GL-06 Metacrisis, the third attractor and the governance of existential risk

The Architecture of Existential Catastrophe: Contemporary Theoretical Frameworks

Contemporary existential risk scholarship proceeds from the observation that industrial civilization has developed asymmetrical physical and technological capabilities without achieving a commensurate maturity in collective governance, institutional antifragility, or moral philosophy[1]. The resulting condition is increasingly characterized not as an inventory of isolated crises—such as biospheric depletion, nuclear brinkmanship, synthetic bioweapons proliferation, and cognitive warfare—but as a systemic, self-reinforcing "metacrisis" generated by core structural dynamics of human organization[1].

In the systemic taxonomy developed by Daniel Schmachtenberger and disseminated through the Civilization Research Institute and *The Consilience Project*, civilizational endangerment is driven by four primary "generator functions"[1]:

1. *Rivalrous Game Dynamics*: Structural configurations where participants are trapped in non-cooperative games, multipolar traps, and competitive races to the bottom[1]. In such environments, individual rational optimization mandates the unilateral extraction of resources or development of destabilizing capabilities, because uncoordinated restraint leads to immediate competitive displacement or geopolitical annihilation[1].
2. *Exponential Technological Leverage*: The acceleration of energetic, material, and computational agency available per capita[1]. As technological agency scales exponentially, the barrier of entry for catastrophic destruction plummets, tipping the offense-defense balance toward systemic vulnerability where the physical capacity of small groups or individuals to cause irreparable harm far exceeds the capacity of defensive architectures to absorb or prevent it[1].
3. *Substrate Consumption and Externalization*: Economic architectures that price finished goods while treating the ecological and social foundations of life—soil biomes, oceanic stability, atmospheric cycles, attentional integrity, and institutional trust—as unpriced externalities[1]. Civilizations organized on linear throughput inevitably liquidate their biophysical substrate faster than natural feedback loops can regenerate it[1].
4. *Epistemic and Sensemaking Degradation*: The deliberate corruption of collective information ecology[1]. Under modern economic incentives, communication channels and information platforms optimize for attentional capture, consumer engagement, and tribal polarization rather than empirical fidelity, paralyzing society's ability to achieve shared perception, deliberate effectively, or execute coordinated policy[1].

Within this theoretical paradigm, civilization is pulled toward two historical attractors[2]. The First Attractor is *Chaotic Collapse*: the ungoverned runout of multipolar competition, resource depletion, and catastrophic technological proliferation leading to socioeconomic

fragmentation, biospheric breakdown, or extinction[2]. The Second Attractor is *Authoritarian Totalitarianism*: the consolidation of centralized, unipolar surveillance states that suppress local defection, enforce defection-proof social compliance, and eliminate catastrophic risk by subordinating human sovereignty to automated surveillance and kinetic enforcement[2].
To evade this civilizational bind, the *Game B* intellectual movement—initiated in the 2010s by Jordan Hall, Jim Rutt, Bret Weinstein, and Daniel Schmachtenberger—articulates the theoretical necessity of a *Third Attractor*[1]. The Third Attractor posits an open, anti-rivalrous civilizational operating system capable of managing exponential technology and biophysical planetary boundaries without collapsing into chaotic fragmentation or locking into an automated police state[1]. In this formulation, local defection is eliminated not through imperial violence or algorithmic subjugation, but through structural coupling: socio-technical architectures where an individual actor's well-being is positively coupled to the integrity of the collective and the biophysical commons[1].

Attractor State	Defining Mechanisms	Primary Failure Modes
First Attractor *(Chaotic Collapse)*	Multipolar traps, competitive deregulation, externalized extraction, information warfare, unconstrained technological dispersion[1].	Kinetic warfare, systemic biospheric collapse, technological release of bioweapons, total socioeconomic devolution[1].
Second Attractor *(Panoptic Leviathan)*	Monolithic global unipolarity, ubiquitous algorithmic surveillance, pre-emptive policing, elimination of privacy, centralized command[12].	Totalitarian ossification, single-point administrative capture by dark-triad elites, loss of evolutionary adaptiveness, irreversible tyranny[14].
Third Attractor *(Polycentric Coherence)*	Structural anti-rivalry, nested self-governance, transparent information commons, reverse dominance hierarchies, cultural coherence[1].	High coordination overhead, cognitive saturation, vulnerability to covert informal hierarchies, scaling friction past watershed thresholds[17].

The Bostromian Panopticon: The Vulnerable World Hypothesis

In direct contrast to decentralized alternatives, Nick Bostrom formalizes the structural rationale for centralized coercion in *The Vulnerable World Hypothesis* (2019, *Global Policy*)[19]. Bostrom models technological discovery as the sequential extraction of balls from a metaphorical "urn of creativity"[7]. While historical draws have yielded white balls (unambiguously beneficial

discoveries like vaccination) and gray balls (mixed technologies with manageable externalities like fossil fuels), Bostrom hypothesizes the existence of "black balls": technologies that, by default, devastate the civilization that extracts them[7]. Bostrom defines civilizational devastation as an event causing the death of at least 15% of the global population or a persistent reduction in global economic output of more than 50% lasting at least a decade[7].

Bostrom classifies civilizational vulnerabilities into four distinct structural typologies:

- *Type 0 (Surprising Strangelets)*: A technology carries a latent terminal systemic risk that cannot be discovered prior to its initial deployment (e.g., initial theoretical concerns that the 1945 Trinity nuclear test might ignite atmospheric nitrogen)[7].
- *Type 1 (Easy Nukes)*: An invention gives small groups or individuals the physical capability to cause mass civilizational destruction using widely available materials and modest technical skills (e.g., table-top synthetic biology platforms capable of engineering pandemic pathogens from commercial oligonucleotides)[7].
- *Type 2a (Safe First Strike)*: Capabilities that grant powerful actors decisive incentives to execute unprovoked, preemptive strikes, destabilizing retaliatory deterrence (e.g., offensive hypersonic or biological platforms that eliminate secure second-strike capacity)[7].
- *Type 2b (Worse Global Warming)*: Severe commons dilemmas where millions or billions of independent actors face rational incentives to execute minor, individually beneficial actions whose aggregate externality produces irreversible civilizational devastation[7].

Bostrom argues that if the urn of creativity contains Type 1 or Type 2 vulnerabilities, human civilization cannot survive within its "semi-anarchic default condition"[11]. This default state is characterized by three empirical facts: limited capacity for preventive policing, limited capacity for global governance, and a diverse distribution of actor motivations that includes an "apocalyptic residual"—a small minority driven by nihilism, ideological fanaticism, or psychopathology who will trigger destruction regardless of deterrence[11].

To stabilize a civilization facing Type 1 black balls, Bostrom argues that society must establish a "High-Tech Panopticon"[12]. This architecture mandates the deployment of real-time, ubiquitous surveillance: every individual is fitted with an audiovisual recording device ("freedom tag") that streams continuous telemetry to machine-intelligence algorithms calibrated to detect illicit activity[11]. When an algorithm identifies an anomaly, human review officers verify the threat and dispatch armed preventive policing units to preemptively intercept the offender[12]. For Type 2 vulnerabilities, Bostrom posits the absolute necessity of unipolar global governance: a centralized planetary authority possessing the coercive capacity to enforce defection-proof compliance across all nation-states[11].

The Precipice and the Institutional Long Reflection

Toby Ord, writing from Oxford's Future of Humanity Institute in *The Precipice: Existential Risk and the Future of Humanity* (2020), calculates humanity's total existential risk over the next century at approximately one in six[25]. Ord categorizes natural existential risks (asteroid impacts, supervolcanic eruptions, stellar explosions) as negligible—quantified at less than 1 in 10,000 per century—while anthropogenic hazards, predominantly unaligned Artificial General

Intelligence (quantified at 1 in 10) and engineered pandemics (quantified at 1 in 30), dominate the risk landscape[25].

Ord's institutional prescription requires achieving "existential security": a stable civilizational condition where anthropogenic hazards are permanently suppressed below tolerable thresholds[25]. Once physical survival is secured, Ord calls for humanity to enter the "Long Reflection": an open-ended epoch spanning centuries or millennia during which humanity intentionally halts planetary colonization and irrevocable technological commitments to deliberate upon ethics, governance, and value theory[25]. Ord acknowledges that navigating the Precipice requires international institutions with robust regulatory mandates, but leaves the operational machinery of both the transitional state and the Long Reflection procedurally underspecified[28].

Concurrently, empirical researchers at the Centre for the Study of Existential Risk (CSER) at Cambridge, including Luke Kemp and Catherine Rhodes, have documented the actual topography of global hazard management[30]. In *The Cartography of Global Catastrophic Governance* (2019/2020), Kemp and Rhodes demonstrate that existing planetary risk governance is deeply fragmented across divergent, non-binding multilateral regimes[27]. Analyzing treaties across nuclear arms control, biological weapons, and climate mitigation, they conclude that centralized top-down governance often introduces systemic fragility[32]. In *Democratising Risk* (2022) and *Goliath's Curse* (2025/2026), Kemp argues that extreme administrative centralism exacerbates catastrophic vulnerabilities by concentrating power within unaccountable managerial elites, disconnecting policy from localized ecological feedback, and creating single points of failure vulnerable to authoritarian co-optation[34].

The Philosophical and Political Critique: Epistemic Blindness, Authoritarianism, and Exploitation

The existential risk governance paradigm has provoked rigorous critiques from political philosophers, intellectual historians, and social theorists, who identify within its foundational premises a technocratic paternalism and vulnerability to despotic capture[14].

The Ideological Critique: The TESCREAL Bundle

Intellectual historian Émile P. Torres, alongside computer scientist Timnit Gebru, identifies existential risk and longtermist governance as the operational core of the "TESCREAL" bundle—an interconnected philosophical matrix composed of Transhumanism, Extropianism, Singularitarianism, Cosmism, Rationalism, Effective Altruism, and Longtermism[37].

Torres demonstrates that longtermism, rooted in Nick Bostrom's 2003 formulation of "astronomical waste," derives its moral imperatives from utilitarian calculations over speculative future populations[41]. Bostrom posits that the local virgo supercluster could support approximately 10^{54} biological or digital subjective human-life-years[41]. Consequently, any action that reduces existential risk by a fraction of a percent yields an expected value equal to billions of times the value of all existing lives[41].

Torres argues that this mathematical totalization creates an ethical justification for tyranny:

1. *Subordination of Living Populations*: When the theoretical welfare of 10^{54} future digital simulations is balanced against the immediate material deprivation of 8 billion living humans, the moral claims of the present are discounted toward zero[41].
2. *Authoritarian Inevitability*: If the prevention of an existential catastrophe carries infinite or near-infinite utility, any action—including pre-emptive wars, totalitarian censorship, the suspension of civil liberties, or the execution of dissidents suspected of illicit technological development—can be rationalized as a moral imperative[12].
3. *Class Blindness and Oligarchic Alignment*: Torres argues that the movement serves as an apologetic ideology for Silicon Valley tech oligarchs, reframing vast inequalities of capital and computational power not as predatory extraction, but as necessary accumulations of agency to steward humanity through its critical bottleneck[14].

The Epistemic and Structural Critique: Moral Math and Depoliticization

Philosopher Alice Crary delivers a foundational critique of the moral epistemology underlying effective altruism and existential risk governance[44]. Crary demonstrates that the reliance on quantifiable consequentialist metrics—such as Quality-Adjusted Life Years (QALYs) or expected-value risk reductions—imposes an abstract welfarist framework that excludes structural, relational, and virtue-ethical considerations[45].

Crary argues that the existential risk agenda produces two structural deformations:

- *Erasure of Structural Injustice*: By treating moral action as an optimization exercise across a surface of neutral choices, the framework abstracts away the historical dynamics of racial capitalism, colonial extraction, and corporate power that generate systemic vulnerability[44].
- *The Dirty Hands Problem*: Because longtermism evaluates moral action strictly via downstream utility, it permits agents to participate in predatory, extractive industries under the doctrine of "earning to give," presuming that tainted capital can be cleansed through strategic redistribution to existential risk foundations[14].

In *Stop the Robot Apocalypse* (2015), Amia Srinivasan deconstructs the political complacency of William MacAskill's *Doing Good Better*[43]. Srinivasan targets the "replacement argument," which suggests that an individual is morally permitted to enter exploitative industries (such as high-frequency trading or fossil-fuel extraction) because an unreflective replacement would otherwise capture the role without directing surplus capital toward high-impact charities[43]. Srinivasan demonstrates that this logic individualizes moral responsibility, dismantles collective solidarity, and obscures the necessity for collective political contestation[43]. Srinivasan highlights the absurdity of redirecting scarce philanthropic capital away from verifiable, present-day human deprivation toward speculative technocratic risks—such as rogue artificial intelligence—a priority skew that disproportionately reflects the ideological insularity and material interests of corporate tech elites[42].

Institutional Specification: Evaluating the Operational

Deficit

When evaluated through the lens of public administration, legal jurisprudence, and institutional design, contemporary existential risk frameworks reveal an operational specification deficit[8]. Neither the Third Attractor nor the Bostromian Panopticon provides actionable administrative blueprints for actual governance.

Framework	Concrete Administrative Specification	Unresolved Institutional Mechanism
Schmachtenberger / Game B	Minimal; relies on evocative concepts like "anti-rivalry", "collective coherence", and "humane sensemaking"[1].	Mechanics of property allocation, boundary enforcement, dispute adjudication, and armed defense against rivalrous external actors[3].
Bostrom (VWH / Panopticon)	Technical mandate: continuous wearable audio-visual tags, algorithmic anomaly detection, pre-emptive policing interception[12].	Constitutional limits, separation of powers, rules of judicial evidence, and administrative safeguards against internal executive capture[14].
Ord (The Long Reflection)	Broad programmatic call for an intergenerational deliberative pause on technological deployment and space colonization[25].	Jurisdictional boundaries, voting procedures, criteria for ending the reflection, and enforcement across dissenting regional factions[28].
Ostrom (IAD / CPR Design)	Fully operationalized: 8 empirical design principles, graduated sanctions, boundary rules, nested enterprises[17].	Validated up to the watershed scale; exhibits institutional friction and coordination limits when applied to planetary commons[17].

The Game B Specification Deficit

The *Game B* paradigm operates primarily as a set of aspirational metaphors[6]. While it diagnoses the structural dynamics of multipolar traps with mathematical accuracy, it does not specify:

- The procedural structure of adjudication: How property disputes, boundary incursions, or kinetic violence are resolved when consensus cannot be attained[6].
- The enforcement mechanism: How a post-rivalrous community defends itself against an external, heavily armed rivalrous actor without adopting the command-and-control hierarchies of Game A[3].
- The transition protocol: How individuals embedded in industrial debt contracts, sovereign legal jurisdictions, and fossil-fuel infrastructures extract physical land and capital without being crushed by state apparatuses[10].

The Panoptic Enforcement Deficit

Bostrom's *Vulnerable World Hypothesis* outlines technical surveillance requirements while omitting the constitutional and administrative architecture necessary to prevent instantaneous capture[12]:

- Jurisdictional Recourse: The framework does not define an independent judiciary, appellate mechanisms, or rules of evidence to challenge false-positive classifications generated by machine-intelligence algorithms[13].
- Custody of the Custodians: Bostrom does not answer the fundamental public-choice dilemma: Who controls the root access keys to the surveillance data streams? How are intelligence directors prevented from weaponizing the apparatus to eliminate domestic political competitors[14]?

The Ostromian Alternative: Institutional Realism

The operational vacuum in existential risk literature contrasts sharply with the empirical institutionalism of Elinor Ostrom[49]. In *Governing the Commons* (1990) and subsequent synthesis with Thomas Dietz and Paul C. Stern (*The Struggle to Govern the Commons*, 2003, *Science*), Ostrom demonstrated that common-pool resources (CPRs) can be stably stewarded over centuries without centralized state command or private enclosure, provided eight core design principles are institutionalized[48]:

1. *Clearly Defined Boundaries*: Clear delineation of who has rights to harvest resources and the physical boundaries of the CPR[17].
2. *Congruence Between Appropriation/Provision and Local Conditions*: Rules governing resource use must match local ecological conditions and operational labor inputs[17].
3. *Collective-Choice Arrangements*: Individuals affected by operational rules must participate in their modification[17].
4. *Monitoring*: Monitors who actively audit resource conditions and appropriator behavior must be accountable to the users or be the users themselves[17].
5. *Graduated Sanctions*: Penalties for violations must start small and escalate progressively based on the severity and context of the infraction[17].
6. *Conflict-Resolution Mechanisms*: Rapid access to low-cost local arenas to resolve disputes between users and officials[17].
7. *Minimal Recognition of Rights to Organize*: External governmental authorities must recognize the statutory autonomy of the community's self-governing institutions[17].
8. *Nested Enterprises*: For CPRs embedded in larger social-ecological networks,

appropriation, provision, monitoring, enforcement, and conflict resolution must be organized into multiple tiers of nested operational units[17].

In a systematic meta-analysis of 91 empirical case studies, Michael Cox, Gwen Arnold, and Sergio Villamayor-Tomas (2010) evaluated the empirical validity of Ostrom's framework, confirming that the presence of these design principles correlates with sustained institutional robustness and ecological stability[17]. The crucial insight of Ostrom's work is that human coordination is secured not through monolithic Leviathans or unstructured moral goodwill, but through concrete, nested, polycentric institutional arrangements that balance power and enforce transparent, escalating consequences for defection[17].

The Capture Problem: Dark-Triad Dynamics in Complex Governance

A critical failure mode across both centralized and decentralized governance models is their vulnerability to subversion by individuals exhibiting Dark Triad personality traits: narcissism (grandiose entitlement, lack of empathy), Machiavellianism (cynical opportunism, deceptive strategic manipulation), and psychopathy (callousness, impulsivity, complete absence of remorse)[15].

In centralized hierarchies, capture operates along an *executive infiltration vector*. A single institutional apex—such as the command node of a unipolar world authority or the intelligence clearinghouse of a Bostromian panopticon—exerts an intense evolutionary selection pressure that attracts and promotes individuals with subclinical and clinical psychopathic profiles[12]. In an empirical study of corporate management personnel, Paul Babiak, Craig S. Neumann, and Robert D. Hare (*Corporate Psychopathy: Talking the Walk*, 2010, *Behavioral Sciences & the Law*) examined a sample of $N = 203$ corporate professionals selected for management development programs across multiple industrial firms[15]. Assessing psychopathic traits using the Psychopathy Checklist-Revised (PCL-R) and Psychopathy Checklist: Screening Version (PCL:SV), the researchers documented a psychopathy prevalence rate of approximately 3.9% among corporate managers[15]—roughly four times higher than the baseline rate of approximately 1% observed in the general community[53].

Babiak, Neumann, and Hare demonstrated that psychopathy scores were positively correlated with in-house ratings of charisma, presentation style, and strategic vision, but negatively correlated with actual managerial competence, team-building capacity, and basic performance responsibility[15]. Corporate psychopaths leveraged superficial charm to build upward patronage while actively exploiting subordinates—an organizational dynamic termed "kissing up and kicking down"[51]. When applied to existential risk governance, this dynamic reveals the fatal vulnerability of the Second Attractor: whoever captures the apex executive keys acquires permanent, unassailable control over the planetary surveillance and enforcement machinery, converting the defensive panopticon into an absolute, irreversible tyranny[14].

Conversely, decentralized and structureless governance models are subject to an *informal hegemony vector*. As Jo Freeman established in *The Tyranny of Structurelessness* (1972), the deliberate elimination of formal rules, codified offices, and constitutional voting procedures

does not abolish power; it merely renders power invisible, unaccountable, and informal. In uncodified, consensus-driven assemblies—such as radical activist networks or early-stage Game B hubs—power concentrates within charismatic cliques and manipulative friendship networks[3]. Narcissistic and Machiavellian actors exploit this ambiguity through performative moral purity, weaponized empathy, and epistemic gaslighting, marginalizing dissenting voices without ever facing the institutional accountability of formal impeachment[8].

To understand how human groups successfully resisted dark-triad capture across evolutionary deep time, anthropologist Christopher Boehm conducted comparative ethnographies of 48 extant egalitarian and mobile hunter-gatherer societies in *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (1999)[16]. Boehm demonstrated that ancestral egalitarianism was neither an accidental baseline nor a product of natural peacefulness; it was an intentional, continuously maintained political equilibrium that Boehm terms a *reverse dominance hierarchy*[16].

In reverse dominance systems, the rank-and-file majority forms a disciplined, cross-lineage coalition that vigilantly monitors potential "alphas"—individuals who exhibit arrogance, entitlement, physical bullying, or despotic ambition[57]. Boehm documents a graduated, four-stage escalation of leveling mechanisms deployed against predatory actors[16]:

1. *Criticism and Ridicule*: Deflating narcissistic aggrandizement through public humor, satire, and collective dismissal[16].
2. *Disobedience and Shunning*: Passive and active refusal to comply with attempted commands; deliberate collective silence and emotional withdrawal[16].
3. *Formal Ostracism and Exile*: Physical expulsion of the unrepentant predator from the group's boundaries and protective perimeter[16].
4. *Coordinated Capital Punishment*: When a psychopath or serial dominator refuses to desist and poses an active physical threat, the community coalition reaches consensus to execute them[16]. To prevent endless cycles of blood revenge, the execution was traditionally carried out by the offender's own immediate patrilineal or matrilineal kinsmen, precluding inter-clan retaliation[16].

Egalitarian self-governance survives only when an institutional structure possesses explicit, collective mechanisms to identify, sanction, and neutralize predatory dominance before it can convert social influence into formal leverage[16].

Historical Archaeology of the "Third Attractor": Four Institutional Precedents

The central empirical question confronting modern existential risk theory is whether the "Third Attractor"—a governance structure that sustains high coordination and ecological equilibrium without collapsing into either chaos or an authoritarian panopticon—has ever existed in historical reality[2]. Historical, ethnographic, and institutional literature demonstrates that polycentric, anti-rivalrous architectures have functioned stably across deep time. However, their survival was never anchored in spiritual enlightenment or uncodified trust; it was sustained by rigorous institutional engineering, bounded memberships, and enforceable counter-dominance mechanisms[16].

The Balinese Subak Water Temple System

Along the volcanic terraces of Bali, Indonesia, irrigated wet-rice cultivation has been governed for over a millennium by an autonomous, polycentric commons known as the *subak*[59]. Detailed by anthropologist J. Stephen Lansing in *Priests and Programmers: Technologies of Power in the Engineered Landscape of Bali* (1991), the subak networks span entire watersheds across hundreds of square kilometers, coordinating thousands of farming households along rivers such as the Oos and the Petanu in the Gianyar regency[59].

Analytical Dimension	Institutional Manifestation
Historical Span and Physical Scope	Documented from the 9th to 11th century CE to the present; island of Bali; coordinates ~1,000 autonomous subaks across ~100,000 hectares of terraced paddies[59].
Emic Claim *(Native Rationalization)*	Ritual propitiation of Dewi Danu (goddess of Lake Batur) and water spirits to preserve cosmological harmony and maintain sacred water flows (*Agama Tirtha*)[59].
Etic Historical Record *(Empirical Operation)*	A decentralized, hierarchical confederation of self-governing irrigation cooperatives coordinating water turns and fallow schedules via temple assemblies[59].
Modern Scholarly Inference	A distributed analog computing architecture that dynamically solves a multi-variable optimization trade-off between river-flow water scarcity and agricultural pest diffusion[59].

The biophysical coordination dilemma in wet-rice cultivation is structurally intractable for either uncoordinated markets or centralized bureaucracies[59]. If upstream farmers maximize their water intake, downstream paddies suffer severe drought[59]. Conversely, if farmers plant asynchronously to minimize water demand, agricultural pests (specifically the brown planthopper, *Nilaparvata lugens*) continuously migrate between paddies, exploding into devastating ecological infestations[59]. If all farmers across the watershed plant synchronously to starve out pests via a simultaneous fallow period, the river flow is instantly exhausted, inducing catastrophic downstream crop failure[59].

Rather than relying on a centralized bureaucratic irrigation ministry (a hydraulic despotism) or a privatized water market, Balinese farmers engineered a decentralized, hierarchical network of regional water temples (*pura tirtha*)[59]. The hierarchy links individual field shrines (*carik*),

canal-diversion shrines (*temuku*), subak-level temples (*pura ulun swi*), regional weir temples (*pura masceti*), and the apex crater-lake temple (*Pura Ulun Danu Batur*)[59]. At each temple node, elected farmer delegates meet on a shared, highly sophisticated calendrical system (the 210-day *pawukon* calendar paired with lunar tracking) to negotiate precise water-release dates and crop-rotation sequences[59]. Through simulations validated by Lansing and James Kremer (1993), this distributed network behaves as an analog computer: it converges naturally on optimal cluster sizes that maximize synchronized pest-control fallow while preventing hydraulic stress downstream[59].

The empirical test of this architecture occurred during the 1970s Green Revolution (*BIMAS* program)[59]. Dismissing temple rituals as primitive superstition, the Indonesian state mandated continuous planting of hybrid rice with synthetic fertilizers and pesticides[59]. The synchronized fallow broke down: pesticide-resistant planthopper populations exploded, river diversions fell into chaotic conflict, and crop yields collapsed across the island[59]. Yields recovered only when the state formally recognized the statutory authority of the water temples in the late 1980s, re-establishing polycentric coordination[59].

Scale Ladder Rung	**Operational Status and Structural Constraints: Balinese Subak**
Household *(4–10)*	Fully functional: Family compounds manage individual field plots and canal maintenance[59].
Hamlet / Band *(30–50)*	Fully functional: Local sub-units (*tempek*) supervise canal maintenance and water turns[59].
Village *(150–500)*	Optimal operational core: The single subak assembly coordinates 100–400 farmers democratically[59].
Parish / District *(1,500–5,000)*	High functionality: *Pura Masceti* regional councils federate multiple adjacent subaks[59].
Town / Watershed *(15,000–50,000)*	High functionality: Lake Batur regional temple assemblies synchronize river systems[59].
Region / Bioregion *(0.5–5M)*	Historically functional across Bali Island through shared sacred geography and

	pilgrimages[59].
Nation / Continent *(50–500M)*	Failed / Inapplicable: Vulnerable to state bureaucratization; easily suppressed by nation-states[59].
Planet *(8 Billion)*	Never tried: Culturally rooted in local animist-Hindu cosmology; cannot translate globally[59].

The subak resists dark-triad capture through biophysical feedback and institutional decentralization[59]. An upstream ruler or rogue farmer attempting to divert excessive water cannot isolate themselves from the biophysical consequences: downstream farmers deprived of water cannot maintain synchronized fallow, producing pest outbreaks that inevitably migrate upstream to ravage the predator's own fields[59]. Furthermore, temple priests hold no judicial command or standing militia; authority is vested strictly in the collective assembly of irrigators (*krama subak*), which can dismiss leaders who exhibit tyrannical traits[59].

However, anthropologists such as Stefan Helmreich emphasize that the subak system must not be romanticized: traditional Balinese society maintained rigid caste hierarchies (*triwangsa*), systematic gender inequalities that excluded women from subak assemblies, and historical records of violent skirmishes between rival subak networks during periods of catastrophic drought[61].

The Haudenosaunee Confederacy: The Great League of Peace

Formed between 1142 and 1450 CE in upstate New York, the Haudenosaunee (Iroquois) Confederacy originally united five distinct nations: the Mohawk, Oneida, Onondaga, Cayuga, and Seneca (joined by the Tuscarora in 1722)[65]. Documented by ethnohistorian Daniel K. Richter in *The Ordeal of the Longhouse: The Peoples of the Iroquois League in the Era of European Colonization* (1992) and anthropologist William N. Fenton in *The Great Law and the Longhouse* (1998), the League constituted a sophisticated, treaty-based federal republic that maintained peace across a massive bioregion for centuries[65].

Analytical Dimension	**Institutional Manifestation**
Historical Span and Physical Scope	Codified c. 1142 or 1450 CE to the present; upstate New York, Pennsylvania, Ontario, and Quebec; federated five (later six) sovereign nations comprising 20,000 to 30,000 individuals[65].
Emic Claim *(Native Rationalization)*	The *Kayanerehkowa* (Great Law of Peace): An ethical and cosmic mandate given by

	Deganawida (the Peacemaker) to replace cycles of blood revenge with condolence ceremonies, wampum exchanges, and collective peace[65].
Etic Historical Record *(Empirical Operation)*	A confederal representative council of 50 civil chiefs (*Hoyaneh*) functioning under bicameral consensus, subjected to matriarchal veto, with no standing army, executive head, or police[65].
Modern Scholarly Inference	A consociational confederal democracy specifically engineered to neutralize male dominance, eliminate despotic executive power, and prevent inter-clan escalation[65].

The constitutional architecture of the Great League of Peace embedded deliberate structural firewalls against authoritarian centralization[65]:

1. *Bicameral Deliberation and Fire-Keepers*: The Grand Council at Onondaga deliberated through structured bicameral chambers[65]. Matters were debated first by the "Older Brothers" (Mohawks and Senecas), then passed across the fire to the "Younger Brothers" (Oneidas and Cayugas)[65]. Only when these chambers achieved agreement was the resolution submitted to the Onondaga ("Keepers of the Fire") for final confirmation[65].
2. *Absolute Unanimity Requirement*: Confederal actions required complete consensus across all 50 sachems; a single dissenting nation or clan could veto collective mobilization or international treaties[65].
3. *Matriarchal Recall and Impeachment*: Civil titles were held by specific matrilineal clans[65]. Clan Mothers (*Otiyaner*) appointed the male chiefs and retained unconditional power to impeach and de-horn (*knocking off the horns*) any sachem who exhibited warmongering, financial corruption, or despotic ambition[68].
4. *Strict Separation of Civil and Military Authority*: Civil chiefs were strictly prohibited from leading military expeditions, and war leaders held no constitutional vote in the Grand Council[65].

Scale Ladder Rung	**Operational Status and Structural Constraints: Haudenosaunee**
Household *(4–10)*	Fully functional: Extended matrilineal longhouses governed internally by senior matriarchs[65].

Hamlet / Band *(30–50)*	Fully functional: Local clan segments shared common garden clearings and storage caches[65].
Village *(150–500)*	Optimal operational core: Village councils resolved day-to-day resource distribution[65].
Parish / District *(1,500–5,000)*	High functionality: Tribal councils (e.g., Mohawk council) governed national affairs[65].
Town / Watershed *(15,000–50,000)*	Confederal apex: The Grand Council at Onondaga federated roughly 25,000 people across watersheds[65].
Region / Bioregion *(0.5–5M)*	Moderately functional via external treaty networks (the Covenant Chain alliance system)[67].
Nation / Continent *(50–500M)*	Failed / Untested: Extreme unanimity requirement paralyzed decision-making against industrial empires[65].
Planet *(8 Billion)*	Never tried: Inapplicable; dependent on face-to-face matrilineal kinship ties and clan lineage rites[65].

The primary test of capture occurred at the founding of the League[65]. Tradition records that Tadodaho, a tyrannical Onondaga warlord and sorcerer, blocked all peace efforts through intimidation and terror[65]. Deganawida and Hiawatha defeated him not through superior violence, but by building an overwhelming coalition of all surrounding clans, "combing the snakes from his hair," and creating a constitutional office for him that integrated the Onondaga as permanent presiding officers of the Council fire, neutralizing his destructive agency within structural checks[65].

The ultimate failure of the League occurred during the American Revolutionary War (1777)[65]. Subjected to intense geopolitical pressure from the British Empire and the American revolutionaries, the confederacy's consensus broke down: the Mohawk, Seneca, and Cayuga allied with the British, while the Oneida and Tuscarora sided with the colonists[65]. Unanimity proved impossible, leading the chiefs to formally "cover the council fire" at Onondaga, which precipitated devastating civil war and the Sullivan Expedition of 1779[65].

Moreover, historians like Francis Jennings (*The Ambiguous Iroquois Empire*, 1984) and Daniel Richter emphasize that while the League maintained internal peace, it projected extreme

violence externally[65]: during the 17th-century "Mourning Wars," Haudenosaunee war bands decimated neighboring Iroquoian and Algonquian peoples (Hurons, Petuns, Neutrals, Eries) through systematic raiding, torture, and forced assimilation to replenish demographic losses from European epidemics[65].

The Swiss Alpine Pastures of Törbel

In the Canton of Valais, Switzerland, the high-altitude communal village of Törbel has maintained continuous self-governance of its communal pastures, forests, and irrigation networks (*bisses*) since the 15th century[70]. Evaluated by cultural anthropologist Robert McC. Netting in *Balancing on an Alp: Ecological Change and Continuity in a Swiss Mountain Community* (1981) and later integrated into Elinor Ostrom's core CPR studies, Törbel represents an empirically documented, long-lived commons that avoided both privatized enclosure and tragic depletion[70].

Analytical Dimension	Institutional Manifestation
Historical Span and Physical Scope	Formally codified in 1483 CE to the present; Valais, Switzerland; manages roughly 1,000 hectares of common alpine territory for a population between 200 and 600 burghers[70].
Emic Claim *(Native Rationalization)*	Preserving ancestral communal patrimony, defending valley autonomy against regional feudal lords, and maintaining Catholic piety and mutual village solidarity[70].
Etic Historical Record *(Empirical Operation)*	Direct democratic assembly of property-holding citizens enforcing strict statutory livestock caps based on private winter hay yield[70].
Modern Scholarly Inference	A dual-tenure socio-ecological system establishing long-term equilibrium by directly coupling common-pool summer forage rights to private wintering carrying capacity[70].

The structural survival of Törbel hinges on the *Überwinterungsregel* (wintering rule), codified in the communal statutes of February 1, 1483[72]: no citizen is permitted to send more cows to graze on the communal alp (*alp*) during the summer than they can adequately support through the winter with hay harvested from their own private meadows[70]. This statute elegantly binds open-access communal grazing rights to the carrying capacity of private hay meadows[70].

Because hay production on steep mountain slopes requires intensive, exhausting family labor, an individual cannot expand their livestock holdings without acquiring scarce, expensive private land, preventing speculative overgrazing[70].
Enforcement is administered through an annual assembly of burghers (*Bürgergemeinde*), which elects an alp committee, an alp warden (*Alpvogt*), and local herdsmen[72]. The warden inspects all herds entering the high meadows, counts livestock tallies against property declarations, monitors timber harvesting in communal protection forests (*Bannwälder*), and levies fines against infractions[72]. Fines are split: half goes into the village communal treasury, and half is granted to the warden, creating an incentive for diligent enforcement.

Scale Ladder Rung	**Operational Status and Structural Constraints: Törbel Alpine Commons**
Household *(4–10)*	Fully functional: Domestic stem households own private chalets, arable terraces, and hayfields[70].
Hamlet / Band *(30–50)*	Fully functional: Kin-based neighborhood clusters coordinate hay transport and canal maintenance[70].
Village *(150–500)*	Optimal operational core: The civic assembly of 200–600 burghers exercises direct democracy[70].
Parish / District *(1,500–5,000)*	High functionality: Federated into parish councils and valley associations (*Talschaften*)[70].
Town / Watershed *(15,000–50,000)*	Moderately functional: Valley confederations coordinate regional floodworks and transit roads[71].
Region / Bioregion *(0.5–5M)*	Cantonal federation: Nested historically into the Republic of the Seven Zenden and the Swiss Confederation[71].
Nation / Continent *(50–500M)*	Subordinated: Local alpine statutes integrated into and constrained by modern Swiss federal jurisprudence[71].

Planet *(8 Billion)*	Inapplicable: Parochial; depends entirely on geographic isolation and bounded property citizenship[70].

Törbel successfully resisted capture by outside feudal lords and wealthy valley merchants through strict citizenship and property alienation laws[70]. Common rights were severed from mere geographic residence: communal pasture rights could not be bought, sold, or leased to non-citizens, and could only be inherited through legitimate birth within a citizen lineage[70]. The village assembly held a mandatory first right of refusal on any private land sales, preventing the accumulation of private estates by external financial interests[70].

As environmental historian Jon Mathieu (*History of the Alps, 1500–1900*, 1998) demonstrates, Törbel's multi-century ecological equilibrium was sustained through severe social exclusion[71]. The village developed a stratified class order: resident non-citizens (*Hintersassen*) resided in the valley for generations without political voice, timber access, or grazing privileges[71]. Furthermore, to prevent land fragmentation, the community practiced strict impartible inheritance, resulting in structural demographic repression: high permanent celibacy rates, delayed female marriage (average age exceeding 28 years), and the systemic economic expulsion of non-inheriting sons into dangerous foreign mercenary service[70].

The Icelandic Commonwealth (930–1262 CE)

The Icelandic Commonwealth (*Þjóðveldið*) established an extensive, stateless, polycentric legal order spanning over three centuries[73]. Documented by medieval historian Jesse Byock in *Viking Age Iceland* (2001) and analyzed by institutional economists such as Thráinn Eggertsson (1990) and legal theorist William Ian Miller (*Bloodtaking and Peacemaking*, 1990), Iceland operated an island-wide judicial and legislative system with an absolute absence of a sovereign king, an executive branch, a centralized bureaucracy, or a public police force[73].

Analytical Dimension	Institutional Manifestation
Historical Span and Physical Scope	930 to 1262 CE; island of Iceland; coordinated a stateless population fluctuating between 40,000 and 70,000 people across a harsh subarctic bioregion[73].
Emic Claim *(Native Rationalization)*	"With law shall our land be built up, but by lawlessness laid waste" (*Með lögum skal land byggja*); defense of ancient Norse liberties against monarchical subjugation[73].
Etic Historical Record *(Empirical Operation)*	A private-property legal market centered on 39 (later 48) non-territorial chieftaincies (*goðorð*), regional courts, and an annual

	national assembly (*Althing*)[73].
Modern Scholarly Inference	A stateless, polycentric private-law order where judicial enforcement was privatized through tort and blood feuds, which eventually collapsed due to capital concentration and oligarchic capture[73].

The Icelandic constitution, codified at Thingvellir in 930 CE, instituted an innovative separation between governance authority and geographic territory[73]:

1. *Non-Territorial Chieftaincies (Goðorð)*: The legal structure rested on 39 (later 48) chieftaincies known as *goðorð*, held by leaders called *goðar*[73]. The *goðorð* was not a geographic fiefdom, but a bundle of personal legal jurisdictions and ceremonial roles that could be sold, divided, loaned, or inherited as private property[73].
2. *Freedom of Association and Legal Competition*: Free farmers (*bóndi*) were legally required to affiliate with a *goði*, but were completely free to transfer their allegiance to *any* chieftain within their quarter at the annual assembly without moving their residence[73]. If a chieftain attempted extortion, broke agreements, or failed to protect his clients, farmers transferred their fees and support to a rival *goði*, disciplining predatory leadership through dynamic market exit[73].
3. *Legislative Council and Quarter Courts*: The *Lögrétta* (the Law Council) served as the legislative core, composed of the *goðar*, their advisors, and the Lawspeaker (*Logsogumaður*), who recited one-third of the extensive unwritten legal code from memory each summer. Four Quarter Courts (*Fjórðungsdómar*) and a supreme Fifth Court (*Fimmtardómur*) adjudicated civil disputes and violent crimes through formal juror panels.
4. *Private Tort Enforcement and Outlawry*: The legal system recognized no crimes against the state; all homicides, thefts, and assaults were treated as private torts compensable by monetary restitution (*wergild*). If a defendant refused to submit to arbitration or pay damages, the courts declared them an outlaw (*skóggangsmaður*), revoking all legal protection and permitting any citizen to kill them with complete legal immunity. The responsibility for executing court judgments rested entirely with the plaintiff, who assembled kin coalitions to enforce outcomes.

Scale Ladder Rung	**Operational Status and Structural Constraints: Icelandic Commonwealth**
Household *(4–10)*	Fully functional: Patriarchal farmsteads (*bær*) operated as autonomous economic and defensive cells[73].

Hamlet / Band *(30–50)*	Fully functional: Neighborhood networks cooperated in seasonal sheep roundups (*réttir*)[73].
Village *(150–500)*	Fully functional: Local mutual-aid communes (*hreppar*) provided mandatory local famine insurance[73].
Parish / District *(1,500–5,000)*	High functionality: Spring Assemblies (*vorþing*) resolved local civil litigations[73].
Town / Watershed *(15,000–50,000)*	High functionality: Quarter assemblies governed regional appellate adjudication[73].
Region / Bioregion *(0.5–5M)*	Historical ceiling: Successfully federated an entire island population of ~50,000 under the *Althing*[73].
Nation / Continent *(50–500M)*	Failed / Untested: Stateless private enforcement requires face-to-face social reputation and cannot scale[73].
Planet *(8 Billion)*	Never tried: Polycentric tort enforcement collapses completely in anonymous, non-kin environments[73].

The Icelandic Commonwealth serves as an illustrative historical case study of how a stateless, polycentric system is captured and dismantled by dark-triad actors[73]:

- *The Capture Vector (Wealth Concentration and Tithe Extraction)*: In 1096, the *Althing* passed the Tithe Law, establishing a formal tax to support the Catholic Church. Crucially, secular *goðar* who built and maintained churches on their private estates retained control over half the collected tithe revenue, transforming chieftaincies into lucrative capital-accumulating assets. Over several generations, marriage alliances, land sales, and financial inheritances concentrated the 39 independent *goðorð* into the hands of just five wealthy families: the Sturlungs, Haukdælir, Oddaverjar, Ásbirningar, and Svínfellingar[73].
- *The Onset of Warlordism*: The market in governance collapsed: farmers could no longer practically defect from abusive chieftains, as an oligopoly commanded entire quarters. During the Sturlung Era (1220–1262), ruthless chieftains—exemplified by Snorri Sturluson and his nephew Sturla Sighvatsson—amassed private armies of retainers, engaged in terror campaigns, burned rival estates, and executed retaliatory massacres across the island (e.g., the Battle of Örlygsstaðir in 1238, where over 2,000 warriors clashed)[73].
- *Terminal Failure (External Subjugation)*: To secure decisive military advantage, rival warlords such as Gizur Thorvaldsson sworn vassalage to King Haakon IV of Norway,

inviting royal intervention. Exhausted by decades of lawless predation and asymmetric civil war, the farmers assembled at the *Althing* of 1262 and formally surrendered their sovereignty to the Norwegian Crown through the Old Covenant (*Gamli sáttmáli*)[73]. Stateless polycentricity collapsed directly into monarchical subjugation[73].

Legal historian William Ian Miller (*Bloodtaking and Peacemaking*, 1990) and medievalist Jón Viðar Sigurðsson (*Chieftains and Power in the Icelandic Commonwealth*, 1999) dismantle romanticized libertarian accounts of Iceland, showing that the legal system was structurally biased toward wealthy power-brokers: an impoverished smallholder with a legitimate murder claim against a chieftain had no physical capacity to enforce a judgment, and was forced to sell the claim at a severe discount to a rival warlord, turning justice into a speculative market for elite violence.

Comparative Scale Ledger and Structural Synthesis

Evaluating the performance of these six governance paradigms across the scale ladder reveals the fundamental trade-offs between centralized surveillance architectures and polycentric commons networks.

Scale Ladder Rung	Bostrom Panopticon	Game B / 3rd Attractor	Balinese Subak	Haudenosaunee	Törbel Commons	Icelandic Commonw.
1. Household *(4–10)*	Highly invasive; abolishes private domestic sphere.	Untested; models as intentional nuclear pod.	Functional: manages domestic plots and canals.	Functional: matrilineal stem longhouse core.	Functional: private chalets, meadows, and hearths.	Functional: patriarchal defensible farmstead (*bær*).
2. Hamlet / Band *(30–50)*	Redundant; prohibitive per-capita hardware cost.	Experimental; communal living pod iterations.	Functional: localized *tempek* canal divisions.	Functional: local clan matrilineage segment.	Functional: neighborhood hay-cutting clusters.	Functional: neighborhood sheep-roundup mutuals.
3. Village	Provokes intense	Experimental;	Optimal core:	Optimal core:	Optimal core:	Optimal core:

(150–500)	resistance and social alienation.	vulnerable to informal capture.	autonomous *subak* direct assembly.	village longhouse council fire.	*Bürgergemeinde* direct assembly.	*hreppur* mutual-aid commune.
4. Parish / District *(1,500–5,000)*	High administrative overhead; early bureaucratization.	Aspirational; lacks formal enforcement apparatus.	Functional: *pura masceti* regional water temples.	Functional: national tribal councils (e.g. Mohawk).	Functional: *Talschaft* regional valley councils.	Functional: regional Spring Assemblies (*vorþing*).
5. Town / Watershed *(15,000–50,000)*	Constant false-positive algorithmic policing shocks.	Fails: uncodified consensus breaks under anonymity.	Optimal apex: crater lake *Batur* temple council.	Confederal apex: Grand Council at Onondaga.	Functional: cantonal watershed alliances.	Functional: Quarter Assemblies and regional courts.
6. Region / Bioregion *(0.5–5 Million)*	Imposes police-state militarization to enforce rules.	Theoretical; lacks proven real-world model.	Island boundary: unified sacred water geography.	Boundary limit: Covenant Chain treaty alliances.	Nested: Seven Zenden and Swiss Confederation.	Historical ceiling: national *Althing* assembly.
7. Nation / Continent *(50–500 Million)*	Requires unipolar military sovereignty over states.	Purely theoretical; no administrative design.	Fails: subjugated by modern Indonesian state.	Fails: fractured and broken by European empires.	Subordinated: absorbed into Swiss federal law.	Fails: warlord oligopoly forces Norwegian rule.
8. Planet	Prescribed: exits	Aspirational;	Inapplicable: tied	Inapplicable:	Inapplicable: relies	Inapplicable:

(8 Billion)	semi-anarchic global condition.	posits planetary cultural shift.	to local volcanic geography.	requires face-to-face clan lineage.	on strict border exclusion.	private tort fails in mass society.

The comparative ledger demonstrates a profound institutional symmetry. Centralized surveillance mechanisms (the Second Attractor) become increasingly powerful as they scale toward planetary unipolarity, but they create single points of apex failure that render them exceptionally vulnerable to permanent capture by dark-triad elites[12]. Conversely, polycentric commons architectures (historical manifestations of the Third Attractor) demonstrate extraordinary ecological stability, antifragility, and resistance to tyranny across small and intermediate scales (from the household up to watersheds of 50,000 people)[17]. However, these systems exhibit severe scaling ceilings: they depend on bounded memberships, face-to-face reputation, and local ecological couplings, and historically fragment or collapse when exposed to external industrial states, trans-regional markets, or imperial warfare[65].

Ten-Line Summary of Robust Findings

1. Anthropogenic existential risks are driven by multipolar traps where unconstrained defection remains the dominant game-theoretic equilibrium[1].
2. Bostrom's panoptic surveillance model reliably suppresses Type 1 catastrophic vulnerabilities only by constructing an unassailable apex authority[7].
3. Apex administrative hierarchies systematically select for Machiavellian and psychopathic personalities at roughly four times baseline community rates[15].
4. Monolithic existential-risk governance creates an irreversible single point of failure vulnerable to total despotic capture[14].
5. The "Third Attractor" lacks operational administrative specification in contemporary literature, relying on rhetorical postulates rather than functional governance codes[6].
6. Ostrom's polycentric commons principles are empirically demonstrated to manage complex common-pool resource dilemmas stably across centuries[17].
7. Historical instances of the Third Attractor were sustained by strict, exclusionary boundaries and biophysical couplings rather than open-ended altruism[59].
8. Uncodified, consensus-driven assemblies inevitably generate informal, unaccountable hierarchies vulnerable to manipulative capture[3].
9. Pre-state egalitarianism survived exclusively through reverse dominance hierarchies: collective coalitions enforcing graduated sanctions up to capital execution[16].
10. Polycentric regimes scale through federation up to watersheds (50,000 people), but historically collapse under external state pressure or internal wealth concentration[65].

Contested or Open Questions

The intersection of existential risk literature and institutional history leaves several critical theoretical questions unresolved. The most pressing dilemma concerns the absolute scalability

of nested commons architectures: while Elinor Ostrom proved that common-pool resources can be managed across watersheds without sovereign intervention, it remains contested whether nested polycentricity can govern global commons dilemmas—such as atmospheric greenhouse gas accumulation or autonomous artificial intelligence arms races—without establishing a coercive global Leviathan[12].

A second open question concerns the irreversibility of technological information. Nick Bostrom's *Vulnerable World Hypothesis* presumes that once an easy destructive capability (a Type 1 black ball) is extracted from the urn of invention, it cannot be un-invented or suppressed in a decentralized world without ubiquitous surveillance[7]. Opponents argue that social coordination, taboo systems, and physical choke points on precision manufacturing can suppress catastrophic hazards without requiring a total surveillance panopticon, but historical examples of sustained voluntary technological renunciation (such as Tokugawa Japan's abandonment of firearms) occurred under unique conditions of geographic isolation and rigid feudal control.

A third major debate centers on the feasibility of Toby Ord's proposed "Long Reflection"[25]. Longtermist theorists presume that humanity can achieve a sustained period of global existential security and deliberately freeze technological development while ethical deliberation proceeds[28]. Critics from political realism point out that any prolonged pause would inevitably be destabilized by rogue factions seeking strategic advantage, while historical sociologists question whether an intergenerational council could maintain moral authority across centuries without devolving into a dogmatic, ossified priesthood.

Finally, researchers contest whether Dark Triad personality traits can be eliminated through institutional design or whether they represent an evolutionary stable strategy that will continuously re-emerge across every governance architecture[15]. While Christopher Boehm's reverse dominance model demonstrates that small-scale egalitarian coalitions can suppress bullies and narcissists through shunning and execution, scaling these counter-dominance mechanisms to large, anonymous populations without unleashing unchecked mob terror or denunciatory witch hunts remains an unsolved challenge of political philosophy[16].

Foundational Sources for Full Study

- **Bostrom, Nick (2019)**, *The Vulnerable World Hypothesis*, *Global Policy*, 10(4), pp. 455–476. The foundational paper establishing the urn model of technological development, categorizing Type 0, 1, 2a, and 2b vulnerabilities, and introducing the controversial argument for wearable freedom tags, machine-intelligence surveillance panopticons, and unipolar global governance to stabilize existential threats[11].
- **Ostrom, Elinor (1990)**, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press. The definitive work of modern political economy dismantling the inevitability of the tragedy of the commons, establishing the eight core design principles that enable long-term self-governance of shared resource systems without central state command[17].
- **Lansing, J. Stephen (1991)**, *Priests and Programmers: Technologies of Power in the Engineered Landscape of Bali*, Princeton University Press. An extraordinary

anthropological, ecological, and computational investigation demonstrating how Balinese water temple networks solve complex, non-linear dilemmas of water scarcity and pest control across entire watersheds through distributed ritual coordination[59].
- **Richter, Daniel K. (1992)**, *The Ordeal of the Longhouse: The Peoples of the Iroquois League in the Era of European Colonization*, University of North Carolina Press. The primary ethnohistorical study of the Haudenosaunee Confederacy, providing a detailed empirical analysis of the Great Law of Peace, consensual bicameral governance, clan-mother recall powers, and the political stresses of colonial warfare[65].
- **Boehm, Christopher (1999)**, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*, Harvard University Press. A seminal anthropological and evolutionary analysis based on ethnographies of 48 foraging societies, establishing that egalitarianism is sustained through intentional reverse dominance hierarchies where subordinate coalitions actively sanction, ostracize, and execute potential despots[16].
- **Netting, Robert McC. (1981)**, *Balancing on an Alp: Ecological Change and Continuity in a Swiss Mountain Community*, Cambridge University Press. The classic deep-time study of Törbel, documenting five centuries of alpine pasture management and demonstrating how coupling private wintering limits to communal summer rights prevents resource exhaustion[70].
- **Byock, Jesse (2001)**, *Viking Age Iceland*, Penguin Books. An exhaustive historical analysis of the medieval Icelandic Commonwealth, examining the operation of a stateless, polycentric legal order based on non-territorial chieftaincies and detailing its eventual collapse into oligarchic civil war and monarchical annexation[73].
- **Babiak, Paul, Neumann, Craig S., and Hare, Robert D. (2010)**, *Corporate Psychopathy: Talking the Walk*, *Behavioral Sciences and the Law*, 28(2), pp. 174–193. The foundational empirical study quantifying a 3.9% psychopathy prevalence among corporate managers, demonstrating how manipulative personalities exploit organizational hierarchies by presenting superficial charm while undermining collective performance[15].
- **Crary, Alice (2023)**, *The Toxic Ideology of Longtermism*, *Radical Philosophy*, 2.14, pp. 49–57. A philosophical deconstruction of the moral mathematics of existential risk and effective altruism, arguing that abstract consequentialism depoliticizes structural injustice, provides moral cover for exploitative capital, and introduces totalitarian ethical hazards[44].

Implementability after the Fall

Operating under post-collapse conditions—defined by the permanent severance of global industrial supply chains, the permanent failure of the electrical grid, the absence of telecommunications networks, and the evaporation of the state apparatus—existential risk governance ceases to be a theoretical abstraction[2]. It transforms into a physical imperative: preventing the community from descending into chaotic collapse (the First Attractor) while guarding against subjugation by predatory warlords or internal despots (the Second Attractor)[2]. Implementing a functional Third Attractor requires resurrecting historically proven polycentric mechanisms tailored to specific population thresholds[17].

Social Tier	Minimum Viable Architecture	What It Needs to Start	First Failure to Expect
Household *(4–10 people)*	Indivisible stem household; shared hearth, communal granary, and unified domestic defense pact[70].	Basic numeracy, physical seed ledger, solar/lunar frost marker, common dining hearth[70].	Domestic despotism by patriarch/matriarch; labor starvation during acute harvest windows.
Village *(300 people)*	Direct assembly of household heads; Ostromian commons charter with wintering limits and elected wardens[72].	Three literate scribes, municipal parchment cadastre, acoustic assembly bell, covered meeting hall[72].	Clandestine nocturnal water or livestock theft by a dominant kin faction; retaliatory feuds[16].
District *(3,000 people)*	Bicameral confederal council; dual-key militia, water-appellate juries, and matriarchal impeachment checks[65].	Trained legal scribes, treaty tally archive, shared market/fair calendar, neutral moot ground[65].	Upstream water diversion during drought; merchant/warlord capital concentration[59].

The Household Level (4 to 10 People)

At the household scale, governance must organize around an indivisible stem estate capable of maintaining seed security, caloric production, and ballistic defense[70]. Internal governance cannot rely on commercial market transactions; it operates as a kin-based moral economy characterized by pooled resources, shared labor, and direct allocation of responsibilities based on physical capacity.

- *Prerequisites to Start*:
  - *Functional Literacy and Numeracy*: Basic reading and accounting capability to maintain crop-yield logs, ration stores, and read ancestral farming texts.
  - *Physical Records*: A hand-inscribed family parchment ledger recording seed inventories, lineage descents, land boundary stone markers, and contractual debt cancellations.
  - *Shared Calendar*: An analog astronomical calendar—such as a stone sun-marker or lunar peg-board—to track frost cycles, planting dates, and drying periods.
  - *Meeting Place*: The domestic hearth or common kitchen table, serving as the daily site of

operational coordination, labor division, and conflict mediation.

- *The First Failure to Expect*: Domestic tyranny, physical abuse, or narcissistic domination by a single family head, paired with labor collapse when key adult laborers are incapacitated during narrow planting or harvest windows.

The Village Level (Approximately 300 People)

At the village scale, governance must institutionalize an Ostromian common-pool resource association to manage communal pastures, woodlots, and irrigation channels[17]. The assembly consists of one voting delegate from each established household, meeting monthly under a binding civic charter[72]. The village adopts strict wintering rules: no family may graze more livestock or cut more timber on common land than they can support with private hay meadows and labor[70]. Monitoring is executed by an elected, rotating warden (*Alpvogt*), empowered to issue graduated fines split between the village treasury and the warden's household[72]. To prevent alpha-capture, the assembly maintains an active reverse dominance hierarchy: any individual displaying arrogance, bullying, or hoarding is subjected to public mockery, social shunning, and, if unrepentant, physical exile[16].

- *Prerequisites to Start*:
  - *Functional Literacy*: A minimum of three non-lineage scribes capable of reading and drafting legal ordinances, maintaining assembly minutes, and calculating tax assessments.
  - *Physical Records*: A village cadastre on linen paper or parchment, containing precise boundary surveys, shared pasture rights, and water-turn schedules, secured in a physical strongbox requiring three independent keys to unlock.
  - *Shared Calendar*: A standardized civic and ecological calendar synchronized by a central tower bell, horn, or water-clock, signaling daily canal diversion windows and assembly dates.
  - *Meeting Place*: A circular municipal council house, covered timber barn, or stone moot-ground capable of seating 60 to 80 household representatives within immediate earshot without acoustic amplification.
- *The First Failure to Expect*: Free-riding and clandestine resource theft (e.g., nocturnal opening of irrigation sluices, grazing extra cows in deep forest) by a powerful family faction, which escalates into armed clan feuds if the village assembly lacks the collective will to enforce graduated sanctions[16].

The District Level (Approximately 3,000 People)

At the district scale—spanning roughly 10 villages across a single watershed—governance must take the form of a representative confederal republic modeled on the Haudenosaunee bicameral council[65]. The district council federates delegates from each village into two deliberative chambers: an Upstream Chamber and a Downstream Chamber, requiring mutual agreement to pass regional water-allocation rules, trade compacts, and inter-village road maintenance duties[59]. To permanently neutralize military capture, the district institutionalizes a strict dual-key defense architecture: civil sachems have zero military command, while elected war chiefs can mobilize forces only with the explicit consent of the village assemblies and can

hold no legislative or judicial office[65]. An independent council of elder women or matrilineal matriarchs holds exclusive power to confirm civil appointments and exercise direct impeachment (*knocking off the horns*) of any leader displaying corrupt or despotic traits[68].

- *Prerequisites to Start*:
  - *Functional Literacy*: A literate diplomatic cohort trained in administrative jurisprudence, contract law, standard weights, and surveying trigonometry.
  - *Physical Records*: A physical treaty archive (inscribed wampum belts, lead tablets, or parchment treaty rolls) recording mutual defense compacts, upstream-downstream water flow guarantees, and regional boundaries[65].
  - *Shared Calendar*: An aligned district-wide market and assembly calendar based on astronomical equinoxes and harvest festivals, synchronizing regional trade fairs, appellate court sessions, and confederal assemblies.
  - *Meeting Place*: A fortified, neutral council compound situated at a central geographical hub (e.g., a river confluence or island) outside the territorial boundary of any single village, preventing local domination[65].
- *The First Failure to Expect*: Upstream defection: an upstream village facing drought constructs illegal timber weirs to hoard the river flow, daring downstream communities to initiate kinetic warfare[59]. Alternatively, wealthy traders amass surplus grain and livestock, purchase impoverished smallholder plots, and form an armed oligarchy that violently overawes the confederal assembly[73].

Principles

1. A common-pool resource regime cannot prevent tragic depletion without clearly defined, defensible boundaries that exclude non-members and restrict internal harvesting to verifiable carrying capacities[17].
2. Centralizing defensive, regulatory, or intelligence authority into a single unipolar apex creates an existential single point of capture that systematically selects for Machiavellian and psychopathic leadership[14].
3. The deliberate elimination of formal offices, written statutes, and structured voting does not prevent hierarchy, but converts governance into an unaccountable tyranny operated by informal, charismatic cliques[3].
4. Egalitarian social stability cannot be preserved through moral exhortation or cultural goodwill alone, but requires an armed, intentional coalition of subordinates enforcing an escalating reverse dominance hierarchy against potential alphas[16].
5. Common-pool appropriation rights remain ecologically stable across generations only when formally coupled to private maintenance liabilities, capping communal consumption by private carrying capacity[70].
6. Asymmetric geographical leverage can be governed without centralized coercion only if downstream actors possess a biophysical or ecological feedback mechanism that renders upstream defection self-destructive to the predator[59].
7. A stateless, polycentric legal order inevitably degenerates into warlordism and foreign subjugation if it permits the unconstrained concentration of capital and tax-extracting

instruments within private lineages[73].

8. Operational resource decisions must be permanently vested in the lowest nested institutional tier that contains both the biophysical boundaries of the resource and the social consequences of its extraction[17].
9. Civil administration and military force must be structurally severed through dual-key architectures, ensuring that civil magistrates possess no standing coercive army and military commanders possess no legislative, fiscal, or judicial power[65].
10. Any governance official, whether civil or military, must be subject to immediate, low-cost recall and destitution by an independent, countervailing assembly, permanently preventing the consolidation of autocratic tenure[65].

Works cited

1. A Complete Knowledge Base Of HUMAN 3.0 - Dan Koe, https://thedankoe.com/letters/a-complete-knowledge-base-of-human-3-0/
2. A Problem Well-Stated Is Half-Solved with Daniel Schmachtenberger, https://www.humanetech.com/podcast/a-problem-well-stated-is-half-solved
3. Game B Wiki, https://www.gameb.wiki/index.php?title=Game_B
4. Generator Functions & the Third Attractor — Schmachtenberger on, https://2014p.com/generator-functions.html
5. In Search of the Third Attractor Part 1 - The Consilience Project, https://consilienceproject.org/media/daniel-schmachtenberger-in-search-of-the-third-attractor-part-1/
6. Understanding Game B Concepts | PDF - Scribd, https://www.scribd.com/document/442863434/On-Game-B-pdf
7. The Vulnerable World Hypothesis, https://notes.aarnphm.xyz/dGhvdWdodHMvVGhlLVZ1bG5lcmFibGUtV29ybGQtSHlwb3RoZXNpcw
8. Beyond the Tipping Point: From Breakdown to Breakthrough - Medium, https://medium.com/disruptive-leaders-journal/were-not-in-an-era-of-change-we-re-in-a-change-of-era-10c9b7655502
9. The Roundtable — the metacrisis, made legible, https://thethirdattractor.org/
10. Civilisational Transition: Game B to Game A | by MR (old) - Medium, https://medium.com/@mjram111/socio-economic-transition-game-b-to-game-a-as-a-parallel-to-civzero-to-civone-78e5de6bc2ba
11. The Vulnerable World Hypothesis - Nick Bostrom, https://nickbostrom.com/papers/vulnerable.pdf
12. (PDF) The Vulnerable World Hypothesis - ResearchGate, https://www.researchgate.net/publication/335661141_The_Vulnerable_World_Hypothesis
13. Summary: The Vulnerable World Hypothesis - Medium, https://rharling.medium.com/summary-the-vulnerable-world-hypothesis-e0fd23126f9a
14. "Effective Altruism, Longtermism, and the Problem of Arbitrary Power,

https://www.thephilosopher1923.org/post/a-mirror-for-tech-bros
15. Corporate Psychopathy Insights | PDF | Behavioural Sciences - Scribd, https://www.scribd.com/document/73187075/Corporate-Psychopathy-Talking-the-Walk-Babiak-Neumann-Hare
16. The Evolution of Equal Liberty Through Capital Punishment, http://darwinianconservatism.blogspot.com/2024/10/the-evolution-of-equal-liberty-through.html
17. How Fisheries Are Made Governable - OSF, https://osf.io/download/zavwc
18. A Journey To GameB - Medium, https://medium.com/@memetic007/a-journey-to-gameb-4fb13772bcf3
19. Vulnerable world hypothesis - Wikipedia, https://en.wikipedia.org/wiki/Vulnerable_world_hypothesis
20. Metatechnological Mapping and the Ye Wenjie Effect: Mitigating, https://doi.org/10.1111/1758-5899.13469?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
21. Vulnerable world hypothesis: Bostrom's idea explained - Electrified, https://electrified.vlvd.net/@miguel/vulnerable-world-hypothesis-bostrom-s-idea-explained
22. Latest Features & Products in Artificial Intelligence, https://analytics248.rssing.com/chan-7760276/all_p79.html
23. A strange ideology is behind the AI pause | Information Age | ACS, https://ia.acs.org.au/article/2023/a-strange-ideology-is-behind-the-ai-pause.html
24. Notes on the Vulnerable World Hypothesis - Michael Nielsen, https://michaelnotebook.com/vwh/
25. The Precipice – Existential Risk and the Future of Humanity.TobyOrd, https://www.researchgate.net/publication/350605374_The_Precipice_-_Existential_Risk_and_the_Future_of_HumanityTobyOrd_2020London_Bloomsbury_Publishing_480_pp_2250_hb
26. Toby Ord on the precipice and humanity's potential futures, https://80000hours.org/podcast/episodes/toby-ord-the-precipice-existential-risk-future-humanity/
27. Minimizing Global Catastrophic and Existential Risks from Emerging, https://www.researchgate.net/publication/256040275_Minimizing_Global_Catastrophic_and_Existential_Risks_from_Emerging_Technologies_Through_International_Law
28. Reflecting on the Long Reflection - Felix Stocker, https://www.felixstocker.com/blog/reflecting-on-the-long-reflection
29. Toby Ord, The Precipice: Existential Risk and the Future of Humanity, https://gcrinstitute.org/papers/064_precipice.pdf
30. Luke Kemp: “Existential Risks: The Biggest Threats to Life as We, https://www.thegreatsimplification.com/episode/153-luke-kemp
31. Governance and Global Justice - CSER, https://www.cser.ac.uk/work/research-themes/governance-and-global-justice/
32. The Cartography of Global Catastrophic Governance (2019), https://globalchallenges.org/library/the-cartography-of-global-catastrophic-gove

rnance-2019/
33. Global policymakers and catastrophic risk - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC8637034/
34. Societal Collapse in a Warming World: A Risk Manager's Lens - GARP, https://www.garp.org/podcast/societal-collapse-in-a-warming-world-a-risk-managers-lens
35. In Search of a Methodology to Study Existential Risk - ResearchGate, https://www.researchgate.net/publication/358164658_Democratising_Risk_In_Search_of_a_Methodology_to_Study_Existential_Risk
36. Goliath's Curse: The History and Future of Societal Collapse — A, https://cisac.fsi.stanford.edu/events/goliaths-curse-history-and-future-societal-collapse-book-talk-and-expert-discussion-luke
37. Bizarre and Dangerous Utopian Ideology Has Quietly Taken Hold of, https://truthout.org/audio/bizarre-and-dangerous-utopian-ideology-has-quietly-taken-hold-of-tech-world/
38. The Dangers of Effective Altruism - Population Balance, https://www.populationbalance.org/podcast/alice-crary
39. TESCREAL — Futures Garden, https://garden.johanneskleske.com/tescreal
40. The TESCREAL Bundle: Eugenics and the promise of utopia, https://montrealethics.ai/the-tescreal-bundle-eugenics-and-the-promise-of-utopia-through-artificial-general-intelligence/
41. Philosophy Against the Present: The Foundations and Critique of, https://umbau.hfg-karlsruhe.de/posts/philosophy-against-the-present-the-foundations-and-critique-of-longtermism
42. Getting Good | The Point Magazine, https://thepointmag.com/examined-life/getting-good-effective-altruism/
43. (summary) Amia Srinivasan- “Stop the Robot Apocalypse” book, https://distortedtimes.wordpress.com/2017/07/05/summary-amia-srinivasan-stop-the-robot-apocalypse-book-review-2015/
44. Alice Crary · The toxic ideology of longtermism (2023), https://www.radicalphilosophy.com/commentary/the-toxic-ideology-of-longtermism
45. Alice Crary · Against 'Effective Altruism' (2021) - Radical Philosophy, https://www.radicalphilosophy.com/article/against-effective-altruism
46. Suggested reading: Stop the robot apocalypse, https://exoticandirrational.blogspot.com/2015/11/suggested-reading-stop-robot-apocalypse.html
47. The Lessons of Effective Altruism - Ethics & International Affairs, https://www.ethicsandinternationalaffairs.org/journal/the-lessons-of-effective-altruism
48. The struggle to govern the commons - PubMed, https://pubmed.ncbi.nlm.nih.gov/14671286/
49. The Struggle to Govern the Commons | Request PDF - ResearchGate, https://www.researchgate.net/publication/8963669_The_Struggle_to_Govern_the_Commons

50. The Struggle to Govern the Commons - Semantic Scholar, https://www.semanticscholar.org/paper/The-Struggle-to-Govern-the-Commons-Dietz-Ostrom/d2e297f50d7df67f1757c0cdf4572a2693a1fa88
51. (PDF) Corporate Psychopathy: Talking the Walk - ResearchGate, https://www.researchgate.net/publication/43346831_Corporate_Psychopathy_Talking_the_Walk
52. Psychopathy: The Dark Side of Humanity, https://craigsneumann.weebly.com/blog/psychopathy-the-dark-side-of-humanity
53. The criminalisation of corporate psychopaths: a holistic inquiry, https://www.tandfonline.com/doi/full/10.1080/10345329.2023.2207233
54. [PDF] Corporate psychopathy: Talking the walk. - Semantic Scholar, https://www.semanticscholar.org/paper/Corporate-psychopathy%3A-Talking-the-walk.-Babiak-Neumann/e39960b1a4b6d43feb7511a6acd6e9f280d4c0ec
55. Corporate law and corporate psychopaths - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC9090396/
56. Hierarchy in the Forest: The Evolution of Egalitarian Behavior, https://opac.uma.ac.id/repository/0674390318.pdf
57. Power in Cultural Evolution and the Spread of Prosocial Norms, https://www.journals.uchicago.edu/doi/pdf/10.1086/700767
58. Purposive Social Selection and the Evolution of Human Altruism - Ovid, https://www.ovid.com/journals/ccres/pdf/10.1177/1069397108320422~purposive-social-selection-and-the-evolution-of-human
59. Stephen Lansing, 1991, PRIESTS AND PROGRAMMERS, Princeton., http://www.andreasaltelli.eu/file/repository/Lansing.pdf
60. The story of the disaster in Bali when a centuries old cultural custom, https://www.reddit.com/r/slatestarcodex/comments/dmp9x1/the_story_of_the_disaster_in_bali_when_a/
61. Foucault and the Water Temples - J. Stephen Lansing, http://www.slansing.org/uploads/4/3/4/1/43417789/lansing_reply_to_helmreich.pdf
62. An analysis of the Lansing-Kremer model of Bali, https://faculty.sites.iastate.edu/tesfatsi/archive/tesfatsi/LansingBaliCritique.Janssen.pdf
63. Emergent Properties of Balinese Water Temple Networks, http://www.islandsoforder.com/uploads/4/3/4/1/43417789/lansing_1993_amanthropol_v95_p97.pdf
64. 1 Foreword to 2nd Edition of Priests & Programmers Steve Lansing, http://www.slansing.org/uploads/4/3/4/1/43417789/clark_lansing_foreword_pp.pdf
65. The Ordeal of the Longhouse: The Peoples of the Iroquois League, https://dokumen.pub/the-ordeal-of-the-longhouse-the-peoples-of-the-iroquois-league-in-the-era-of-european-colonization-0807843946-9780807843949.html
66. An Analysis Of The 1613 Tawagonshi Treaty – Page 2, https://www.onondaganation.org/history/2012/an-analysis-of-the-1613-tawagonshi-treaty/2/
67. Settlers in Indian Country - Treatied Spaces Research Group, https://treatiedspaces.com/wp-content/uploads/2023/09/Prior-2020.pdf

68. Otiyaner: The "Women's Path" Through Colonialism, https://atlantisjournal.ca/atlantis/en/article/download/1050/1007/1321
69. law and politics - Dentons, https://www.dentons.com/import/~/media/D3899D236B424013B3FF2F41A56BB8DB.ashx
70. Analyzing long-enduring CPRs, https://faculty.washington.edu/stevehar/GtC3.pdf
71. History of the Alps, 1500-1900: Environment, Development, and, https://dokumen.pub/history-of-the-alps-1500-1900-environment-development-and-society-9781935978138-9781933202341-9781933202419.html
72. NTNU, Trondheim Fall 2003 - Erling.Berge, https://erlingberge.no/11%20DesignPrinciplesII.pdf
73. The Age of the Sturlungs - Jesse Byock, https://viking.ucla.edu/publications/articles/age_of_the_sturlungs.pdf