trajectories Σ is evaluating. A civic simulator that ignores this loop models a world that does not exist.

Architecture: Four Layers

A full civic simulator requires at least four coupled layers. The coupling is essential: most civic discourse fails precisely because it operates on one layer while ignoring the constraints imposed by others.

Layer 1: Physical Substrate

The lowest layer encodes the physical constraints within which any civilization must operate. These include energy production, storage, and transmission capacity; transportation infrastructure and throughput; material extraction and processing rates; construction and manufacturing capacity; communication bandwidth and latency; ecological regeneration rates and sinks; and maintenance costs and degradation rates. These quantities are not infinitely negotiable. They obey conservation laws, thermodynamic limits, and engineering constraints. A proposal that requires more energy than the grid can produce is not a bad political idea. It is a physically inadmissible trajectory.

The physical layer serves as the domain of hard constraints. Models that violate physical substrate limits are not wrong in a political sense. They are inadmissible in a mathematical sense: they describe trajectories that the system cannot execute.

This observation generalizes. The question of which constraints are truly binding—as opposed to contingent features of the current institutional settlement—is not a political judgment but a structural one. The simulator classifies constraints not as simply "required" or "optional" but by the minimum repair operator needed to remove them. A constraint c is paired with a repair cost $\rho(c)$: the resources, time, and coordination capacity required to relax it.

Layer 2: Institutional Structure

Above the physical layer sits the institutional layer: the formal and informal systems through which societies coordinate action. This encompasses regulatory frameworks and enforcement mechanisms, property regimes and contract enforcement, taxation and redistribution structures, educational and credentialing systems, legal frameworks and adjudication, governance architectures at multiple scales, and financial and monetary systems.

Institutional structures are slower to change than preferences and faster to change than physical infrastructure. They act as medium-term constraints on what interventions are implementable. A policy that is physically possible but institutionally inadmissible—requiring bureaucratic capacity that does not exist, enforcement mechanisms that have not been built, or coordination among institutions that do not currently communicate—occupies a different category of failure from one that is simply politically unpopular.

The institutional layer also encodes path dependence. Institutions are not blank slates. They have histories, stakeholder structures, embedded incentive gradients, and existing capabilities. Models that treat institutions as infinitely malleable are making a strong assumption that the simulator should expose and test.

Layer 3: Preference Fields

The third layer represents what populations actually want, tolerate, resist, imitate, or coordinate around. Preferences are not fixed. They respond to experience, social influence, material conditions, and available options. A central mistake of many social models is treating preferences as exogenous parameters rather than endogenous states.

The civic simulator treats preferences as a dynamic field:

$$P(x,t) : \mathcal{X} \times \mathbb{R}_{\geq 0} \to \Delta(\mathcal{A}) \tag{1}$$

where $\mathcal{X}$ is a population space, t is time, and $\Delta(\mathcal{A})$ is a probability distribution over available actions. A purely behaviorist treatment would allow preferences to evolve only through observed outcomes O_i, social influence N_i, and material conditions M_i. But preferences also evolve through reasoning: through the internal consideration of arguments, the imagining of consequences that have not yet occurred, and the revision of beliefs in response to counterfactual exploration.

The simulator introduces a reflective state variable $R_i(t)$ for each agent, capturing the agent's current model of how the world works and what interventions are available. Preference evolution then takes the form:

$$\frac{dP_i}{dt} = f(P_i,\, O_i,\, N_i,\, M_i,\, R_i) \tag{2}$$

where the reflective term R_i allows preferences to shift in response to counterfactual reasoning rather than only in response to experienced events. This has an important recursive consequence: the simulator's own outputs—the trajectory comparisons, dis-

agreement localizations, and admissibility assessments it produces—feed into R_i and therefore into preference evolution. The civic simulator is not a neutral observation platform. It is a participant in the preference field it models. Making this recursion explicit rather than ignoring it is both more honest and more useful: it allows the simulator to track how its own outputs are reshaping the space of collective choice.

This layer is where models diverge most dramatically. A neoliberal model and a communitarian model may agree about physical constraints and differ primarily in their assumptions about how preferences form and aggregate. Making those assumptions explicit and executable allows the simulator to identify exactly where theoretical traditions part ways. [O The preference layer also enables the simulator to function as a *preference revelation* mechanism. Rather than asking people what they want in the abstract—the standard survey or voting mechanism—the simulator allows people to interact with trajectories and reveal preferences through their choices within consequence-bearing environments.

Layer 4: Narrative and Belief Systems

The top layer represents the models that actors inside the simulation hold about the simulation itself. Civilizations do not operate on raw data. They operate on interpretations, narratives, theories, and ideologies that filter which signals are attended to, which causal structures are inferred, and which futures are considered possible.

This layer feeds back into all lower layers. A widely held narrative that resource scarcity is permanent depresses investment in production capacity. A widely held narrative that institutions are captured by elites undermines institutional legitimacy and enforcement capacity. A widely held narrative that preferences cannot change forecloses adaptive response.

Including this layer makes the simulator recursive in an important sense. The beliefs actors hold about what is possible constitute part of what is possible. Admissibility is therefore not purely a physical or institutional quantity. It is partly a function of the models held by actors operating within the system.

The Theoretical Stack: Five Questions, Five Answers

The four-layer architecture and the formal derivation chain that follows are related but distinct. The four layers—physical, institutional, preference, narrative—are *im-*

plementation layers: they describe the ontological structure of the world the simulator must represent. The derivation chain—constraints, admissibility, repair, persistence, collective choice—is an *explanatory stack*: it describes the logical order in which each component depends on the ones beneath it. The layers tell the engineer what to build. The stack tells the theorist why each piece is necessary. A reader may notice that the two hierarchies are not identical: the preference layer appears in both, but resource geometry $(\Phi, \vec{v}, S)$ underlies the physical layer rather than being a layer itself, and narrative appears as an implementation layer without a direct analogue in the explanatory stack. These asymmetries are features. The implementation layers are organized by what can be modified. The explanatory stack is organized by what is logically prior.

The four-layer architecture raises an immediate question: why exactly those layers, in that order? The answer is not arbitrary. Each layer corresponds to a distinct failure mode that a civilization simulator must be able to diagnose, and each failure mode is answered by a different component of the underlying formal theory. Together they form a derivation chain in which each level provides the substrate for the one above it. A civilization simulator must answer five questions: which futures survive; how failed futures recover; how histories remain identifiable; how collective decisions are formed; and what constrains all of the above. These correspond, in order, to admissibility, repair, persistence, preference fields, and resource geometry. The following subsections derive each answer formally.

I. Admissibility Answers: Which Futures Survive?

Let $x \in X$ be a world state and let $T(x)$ denote the set of states reachable from x under the world's transition rules. Define the *admissible region* $\mathcal{A} \subseteq X$ as the set of states satisfying all required physical, institutional, and ecological constraints.

A future is viable if its continuation remains admissible. Define viability over horizon H:

$$V_H(x) = \frac{1}{H} \sum_{t=0}^{H} \mathbf{1}\big(T^t(x) \in \mathcal{A}\big) \tag{3}$$

A civilization survives when $V_H(x) \approx 1$. A civilization collapses when $V_H(x) \to 0$. Viability is therefore not a primitive quantity to be stipulated. It emerges from admissibility:

$$\boxed{\text{Viability} = \text{Persistence of Admissibility}} \tag{4}$$

This gives the simulator its objective function. Not wealth, not power, not population: the preservation of admissible continuation.

II. Layered Admissibility

To unpack this structure, admissibility must be decoupled from a simplistic binary property. Civic disagreements frequently conflate that which is physically impossible with that which is merely institutionally stalled or narratively resisted. We formalize this separation by defining three nested admissible regions within the world state space X:

$$\begin{aligned}
\mathcal{A}_P &= \{x \in X : \text{all physical constraints satisfied}\} \\
\mathcal{A}_I &= \{x \in \mathcal{A}_P : \text{all institutional constraints satisfied}\} \\
\mathcal{A}_N &= \{x \in \mathcal{A}_I : \text{all narrative constraints satisfied}\}
\end{aligned}$$

such that

$$\mathcal{A}_P \supseteq \mathcal{A}_I \supseteq \mathcal{A}_N \tag{5}$$

A candidate trajectory or future state is not simply admissible or inadmissible. It is admissible at a particular *repair depth*. States within $\mathcal{A}_P \setminus \mathcal{A}_I$ are physically executable but forbidden under the current institutional settlement, requiring structural or regulatory modifications to unlock. States within $\mathcal{A}_I \setminus \mathcal{A}_N$ are both physically possible and institutionally viable, but are foreclosed by prevailing public beliefs, cultural resistance, or ideological paradigms.

By stratifying admissibility, the simulator prevents institutional inertia or political friction from being misclassified as thermodynamic or engineering impossibilities, systematically auditing the true horizon of available human futures.

III. Repair Answers: How Do Failed Futures Recover?

A viable civilization must survive perturbations. Suppose a disturbance moves the system from x to $x + \delta$, potentially pushing it outside admissibility: $x + \delta \notin \mathcal{A}$.

A repair operator $R : X \to X$ satisfies $R(x + \delta) \in \mathcal{A}$. Define repair efficiency:

$$\eta_R = \frac{d(x + \delta,\, \partial\mathcal{A}) - d(R(x + \delta),\, \partial\mathcal{A})}{d(x + \delta,\, \partial\mathcal{A})} \tag{6}$$

where $\partial\mathcal{A}$ is the admissibility boundary. Here $\eta_R = 1$ denotes complete recovery and $\eta_R = 0$ denotes no repair.

Adaptation is therefore not equilibrium maintenance. It is successful reintegration:

$$\boxed{\text{Adaptation} = \text{Repeated Repair Under Perturbation}} \tag{7}$$

A civilization is adaptive precisely when it possesses operators that move damaged states back into admissible regions. The institutional layer of the simulator encodes the capacity for such operators to exist and function.

IV. Persistence Answers: How Do Histories Remain Identifiable?

Classical models assume historical identity. Persistence theory derives it. Let $x_0, x_1, \ldots, x_t$ be a history and define a reconstruction map ρ_t that recovers earlier states from later ones. Define recoverability:

$$P(t) = \Pr(\rho_t(x_t) = x_0) \tag{8}$$

Historical identity persists when $P(t) > 0$. A civilization remains itself not because its components are unchanged but because its history is reconstructable:

$$\boxed{\text{Historical Continuity} = \text{Recoverable Distinction Through Time}} \tag{9}$$

This justifies the simulator's requirement for event logs, provenance records, and historical state reconstruction. Without these, the simulator cannot maintain the identity of the civilization being studied across repair events and institutional transitions.

V. Preference Fields and Reflective Dynamics

Suppose agents occupy a state space X. Each agent i possesses a preference potential $\Phi_i(x)$, whose gradient $\nabla\Phi_i(x)$ defines the direction of desired movement. Collective

preference aggregates as:

$$\Phi_C(x) = \sum_i w_i \, \Phi_i(x) \tag{10}$$

The resulting social pressure field is $F(x) = \nabla \Phi_C(x)$. The society does not choose a future directly: it moves along preference gradients. Stable institutions correspond to attractors satisfying $\lambda_{\max}(\nabla^2 \Phi_C) < 0$.

Collective choice therefore emerges from field geometry rather than discrete aggregation:

$$\boxed{\text{Collective Choice} = \text{Motion Through Preference Landscapes}} \tag{11}$$

To capture how preferences evolve through deliberate internal processing rather than simple sociological exposure, the internal reflective operator $R_i(t)$ of an agent is defined as a function of counterfactual evaluation over the stratified admissible landscape:

$$R_i(t) = \mathbb{E}_{s \sim \mathcal{A}_P}[\mathcal{U}_i(s) \mid \neg\tau\,] \tag{12}$$

where $\mathcal{U}_i$ is the agent's internal utility metric and $\neg\tau$ denotes alternative, unrealized histories. The preference field deforms dynamically via:

$$\frac{dP_i}{dt} = f(P_i,\, O_i,\, N_i,\, M_i,\, R_i) \tag{13}$$

Because the simulator's explicit evaluations directly populate the arguments of $R_i(t)$, the simulator acts recursively as a reflective operator on preference fields rather than merely an observational environment. Agents do not only change their minds because things happen to them. They change their minds because the simulator makes unrealized possibilities legible.

VI. Resource Geometry Answers: What Constrains Everything Else?

The substrate beneath admissibility, repair, persistence, and preference is resource geometry. Let $\Phi(x)$ denote available capacity, $\vec{v}(x)$ denote transport and coordination flow, and $S(x)$ denote entropy or obligation complexity. The admissible region is then:

$$\mathcal{A} = \{x : \Phi > \Phi_{\min},\ \|\vec{v}\| > v_{\min},\ S < S_{\max}\} \tag{14}$$

Resource failures occur when $\Phi \to 0$. Coordination failures occur when $\|\vec{v}\| \to 0$.

Complexity failures occur when $S \to \infty$. The reachable future volume is:

$$\mathrm{Vol}_R(\mathcal{A}) = \int_{\mathcal{A}} w(x)\, dx \tag{15}$$

Civilizations expand when $\frac{d}{dt}\mathrm{Vol}_R(\mathcal{A}) > 0$ and contract when this derivative is negative. The triple $(\Phi, \vec{v}, S)$ is the resource geometry that underlies all other constraints:

$$\boxed{\text{Constraints} = (\Phi,\ \vec{v},\ S)} \tag{16}$$

The Complete Stack

The six answers compose into a single derivation chain:

$$(\Phi, \vec{v}, S) \ \longrightarrow\ \mathcal{A} \ \longrightarrow\ R \ \longrightarrow\ P \ \longrightarrow\ \Phi_C \tag{17}$$

In words:

$$\boxed{\text{Constraints} \ \to\ \text{Admissibility} \ \to\ \text{Repair} \ \to\ \text{Persistence} \ \to\ \text{Collective Choice}} \tag{18}$$

A civilization simulator built on this architecture does not ask whether a policy is good. It asks five questions: Does it remain admissible? Can it repair disturbances? Does it preserve historical continuity? How does it reshape preference fields? What happens to reachable future volume? Those five questions are sufficient to compare entire civilizations as dynamical systems rather than as competing ideologies.

The Model Comparison and Disagreement Localization Mechanism

The simulator's most distinctive feature is not any single world model. It is the machinery for comparing multiple world models simultaneously.

Let $\mathcal{M} = \{M_1, M_2, \ldots, M_n\}$ be a set of models, each encoding different assumptions about human behavior, incentive response, institutional dynamics, technological change, and resource constraints. All models receive identical observations. All models propose interventions. The simulator executes the interventions and observes

consequences.

Divergences between models are not treated as errors to be eliminated. They are the primary object of study. Where models agree, their shared predictions carry evidential weight. Where they disagree, the simulator asks: *what is the minimal set of assumption differences that generates this divergence?*

Definition 4 (Disagreement Localization)**.** *Given two models M_i and M_j that produce divergent predictions $O_i \neq O_j$ for a shared intervention ι and initial state s_0, the* disagreement localization problem *is to find the smallest subset $\Delta A \subset A_i \triangle A_j$ of differing assumptions such that fixing ΔA in one model causes its predictions to converge with the other.*

This reframes the purpose of civic deliberation. The goal is not to determine which model is correct. The goal is to repair the disagreement into a smaller set of testable distinctions.

However, not all disagreements decompose into independent assumptions. Many divergences are emergent: no single assumption difference generates the prediction gap, but a cluster of assumptions interacting in the model's dynamics does. A naive localization procedure that searches for the smallest individual assumption will fail in these cases and produce false confidence that a disagreement has been reduced to an empirical question when the underlying models are operating in genuinely different dynamical regimes.

Definition 5 (Interaction Localization)**.** *When no individual assumption difference accounts for a divergence $O_i \neq O_j$, the simulator searches for the smallest* interacting subset *of differing assumptions:*

$$\Delta^\star = \arg \min_{\Delta \subseteq A_i \triangle A_j} (M_i \restriction_\Delta \approx M_j \restriction_\Delta)$$

where $M_k \restriction_\Delta$ denotes model M_k with the assumptions in Δ held fixed. The set $\Delta^\star$ is the minimal connected subgraph of assumption differences whose joint modification eliminates the divergence.

When interaction localization returns a large $\Delta^\star$, this is itself informative: the two models are not merely disagreeing about parameters but about the structural coupling of mechanisms. That is a deeper form of disagreement, and the simulator flags it as such rather than forcing a spurious reduction.

To take a concrete case: consider two models that disagree about whether nuclear

energy deployment can close a carbon gap by 2045. One predicts success; the other predicts failure. Disagreement localization might reveal that both models agree on reactor output per unit and on current grid capacity, but that one assumes regulatory approval in three years while the other assumes twelve, and that one draws construction cost learning curves from South Korean deployment history while the other draws them from recent European projects. The disagreement is now much smaller than "nuclear works" versus "nuclear doesn't work." It is localized to two empirical assumptions—regulatory timelines and cost learning curves—that can in principle be investigated, updated, and tested against historical data. This is the simulator's *didactic repair operator*: it transforms binary ideology conflicts into decomposed empirical questions.

The Admissibility Objective

Most civilization simulators optimize the wrong thing. They maximize wealth, population, military power, scientific output, or composite happiness indices. These objectives invite pathological strategies: strip-mine the future to win the present, convert all social capital into measurable metrics, optimize every system for the metric rather than the underlying reality.

The civic simulator proposes a different objective function: *admissibility maintenance*. As established in the preceding section, a state is admissible when it lies within a trajectory from which recovery remains possible, and reachability volume $\mathrm{Vol}_R(\mathcal{A})$ measures the space of futures accessible from the current state under repair-bounded intervention. The simulator's objective is to maintain and expand this volume:

$$\mathrm{Vol}_R(\mathcal{A}(s_t)) > 0 \quad \text{and ideally} \quad \frac{d}{dt}\mathrm{Vol}_R(\mathcal{A}(s_t)) \geq 0 \tag{19}$$

This objective captures something that GDP, happiness indices, and military power do not: the distinction between a civilization that is doing well now and one that has preserved its ability to adapt, recover, and choose. A civilization that has maximized present consumption at the cost of ecological regeneration capacity, institutional flexibility, and knowledge infrastructure has high present-state metrics and low reachability volume. The civic simulator treats this as a failure, not a success.

Reachability volume is not directly computable in the full simulation. But it can be

approximated by tracking the diversity of viable futures remaining after an intervention, the reversibility of interventions that have been taken, the repair capacity available for addressing failures that emerge, and the knowledge infrastructure required to navigate the available option space. A good intervention, under this objective, is one that maintains or increases the number and quality of futures the civilization can still navigate toward. A bad intervention is one that forecloses options, even when it improves present metrics.

Repair Operators

The civic simulator is not a passive observation environment. It includes explicit repair operators: mechanisms for identifying failures in models, institutions, and infrastructure, and for initiating recovery processes.

Definition 6 (Repair Operator)**.** *A repair operator $\mathcal{R}$ is a function that takes a failing state s, a diagnosis of the failure mode d, and a set of available resources ρ, and returns a set of intervention sequences $\mathcal{I}$ that are expected to restore admissibility:*

$$\mathcal{R}(s, d, \rho) = \mathcal{I} \subseteq \mathit{Interventions} \tag{20}$$

Repair operators operate at multiple scales. Infrastructure repair addresses physical layer failures—maintaining roads, restoring grid capacity, rebuilding ecological sinks—and these repairs are resource-intensive and time-bound. Institutional repair addresses coordination failures: updating regulatory frameworks that have become misaligned with actual conditions, rebuilding enforcement capacity, revising coordination mechanisms that have become brittle.

Model repair is the most distinctive feature of the civic simulator. When a model produces systematically wrong predictions, the simulator does not merely flag the error. It initiates a repair process: identifying which assumptions generated the error, proposing revised assumptions consistent with the evidence, and updating the model's predictions going forward. Model repair is iterative. A repaired model is a better model, not a correct one. The simulator maintains a provenance record of how each model has been revised, under what evidence, and by whom. This makes model evolution a public, traceable, auditable process rather than an invisible shift in expert consensus.

Preference repair addresses situations where collective preferences have been systematically distorted—by information asymmetry, by manipulation, by material deprivation that narrows the menu of tolerable options. Preference repair operators provide experiences and information that expand the considered option space.

The Multi-Agent Configuration

The simulator becomes most powerful when populated with agents that embody different theoretical frameworks. Each agent holds a theory about how the world works, proposes interventions consistent with that theory, and updates its theory when interventions fail. Each agent class can be characterized by an objective function family $J_i : X \to \mathbb{R}$ that determines which states the agent regards as improvements.

An *optimization agent* has $J_i(x) = u(x)$ for some utility function u, accepting any trade-off that increases u regardless of variance or reversibility. A *stability agent* has $J_i(x) = -\text{Var}(x)$, penalizing deviation from a reference trajectory even when expected utility is positive. An *exploration agent* has $J_i(x) = \text{Vol}_R(\mathcal{A}(x))$, treating expansion of reachable option space as the primary objective. A *conservation agent* has $J_i(x) = -\mathbf{1}[x \notin \mathcal{A}_P]$, imposing hard constraints on physical throughput and ecological regeneration as prior to any optimization. An *innovation agent* has $J_i(x) = \mathbb{E}[\text{Vol}_R(\mathcal{A}(x+\delta_{\text{tech}}))]$, deferring near-term resource constraints in expectation of future technological substitution. A *repair agent* has $J_i(x) = \eta_R(x)$, prioritizing restoration of admissibility over any other metric when the system is in a damaged state.

Contemporary political traditions instantiate different mixtures of these objective function families rather than corresponding cleanly to any one. The abstract classification allows the simulator to identify which underlying dispositions are generating disagreement rather than inheriting the accidental boundaries of current ideological labels.

Each agent receives identical observations. Each proposes interventions consistent with its framework. The simulator executes them in separate branches, or aggregates them through a governance mechanism, and observes the resulting trajectories. The question the multi-agent configuration answers is not which ideology wins, but which explanatory framework generates futures that remain navigable under perturbation. This is a much harder question than the usual ideological competition, because it asks theories to demonstrate not just that their preferred interventions produce good

outcomes in ideal conditions, but that the systems they build maintain repair capacity when conditions depart from the ideal. A theory that produces impressive outcomes in its preferred scenario but collapses the option space when subject to an unexpected shock is not a good theory. The simulator makes this failure mode visible rather than hiding it behind a rhetorical victory.

Persistence and Historical Memory

A critical feature of the civic simulator is that it is *persistent*. It does not reset between sessions. Its history accumulates. Decisions made in one period constrain options in the next. Failures that occur leave residues that must be repaired rather than deleted.

This is not a technical choice. It is an epistemological requirement. The world we actually inhabit is historically persistent. Path dependence is real. The institutional structures we inherit were built under conditions that no longer obtain, by actors whose motivations we can only partially recover. A simulator that resets would model a world of consequence-free experimentation that does not exist.

Historical persistence enables several features that reset-based simulations cannot provide. Accountability requires that agents and models be held responsible for the consequences of predictions they made and interventions they proposed; a theory that looked attractive five simulated years ago and has since produced cascading failures is visibly traceable to its originators. Institutional learning requires that organizations within the simulation accumulate memory, refine procedures, and develop the tacit knowledge that distinguishes effective institutions from formally similar but operationally incompetent ones. Repair history requires that each repair performed leave a record, building over time a map of what kinds of failures occur, under what conditions, with what available repair pathways and at what cost—a form of civilizational knowledge that has no equivalent in current institutions. Preference evolution requires that preferences visible in one period be comparable with preferences visible in later periods, after those preferences have been subjected to experience, allowing the simulator to distinguish between preferences stable under reflection and those that shift dramatically once their consequences become apparent.

A subtler problem concerns initial conditions. Every persistent simulation begins somewhere, and where it begins constrains which futures are visible. A simulation

initialized with highly unequal resource distribution will produce different admissibility conclusions than one initialized with equitable distribution, even under identical models. There is no politically neutral starting point. The solution is not to find the correct initial conditions but to abandon singular initialization entirely. A civilization hypothesis should survive across an *ensemble* of starting states:

$$S_0 = \{s_1, s_2, \ldots, s_n\}$$

The relevant question becomes $\Pr(\text{admissible future} \mid C, S_0)$ across historically plausible initial distributions, not admissible future(C, s_0) for a single privileged starting point. A civilization model that is only viable from one particular distribution of initial conditions is a fragile hypothesis, regardless of how impressive its trajectory looks from that starting point.

Conditions for Civic Infrastructure

A civic simulator that is technically well-designed but institutionally isolated is not yet civic infrastructure. The transition from a simulation environment to an epistemological institution requires several additional conditions.

Accessibility requires that the simulator be usable by people who are not specialists in economics, political science, or computational modeling, through an interface that allows non-experts to propose interventions in natural terms while the simulator handles the translation to executable model parameters. Transparency requires that the assumptions embedded in each model be inspectable by any participant; a black-box model that produces outputs without auditable assumptions is not a tool for civic epistemology but a tool for technocratic authority. Multiplicity requires that the simulator maintain multiple models simultaneously, none of which is designated as the correct one; a single-model simulator is a propaganda tool while a multi-model simulator is an epistemic environment. Connection to consequences requires that the simulator be linked, however imperfectly, to real decision processes, since a simulator entirely disconnected from governance produces only entertainment and academic publication. Open contribution requires that any participant be able to propose new models, new agents, new interventions, and new evaluation criteria; the institutional knowledge encoded in the simulator must not be the exclusive property of any single research group, government agency, or corporation.

These conditions are demanding. They are not fully met by any existing institution. But they are not utopian. Wind tunnels meet analogous conditions for aeronautical engineering. Clinical trial registries meet analogous conditions for medical research. The question is whether civic epistemology is willing to invest in the equivalent infrastructure.

A final condition concerns the relationship between the simulator and governance. The simulator is neither a governor nor an oracle. It does not determine which policies are permissible. It does not replace democratic deliberation. The appropriate analogy is an *auditor*: an institution with standing to produce assessments that are public, traceable, and formally recorded, but not binding. Environmental impact review, fiscal scoring, and engineering certification operate on this model. Policy may proceed despite negative findings. But negative findings become visible public objects rather than disappearing into the private reasoning of decision-makers. The civic simulator should be understood as extending this audit function to the full trajectory space of civilization-scale proposals.

Scope and Limitations

The proposal advanced here is a research program and a conceptual architecture, not an implementation plan. Several serious limitations deserve explicit acknowledgment.

Computational feasibility. The civic simulator as described makes demands that exceed current technical capacity. Coupling models across physical, institutional, preference, and narrative layers—each with different dynamics, timescales, and mathematical structures—is an open research problem. Computing reachability volumes in high-dimensional state spaces is in general intractable; the approximations described in §8 are placeholders for techniques that do not yet exist at the required scale. The honest framing is that this essay identifies what a civic simulator would need to do, not that it demonstrates such a simulator can be built. The gap between the two is large and should not be minimized.

Preference aggregation. The collective preference field $\Phi_C(x) = \sum_i w_i \Phi_i(x)$ assumes linear aggregation of scalar potentials. This is tractable but faces well-known impossibility results in social choice theory [2, 14]. Arrow's theorem and Sen's liberal paradox establish that no aggregation procedure satisfying reasonable consistency conditions can always produce coherent collective preferences from diverse individual

ones. The preference field formalism does not dissolve these results; it defers them. A complete treatment would either accept a specific aggregation rule and defend its normative justification, or replace the scalar potential with a representation that handles non-comparability explicitly. The weights w_i are themselves a political object: their determination requires a decision procedure that this essay does not provide.

The capture problem. The essay draws an analogy between the civic simulator and environmental impact review. The analogy is instructive but cuts in both directions. Environmental impact review has frequently been captured by the industries it regulates, used to delay rather than inform, and converted from an epistemic institution into a procedural formality [1]. A civic simulator connected to real governance would face the same pressures. Model inclusion, evaluation criteria, and output interpretation are all sites where interested parties could exert influence. The open-contribution condition is necessary but not sufficient: access to modeling skills and computational resources is unequally distributed, and the marginal contributions of well-resourced actors will systematically exceed those of poorly-resourced ones. The simulator's governance mechanisms—who adjudicates disputes about model validity, who sets evaluation criteria, who controls access—are themselves political problems requiring political solutions that lie outside the scope of this essay.

The fact-value entanglement. The essay distinguishes disagreements about facts and mechanisms from disagreements about values, and proposes the simulator as a tool for decomposing the mixture. This distinction is real and useful, but it is not clean [11]. Value commitments are embedded in the choice of which physical constraints to model, which institutional structures to represent, which preference dynamics to track, and what counts as repair. The admissibility objective $\mathrm{Vol}_R(\mathcal{A})$ itself embodies a substantive normative commitment: that option preservation is preferable to option exploitation, and that future generations have a claim on the option space available to present ones. These are contestable intergenerational ethics claims [12], not technical defaults. The simulator externalizes value commitments that were previously hidden inside informal assumptions; it does not eliminate them.

Unknown unknowns. The simulator can only evaluate models that have been explicitly formulated. It cannot discover frameworks that no participant has thought to propose. This is not a defect of the simulator specifically—it applies to any formal evaluation system—but it means the simulator's outputs are bounded by the imaginative horizon of its contributors. Systematic exclusion of certain modeling traditions from contribution, whether by resource constraints, cultural barriers, or governance

capture, will therefore produce systematic blind spots in what the simulator can see.

None of these limitations is fatal to the program. Computational intractability can be addressed incrementally; reduced-scale implementations that test subcomponents are already feasible. Aggregation problems can be addressed by explicit normative argument rather than formal magic. Capture risks can be mitigated by governance design, though not eliminated. Fact-value entanglement can be made explicit and auditable rather than hidden. Unknown unknowns can be partially addressed by actively soliciting contributions from underrepresented modeling traditions.

The appropriate relationship between this essay and these limitations is not confidence that they have been solved, but commitment to treating them as the next generation of problems to address. The proposal is a direction. The engineering and political work of building toward it remains largely ahead.

What the Civic Simulator Is Not

Several important clarifications prevent the proposal from being assimilated to existing categories.

It is not a prediction market. Prediction markets aggregate existing beliefs into probability estimates [15]. The civic simulator generates consequences from explicit models and exposes the mechanistic assumptions underlying different predictions. It aims to repair theories, not merely price them.

It is not a policy recommendation engine. The simulator does not tell users what to do. It shows users what their models predict will happen if they do what they are considering. The normative choice remains with the deliberating community.

It is not a game. Games are designed to be won. The civic simulator is designed to be navigated. The objective is maintaining admissibility, not accumulating points. A civilization that scores highly by all internal metrics while foreclosing its future option space has lost in the only sense that matters.

It is not a replacement for democratic deliberation. The simulator is a substrate that deliberation can operate on [5]. It replaces the blank white room of narrative assertion with a consequence-bearing environment. The normative preferences, value weightings, and fundamental political choices remain irreducibly human.

It is not neutral. The choice of which physical constraints to include, which institutional structures to model, and which preference dynamics to represent are themselves theoretical commitments. The simulator does not eliminate the sociology of knowledge [8]. It makes it explicit and auditable rather than hiding it inside methodological assumptions that ordinary participants cannot see.

Conclusion: The Epistemology We Need

The deepest problem facing contemporary civilization is not a shortage of values, vision, or political will. It is a shortage of the right kind of knowledge: knowledge about what interventions will actually produce, under what conditions, with what failure modes and repair requirements.

Narrative coordination produces conviction without mechanism. Markets produce coordination without deliberation. Expert systems produce conclusions without auditability. Democratic voting produces legitimacy without consequence tracing.

None of these institutions, individually or in combination, is producing what is needed: a shared environment in which competing models of collective organization can be instantiated, tested, compared, repaired, and improved by any participant with something to contribute.

The civic simulator is not a technology for resolving political disagreement. Political disagreement is legitimate and unlikely to disappear. It is a technology for separating the disagreements that are about facts and mechanisms—and that could in principle be resolved by evidence—from the disagreements that are about values—and that must be resolved by democratic choice.

That separation is itself an enormous contribution. Most contemporary political conflict is a mixture of factual disputes, mechanistic disputes, and value disputes, all conducted in the same rhetorical register, all appearing equally irresolvable. The simulator provides the machinery for decomposing the mixture.

What remains after that decomposition is the political task in its proper form: not which model of the world is correct, but given what the models agree on, and given the genuine remaining uncertainties, and given the testable predictions where models diverge, what kind of future do we want, and how much of our current option space are we willing to spend to pursue it?

That is a question that belongs to deliberating communities, not to simulations. But it is a much better question than the ones we are currently asking.

The central question of civilization is not which future to choose. It is which futures remain reachable after the choices have been made. A civilization that cannot answer that question is navigating without a model. A civilization that can answer it has transformed itself from a collection of beliefs into a hypothesis capable of learning from its own consequences.

$$\boxed{\text{A civilization is an adaptive hypothesis about how futures are generated.}} \tag{21}$$

This essay is part of an ongoing program connecting admissibility theory, repair dynamics, and the architecture of persistent computational worlds. Earlier treatments of related themes appear in Simulation as Civic Infrastructure *and* The Elements of Computational Worlds.

References

---

[1] Andrews, R. N. L. (1976). Environmental policy and administrative change: Implementation of the National Environmental Policy Act. *Lexington Books*.

[2] Arrow, K. J. (1951). *Social Choice and Individual Values*. Yale University Press.

[3] Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

[4] Epstein, J. M., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press.

[5] Habermas, J. (1984). *The Theory of Communicative Action*. Beacon Press.

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

[7] Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

[8] Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Harvard University Press.

[9] Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. (1972). *The Limits to Growth*. Universe Books.

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

[11] Putnam, H. (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Harvard University Press.

[12] Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

[13] Schelling, T. C. (1978). *Micromotives and Macrobehavior*. W. W. Norton.

[14] Sen, A. K. (1970). The impossibility of a Paretian liberal. *Journal of Political Economy*, 78(1), 152–157.

[15] Surowiecki, J. (2004). *The Wisdom of Crowds*. Doubleday.

[16] Tesfatsion, L., & Judd, K. L. (Eds.). (2006). *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*. North-Holland.