

# The Leviathan Hypothesis: Why Evil Movements Become Unstoppable

## A Framework for Understanding and Countering Fascism, Extremism, and Authoritarian Capture



---

### Richard Dobson

Independent Researcher, Clara Futura, Andorra and Research Institute Netherlands for Harmonizing Human and Artificial Intelligence (RINHUMAI), The Netherlands Richard Dobson. Email:

[richard@rinhumai.org](mailto:richard@rinhumai.org)

### Dirk K.F. Meijer, PhD

em. Professor in Pharmacokinetics and Pharmacotherapy, Research Institute Netherlands for Harmonizing Human and Artificial Intelligence (RINHUMAI), Groningen, The Netherlands. Email:

[meij6076@planet.nl](mailto:meij6076@planet.nl)

## Summary

Some political movements spread like wildfire and then fade away. Others cross a dangerous threshold and become self-reinforcing systems that resist suppression, reshape identities, and force entire societies to respond. What separates a passing trend from an extremist movement that captures institutions and threatens democracy? This paper introduces a testable framework for identifying when movements—from fascism to religious extremism to digital authoritarianism—become governance crises rather than manageable social trends. Drawing on research across psychology, political science, and information network theory, it identifies four key warning signs: first, how movements spread and resist suppression; second, why participants commit despite personal cost; third, how symbols and rituals create self-reinforcing allegiance, implying loyalty to a leader, a group, or cause; and fourth, how movements capture or overwhelm institutions. Urgently relevant to today's resurgent authoritarian movements, this framework offers early warning indicators that can be monitored in real-time, including online radicalization. It provides practical countermeasures targeting the psychological, social, and institutional levels where dangerous allegiances form. It includes a typology distinguishing grassroots movements from engineered propaganda campaigns. And it generates testable predictions that can be validated or disproven through empirical research. Rather than asking "why do movements spread," this work asks the

harder question: "when does a movement become impossible to suppress—and what can we do before that threshold is crossed?" The answer has profound implications for protecting democracy in the digital age.

**Keywords:** emergence, threshold dynamics, allegiance, costly commitment, governance response, social movements, fascism, authoritarianism, memetics, consciousness, socio-technical systems

# 1. Introduction

## 1.1 The Problem: Spread is Not Impact

Societal movements vary not only in how far they spread, but in what they do to the systems that encounter them. Some movements trend, surge, and fade. Others generate institutions, reshape moral order, and compel governance reactions that persist long after the initial burst of attention. In empirical terms, these are different phenomena.

A familiar error in contemporary discourse treats "viral spread" as equivalent to "civilizational reorganization." Yet they are not the same. Spread, rather, concerns replication, and impact concerns transformation of identity, obligation, as well as governance. The present paper asks a narrower question than "Why do movements spread?" It asks: Under what conditions does a movement in our world become difficult to suppress and force a change in governance response?

Contemporary examples are urgent and visible. Across democracies and semi-democracies, authoritarian and fascist formations are resurgent [1][2]. These movements exhibit distinctive patterns: they spread rapidly through digital networks, generate intense identity fusion among participants, resist factual correction, trigger polarization cascades, and force institutional responses ranging from platform moderation to constitutional crisis [3][4]. Standard diffusion models explain their propagation; standard incentive theories explain opportunistic participation. But something remains under-explained: why do some movements cross a threshold into self-reinforcing allegiance systems that governance cannot manage from the outside?

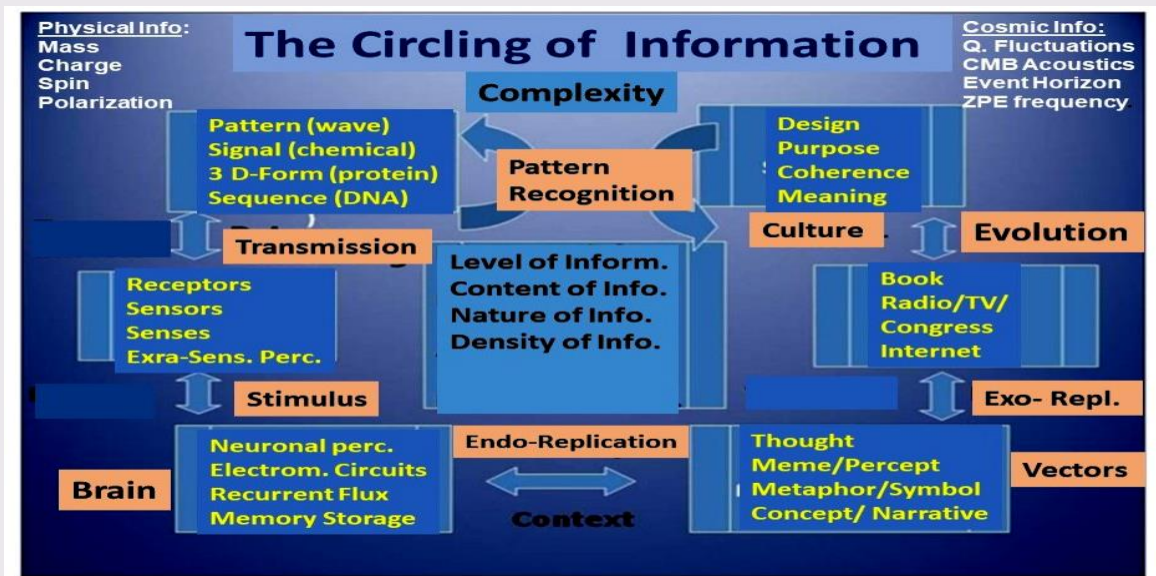


Figure 1: The Different Guises of Information and Information Processing in Human Culture

## 1.2 Why "Leviathan"?

The term "Leviathan" is used as a structural label for phenomena that exceed ordinary governance tools. In ancient Near Eastern and biblical tradition, Leviathan appears as an untamable creature (Job 41:1-34)—powerful, resistant to control, beyond ordinary management. Thomas Hobbes (1651) adapted the image to describe the sovereign state as a constructed power that transcends individual wills. This paper uses "Leviathan" in a third sense: as a designation for movements or participatory systems that become structurally ungovernable through ordinary means, forcing governance escalation into containment, capture, or systemic reconfiguration.

This is not a theological claim. It is a methodological designation for a distinct pattern: threshold crossing combined with self-reinforcing allegiance combined with governance disruption. Throughout this paper, "Leviathan" can be read interchangeably with "threshold allegiance system"; the term is retained only as a mnemonic shorthand, not as a metaphysical entity. The term captures the encounter with emergent reality that resists external control. Annex D provides symbolic background for interested readers; the model does not depend on it.

## 1.3 Central Scientific Question

When does a movement cross a threshold into a self-reinforcing allegiance system that governance cannot manage from the outside?

Our model proposes that a key diagnostic is not merely growth, but governance response. When institutions shift from ordinary policing or administration into specialized containment, doctrinal compression, legal consolidation, institutional capture, or exceptional measures, something structurally important has occurred.

## 1.4 Contribution and Structure

The present paper contributes operational definitions of key constructs including allegiance, threshold, field, and Leviathan-class. It provides indicator families for empirical coding across four dimensions: Threshold (T), Allegiance (A), Symbolic-Ethical recursion (SE), and Power (P). It offers a movement typology distinguishing emergent, constructed, and hybrid phenomena. It specifies falsifiability criteria indicating how the model could be wrong. It includes a strict-mode scoring sheet in Annex A to reduce interpretive drift. It presents an empirical protocol for real-time testing in online debates. And it applies this framework to contemporary authoritarianism and socio-technical systems.

The Leviathan Hypothesis is operationalized as a middle-range diagnostic framework with a strict-mode scoring procedure, designed as much to reject borderline cases as to classify positive ones. This is not a grand unifying theory but a structured tool for comparative inquiry.

The aim is to make the model testable, comparative, and useful for both researchers and practitioners concerned with countering dangerous allegiance dynamics.

The document proceeds as follows. Section 2 positions the model within existing literature. Section 3 presents the core model. Sections 4 and 5 define constructs and indicators. Section 6 develops the typology. Section 7 specifies falsifiability criteria. Section 8 presents application tests, including the online debate protocol. Section 9 outlines implications for governance and countermeasures. Section 10 addresses limitations. Section 11 presents memetics as a theory for understanding evil information spreading. Section 12 concludes.

## 2. Positioning the Model Within Existing Frameworks

### 2.1 What Existing Frameworks Explain Well

The Leviathan Hypothesis is not offered as a replacement for established theories on information diffusion, but as a complement aimed at a specific explanatory remainder. Existing frameworks cover substantial ground.

Network diffusion models as developed by Rogers (2003) and Centola (2018) explain propagation through social ties, threshold effects in adoption, and the role of weak ties versus strong ties. These models predict how innovations, behaviors, and beliefs spread through populations via network structure [5][6].

Charismatic authority as analyzed by Weber (1922) explains how leadership and perceived exceptional legitimacy mobilize participation. Charisma operates through emotional identification with a leader who embodies transcendent values, enabling rapid mobilization and devotion [7].

Social cohesion and ritual as theorized by Durkheim (1912) explain how sacred/profane boundaries and collective effervescence strengthen group identity. Rituals create solidarity by synchronizing emotion and attention, marking in-group boundaries [8].

Mass psychology and totalism as examined by Le Bon (1895), Arendt (1951), and Lifton (1961) explain susceptibility under conditions of social fragmentation, identity threat, and atomization. These frameworks illuminate how ideological systems exploit psychological needs for belonging, certainty, and purpose [9][10][11]

Memetics, as originated by Dawkins (1976), developed by Blackmore (1999), and further elaborated by Meijer in his information-universe and meme–virus work, explains replication of cultural units via cognitive and social selection pressures. Memes compete for attention and transmission based on emotional salience, simplicity, and cultural fitness

Identity fusion theory as developed by Swann et al. (2012) and Whitehouse et al. (2017) explains extreme commitment through fusion of personal and group identity, predicting willingness to sacrifice for the group [14][15].

Radicalization pathways as mapped by McCauley and Moskalenko (2017) trace mechanisms by which individuals and groups move toward extremism, including personal grievance, group polarization, and competition dynamics [16].

These are all powerful toolsets. Many societal movements can be explained adequately within them.



**Figure 2:** *How Will We Know when AI Becomes Conscious? One Aspect could be the Ability to Recognize and Register Underlying Holistic Patterns in the Multi-layer Information Processing (right above), Coupled with Predictive Information Processing and Sudden Misalignment with Ethical Values (norm, honesty, respect and truth).*

## 2.2 The Explanatory Remainder

The Leviathan Hypothesis addresses a narrower observation: some movements exhibit a constellation of features not always explained well by propagation, incentives, or charisma alone.

First, costly allegiance before incentives: participants accept risk, such as loss of status, livelihood, relationships, or life, without immediate material advantage or coercion. Second, intergenerational persistence under pressure: durability across generations despite repression, marginalization, or hostility. Third, ethical reorganization: new norms restructure kinship, violence, sexuality, property, or authority at deep levels. Fourth, suppression backfire: external pressure increases internal cohesion and commitment rather than dissolving the movement. Fifth, governance escalation: institutions shift strategies from ordinary management to specialized containment, capture, or exceptional measures because standard control fails.

When these features co-occur above a certain threshold, the movement becomes a governance problem rather than simply a social trend. The hypothesis does not claim these features are unique to any one tradition; it claims that when they converge, a distinct structural dynamic emerges.

Contemporary authoritarian and fascist movements exemplify this pattern. Digital propaganda amplifies memetic spread (memetics explains transmission), elite bargains enable institutional capture (power mechanics explain leverage), but the durability under counter-messaging and the willingness of participants to accept social cost for participation (job loss, family estrangement, legal risk) suggest symbolic-ethical recursion and identity fusion mechanisms at work [17][18][19]. The Leviathan framework integrates these aspects.

## 2.3 Consciousness, Memetics, and Allegiance Formation

A critical bridge between individual psychology and collective dynamics lies in understanding how consciousness-level processes enable memetic transmission and allegiance formation. Memetics explains what spreads; consciousness research explains why certain minds are vulnerable.

Cognitive autonomy—the capacity to evaluate information independently, tolerate ambiguity, and resist totalizing narratives—varies across individuals and contexts (Cook-Greuter 2005; Kegan 1982) [20][21]. When cognitive autonomy is low (due to developmental stage, stress, social fragmentation, or information overload), individuals become susceptible to several interrelated dynamics.

Memetic hijacking occurs when simple, emotionally charged narratives bypass critical evaluation. Identity fusion involves the absorption of personal identity into group symbols and obligations. Ultimacy encoding means the adoption of totalizing frameworks that answer, "what matters most" and pre-empt competing values.

This process operates at scale in digital environments where attention is commodified, algorithmic amplification rewards emotional intensity, and echo chambers reinforce boundary mechanisms [22][23]. The result is not mere "belief" but participatory formation—repeated engagement that trains perception, emotion, and action toward alignment with the movement's symbolic-ethical system.

Understanding this consciousness-memetics-allegiance pipeline is essential for designing countermeasures (Section 9.2).

## 2.4 Scope: What This Paper Does Not Claim

To reduce category confusion, the Leviathan Hypothesis explicitly does not claim that a Leviathan-class phenomenon is morally good; that it is metaphysically true; that it is inevitable; that it is always "bottom-up" (construction and coercion matter); or that growth alone implies Leviathan dynamics.

Thus, it is a pattern model: a way to identify when a movement becomes structurally difficult to govern externally, regardless of the movement's moral valence or truth claims.

# 3. Model Overview

## 3.1 Plain-English Definition

A Leviathan-class phenomenon is a movement or participatory system that crosses a threshold into self-reinforcing allegiance, such that external authorities cannot suppress it or manage it through ordinary means, and must shift into containment, capture, or systemic reconfiguration. The point is not the metaphor. The central point is the threshold and the governance response.

## 3.2 The Unit of Analysis Must Be Specified

A frequent failure in debate is to treat broad labels, such as "Christianity," "Islam," "Communism," "liberalism," or "the internet," as if they were single movements. For strict analysis, the unit must be specified: Which lineage or sub-movement? In what time window? In what region? Under what governance conditions?

Without this specification, the model becomes unfalsifiable because it can always be retrofitted after the fact. For example, "early Christianity in the Roman Empire, 30–313 CE" is a valid unit; "Christianity in general" is not.

### 3.3 Governance Response as an External Diagnostic

The Leviathan Hypothesis uses governance response as a partial proxy for threshold crossing. This is useful because governance responses are often documented: specialized laws, councils, decrees, bans, and surveillance regimes; doctrinal compression specifying "what must be believed" and "what is excluded"; integration into state structures (capture) or parallel structures (alternative institutions); and exceptional measures including emergency powers and mass mobilization.

Governance response does not prove the movement is "true" or "good." It indicates that the movement has become too structurally costly to ignore.

### 3.4 Model Components Summary

The model treats Leviathan-class phenomena as emerging when five conditions converge. First, participation becomes dense and repeated, as formation environments shape perception and action. Second, allegiance becomes costly and identity-linked, meaning commitment persists under risk. Third, symbols and ethics recurse through self-reinforcing meaning and obligation loops. Fourth, the movement replicates through local nodes via decentralized reproduction without central control. Fifth, governance escalates, shifting from ordinary management to containment, capture, or reconfiguration.

*These are not mystical claims. They rather represent empirical, comparative indicators.*

## 4. Definitions and Operational Constructs

### 4.1 Allegiance

Allegiance is not mere belief. It is a durable pattern comprising commitment (what one will do and endure), identity (who one becomes by participating), and obligation (what one "must" do, even under cost). Belief may be part of allegiance, but allegiance is measured primarily by behavior under pressure.

### 4.2 Costly Allegiance

Costly allegiance is allegiance that persists when participation brings real risk: loss of status, livelihood, safety, relationships, or life. Costly allegiance distinguishes superficial affiliation (cheap signals, bandwagon effects) from durable commitment (risk-bearing participation). This distinction is critical for differentiating Leviathan-class dynamics from opportunistic or coerced participation.

### 4.3 Participation

Participation means repeated engagement in a shared practice environment that forms perception and action. Participation can include ritual and liturgy; meetings and training regimes; media consumption habits such as news sources, podcasts, and social feeds; public symbols including flags, slogans, and dress codes; mutual aid networks; and disciplinary norms such as codes of speech,

behavior, and loyalty tests. The model treats participation as formative: it trains what participants notice, value, and do.

## 4.4 "Field" as a Methodological Term

The term field is used functionally to describe a distributed system of interactions that shapes behavior across participants. It does not imply a literal physical field. A "field" here is shorthand for distributed reinforcement (mutual encouragement, social proof), mutual amplification (network effects, cascades), shared symbols (common language, icons, narratives), coordinated obligation (reciprocal commitments, loyalty expectations), and recursive norms (feedback loops between identity and action). If "field" is distracting, it can be replaced with allegiance system or participatory ecology.

## 4.5 Threshold Behavior

A threshold is the point at which a system shifts from linear change to non-linear dynamics: feedback loops dominate. In social systems this can appear as sudden cascades of idea recruitment (exponential growth phases), rapid mental polarization (us/them boundaries harden), suppression that increases commitment (martyrdom effects, persecution narratives), self-reinforcing replication through local nodes (decentralized spread), and governance escalation (institutional responses shift mode). Thresholds are not necessarily instantaneous. They are detectable through discontinuities in rate, intensity, or institutional reaction.

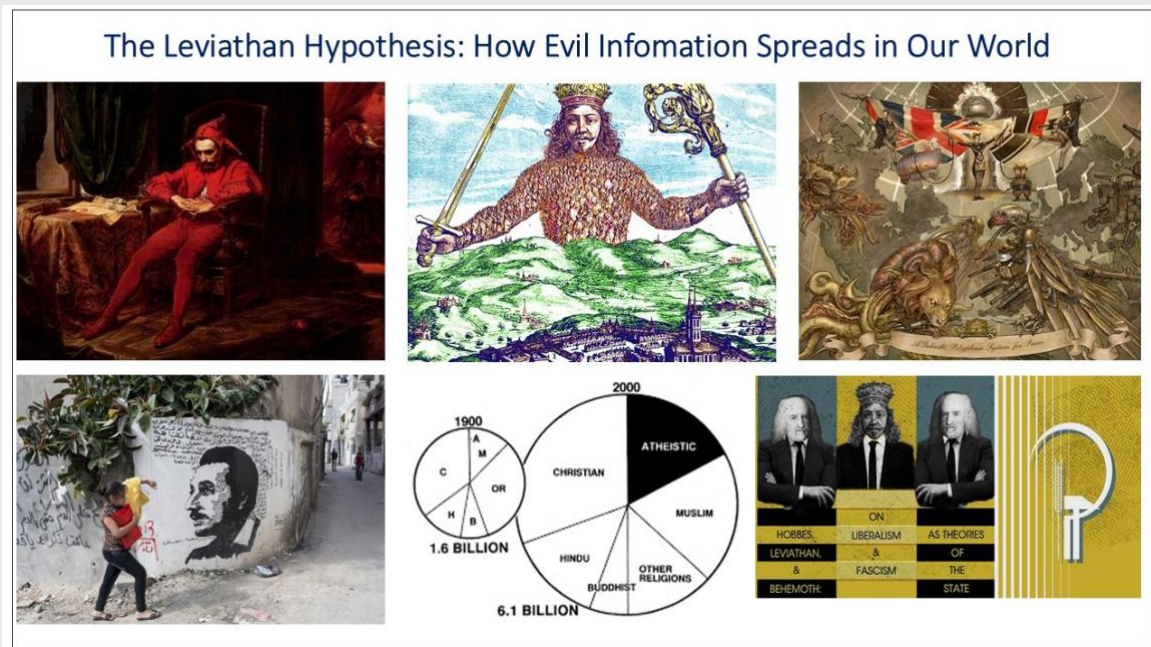
## 4.6 Governance Response Modes

When systems encounter Leviathan-like dynamics, governance tends to shift into recognizable modes. Ordinary management involves routine policing, administration, and minimal special measures. Containment involves special doctrine, law, or structure designed to stabilize the phenomenon, such as religious tolerance laws or platform moderation regimes. Capture occurs when the movement takes over institutions or becomes the institution, as with Constantine and Christianity or the Bolsheviks and the Soviet state. Suppression involves attempted eradication acting through force. Accommodation involves negotiated coexistence with boundary rules. Fragmentation management involves treating a movement as multiple competing lineages to reduce monolithic threat. The model predicts that Leviathan-class phenomena tend to provoke containment or capture when suppression fails, (Fig.3)

## 4.7 Construction vs. Emergence

Not all "powerful movements" are emergent. Rather, some are heavily engineered. The model therefore distinguishes three types. Emergent Leviathans involve bottom-up, self-reinforcing participation with limited central design, such as early Christianity or grassroots fascist cells. Constructed Leviathans involve top-down engineered allegiance using coercion and propaganda, such as Stalinist USSR or Maoist China. Hybrid Leviathans involve a feedback loop between bottom-up enthusiasm and top-down control, such as Nazi Germany or contemporary digital authoritarianism.

This distinction prevents the model from romanticizing destructive regimes or over-attributing agency to symbolism alone. It also clarifies intervention points: emergent systems require different counterstrategies than constructed ones.



**Figure 3:** *The Leviathan Appears in Many Forms, (Hobbs, right below), Being Sometimes Hidden in (Pseudo)-Religion*

## 5. Operational Indicators

The Leviathan Hypothesis becomes testable through specification of indicator families. In strict mode, a case is evaluated by whether the indicator constellation is present, not by rhetorical fit. (Fig. 3)

### 5.1 Threshold Indicators (T)

A movement shows threshold dynamics when several of the following occur. T1, nonlinear growth, means recruitment accelerates beyond linear expectation through exponential or super-exponential phases. T2, local-node replication, means new groups form semi-independently, without central coordination, through decentralized reproduction. T3, suppression backfire, means repression increases cohesion and commitment rather than dissolving the movement. T4, polarization, means identity boundaries harden, neutrality becomes costly, and "with us or against us" dynamics emerge. T5, governance escalation, means authorities shift from ordinary management to special measures including new laws, emergency powers, or institutional redesign.

### 5.2 Allegiance Profile Indicators (A)

Leviathan-class phenomena typically contain multiple allegiance types. The key is which dominate. A1, affective allegiance, involves emotional intensity, belonging, and enthusiasm, which is high but often unstable. A2, symbolic allegiance, involves identity fusion with symbols and narratives, which is core to Leviathan dynamics. A3, instrumental allegiance, involves opportunistic affiliation for material advantage, which is present but insufficient alone. A4, constrained allegiance, involves coerced participation driven by fear, which is typical of constructed regimes. A5, costly allegiance, involves risk-bearing commitment without immediate payoff, which is diagnostic for emergent Leviathans.

A Leviathan-class phenomenon normally requires meaningful levels of A5 (costly allegiance) or a powerful hybrid of A2 combined with A4 (symbolic identity plus coercive constraint).

### 5.3 Symbolic-Ethical Recursion Indicators (SE)

This indicator family describes whether a movement internalizes governance—whether its "codes" obligation into the participant. SE1, ritual density, involves repeated practices that train perception and action such as rallies, liturgies, and daily media rituals. SE2, ethical recursion, means norms reinforce identity through the cycle "who we are" leading to "what we do" leading back to "who we are." SE3, ultimacy encoding, means the movement answers "what matters most," pre-empting competing values. SE4, boundary mechanisms, means clear in/out membership practices such as purity tests, loyalty oaths, and excommunication. SE5, formation regime, involves training under pressure that stabilizes behavior through disciplines of attention, speech, obedience, and restraint.

This family captures the difference between mere opinion and lived allegiance.

### 5.4 Power Mechanics Indicators (P)

To remain credible, the model must explicitly integrate power. P1, coercion capacity, involves ability to punish dissenters through state violence, mob enforcement, or social ostracism. P2, propaganda capacity, involves mass narrative shaping through media control and information dominance. P3, institutional capture, involves penetration of state, media, education, and judiciary. P4, elite bargains, involves alliances with power brokers such as financial elites, military, or religious authorities. P5, administrative enforcement, involves surveillance, bureaucracy, and legal infrastructure.

A movement can be Leviathan-like through emergence, construction, or hybridization. Power mechanics help classify which type.

### 5.5 Minimal Operational Rule

To avoid "everything fits," the model uses a minimal rule: A candidate Leviathan-class phenomenon should show multiple threshold indicators (T), some costly or identity-linked allegiance (A), symbolic-ethical recursion (SE), and a governance response shift. The strict-mode scoring sheet (Annex A) formalizes this.

### 5.6 Leading and Lagging Indicators

For early-warning utility, the model distinguishes leading from lagging indicators. Leading indicators signal rising allegiance viscosity before overt governance disruption appears. These include identity-charge escalation, ultimacy framing, and exit-and-return patterns in participation, representing early intensification of A and SE indicators. Lagging indicators confirm that threshold crossing has occurred. These include governance escalation, legal exceptionalism, and institutional capture, representing P indicators and documented governance response.

Threshold crossing is most plausibly inferred where leading indicators converge before, or alongside, clear shifts in governance response. This distinction enables intervention before consolidation rather than reaction after the fact.

# 6. Typology of Leviathan-Like Phenomena

## 6.1 Why Typology Matters

A common objection asks: "If you can call Christianity, Nazism, and Beatlemania 'Leviathan,' your category is meaningless."

The typology answers: not all threshold surges are Leviathans in the same sense. Key discriminators are duration (does it persist across generations?), recursion depth (does it encode obligation and ultimacy?), and governance entanglement (does it force long-term institutional response?).

## 6.2 Typology Axes

The typology operates across three axes. The duration axis distinguishes short-duration surges from long-duration systems. The production mode axis distinguishes emergent, constructed, and hybrid phenomena. The allegiance core axis distinguishes affective, ideological-coercive, ethical-existential, and administrative types.

Axis	Values
Duration	Short-duration surge / Long-duration system
Production mode	Emergent / Constructed / Hybrid
Allegiance core	Affective / Ideological-coercive / Ethical-existential / Administrative

**Table 1:** Typology axes for classifying Leviathan-like phenomena

## 6.3 Four Illustrative Categories

The first category is the Affective Threshold Surge, characterized by high A1 (affective), moderate T (threshold), low SE (recursion), and low duration. Examples include pop-culture waves and some protest movements. The model predicts surge without long-duration governance entanglement.

The second category is the Ideological-Constructed Leviathan, characterized by high P (power), high A4 (constrained), significant A2 (symbolic), with SE that may be enforced. Durability depends on coercion and collapses when coercion fails. Examples include totalitarian regimes such as Stalinist USSR and Maoist China. The model predicts collapse when coercion is removed, unless deep SE recursion has developed.

The third category is the Ethical-Existential Leviathan, characterized by higher A5 (costly) combined with SE depth, and durable across generations. Governance response tends toward containment or accommodation rather than suppression. Examples include long-duration religious movements and some revolutionary ideologies. The model predicts persistence under pressure and intergenerational transmission.

The fourth category is the Hybrid Pathological Leviathan, characterized by emergent enthusiasm fused with coercion and capture, with catastrophic outcomes possible. It shows high T, mixed A (affective plus constrained plus symbolic), high P, and moderate-to-high SE. Examples include Nazi

Germany and contemporary authoritarian-populist movements. The model predicts rapid escalation, governance crisis, and violence.

This typology prevents overgeneralization and enables comparison without moral confusion.

## 6.4 Contemporary Fascism as Hybrid Pathological Type

Contemporary authoritarian and fascist movements fit the Hybrid Pathological category.

Regarding T indicators, they show nonlinear growth via social media (T1), local-node replication in cells and online communities (T2), suppression backfire where fact-checking increases commitment (T3), extreme polarization (T4), and governance escalation from platform moderation to constitutional crisis (T5).

Regarding An indicator, they show high affective allegiance at rallies (A1), strong symbolic fusion with leader/nation symbols (A2), instrumental opportunism among elites (A3), increasing coercion against dissenters (A4), and costly allegiance among core participants willing to risk jobs and relationships (A5).

Regarding SE indicators, they show ritual density in rallies and media consumption (SE1), ethical recursion around loyalty/betrayal (SE2), ultimacy encoding around "civilization at stake" narratives (SE3), boundary mechanisms distinguishing "real citizens" from "traitors" (SE4), and formation regimes in training camps and online radicalization pathways (SE5).

Regarding P indicators, they show growing coercion capacity (P1), massive propaganda infrastructure (P2), institutional capture of courts/media/police (P3), elite bargains with oligarchs/military (P4), and administrative enforcement through surveillance/bureaucracy (P5).

This pattern predicts escalation risk and intrinsic governance crisis. Countermeasures must address all four indicator families (Section 9.2).

## 7. Falsifiability and Disconfirmation Criteria

A testable model should specify how it could be wrong. Annex A operationalizes these criteria in a scoring sheet intended to constrain overextension and retrospective fitting.

### 7.1 Criterion 1: Comparative Counterexamples

If other movements show the same indicator constellation—costly allegiance, intergenerational persistence, ethical reorganization, governance escalation—yet can be fully explained without Leviathan dynamics, then the Leviathan label adds little.

### 7.2 Criterion 2: Full Reduction to Standard Mechanisms

If the best explanation of a case is fully captured by network diffusion, incentives, elite bargains, or opportunism, with no explanatory remainder, then Leviathan is unnecessary.

### 7.3 Criterion 3: Symbolic-Ethical Recursion Proves Epiphenomenal

If symbolic and formation variables do not predict durability or costly commitment—if they are decorative rather than causally relevant—then the model's distinctiveness collapses.

## 7.4 Criterion 4: Absence of Threshold Indicators

The hypothesis predicts threshold behavior. If historical records show only linear diffusion and ordinary governance management (no escalation, no backfire, no node replication), the model weakens.

## 7.5 Criterion 5: Constructed Compliance Mistaken for Allegiance

If what appears as "allegiance" is primarily coerced compliance (high P, high A4, low A5) and collapses once coercion is removed, then the phenomenon is better classified as a constructed regime rather than Leviathan-class allegiance.

## 7.6 Criterion 6: Negative Cases

A strong model must identify non-Leviathans: movements that grew but remained governable, movements that surged but dissipated, and movements that were absorbed without governance escalation. For example, many intense but short-lived online campaigns or pop-cultural boycotts show sharp affective surges and polarization but no sustained costly allegiance, no deep symbolic-ethical recursion, and no durable governance escalation; under strict-mode scoring they should be classified as non-Leviathan, despite alarmist rhetoric at the time. Annex A (scoring sheet) is designed to help generate and track negative cases.

# 8. Application Tests and Empirical Protocol

## 8.1 Case Type: Affective Threshold Surge

The example class includes pop-cultural waves such as Beatlemania and Swiftie fandom. These cases often show T1 (nonlinear growth) and A1 (affective allegiance). They sometimes provoke governance reaction (crowd control, media regulation), but they usually have low SE depth and low intergenerational durability. The model predicts surge without long-duration governance entanglement.

## 8.2 Case Type: Hybrid Pathological Leviathan

The example class includes totalizing political movements such as Nazism 1933–1945 and contemporary authoritarian-populism. Such cases show strong threshold dynamics and intense identity fusion, but durability may depend heavily on coercion and administrative enforcement. Here the Leviathan model must not romanticize power mechanics (P) are central. The typology distinguishes emergent enthusiasm from constructed enforcement. The model predicts rapid escalation, governance crisis, and high violence risk.

For contemporary application, resurgent fascist and authoritarian movements from 2016 to the present exhibit this pattern across multiple democracies. Indicators T, A, SE, and P are all elevated. Governance responses range from platform moderation (containment) to attempted institutional capture [24][25][26].

### 8.3 Case Type: Ideological-Constructed Leviathan

The example class includes revolutionary regimes such as Stalinist USSR and Maoist China. These cases show strong P indicators and a mix of constrained and symbolic allegiance. The model predicts that once constructed enforcement becomes the primary sustaining mechanism, collapse is likely when coercion fails or legitimacy erodes, unless SE recursion is deep enough to persist independently. The model predicts collapse when coercion is removed (cf. USSR 1991).

### 8.4 Case Type: Ethical-Existential Leviathan

The example class includes long-duration religious movements such as early Christianity 30–313 CE and early Islam 622–750 CE. These cases often show higher SE recursion and formation regimes that sustain commitment under pressure and across generations. The model predicts that governance response will shift from suppression to containment/accommodation/capture because ordinary management fails. The model predicts intergenerational persistence and eventual institutional integration or accommodation.

As a methodological note, for traditions with multiple competing lineages, the unit of analysis must be disaggregated. "Christianity" is not one movement; it is a field ecology of competing sub-movements with different thresholds and power relations.

### 8.5 Empirical Protocol: Testing the Leviathan Hypothesis in Real-Time Online Debates

To make the Leviathan Hypothesis operationally testable, we developed a protocol for coding emergent allegiance dynamics in live online debates (Dobson 2025) [27]. This protocol bridges theoretical indicators to observable micro-level phenomena.

The design involves seeded topic debates in moderated online forums where participants discuss contentious issues such as immigration, climate policy, and cultural identity. Trained coders identify several phenomena: evidence posture (how participants treat factual claims), identity charge (degree of self/group fusion with position), governance talk (references to authority, legitimacy, "what must be done"), exit-and-return (participants who disengage but return with increased commitment), boundary marking (in-group/out-group language, purity tests), and ultimacy framing ("civilization at stake," "existential threat").

A thread is classified as "Leviathan-class" if evidence postures shift from inquiry to defense (refusal of falsifiability), identity charge escalates (personal stakes increase), governance talk emerges (calls for enforcement, exclusion), participants return after exits with higher commitment, and moderation attempts trigger suppression backfire.

The protocol includes negative controls (debates that remain deliberative) and moderation experiments (testing whether intervention style affects threshold crossing).

Protocol Code	Model Indicator
Evidence posture (inquiry to defense)	SE3 (ultimacy encoding), T4 (polarization)
Identity charge escalation	A2 (symbolic allegiance), A5 (costly)
Governance talk	T5 (governance escalation)
Exit-and-return with increased commitment	T3 (suppression backfire)
Boundary marking	SE4 (boundary mechanisms)
Ultimacy framing	SE3 (ultimacy encoding)

**Table 2:** Mapping online debate codes to Leviathan indicators

The protocol enables real-time early-warning detection of allegiance formation online. Platforms, educators, and civil-society actors can use adapted versions to monitor for threshold crossing indicators before radicalization consolidates, test moderation interventions (does pre-bunking reduce identity charge? does counter-messaging trigger backfire?), and design formation environments that strengthen cognitive autonomy and reduce susceptibility to totalizing narratives.

This addresses Dirk Meijer's call for practical countermeasures against "the spreading of evil information" (personal communication, 2026).

## 9. Implications

### 9.1 For Historical Sociology

The Leviathan Hypothesis reframes disputes by offering a third option between "It spread because it was true" (theological or ideological apologetics) and "It spread because elites engineered it" (pure instrumentalism). A Leviathan-class phenomenon can be neither purely engineered nor purely rational. It can be a participatory system whose recursion depth produces durability and governance entanglement.

### 9.2 For Political Theory and Governance

The model highlights a recurring governance problem: Some systems cannot be governed only from the outside, because they encode the inside—identity, obligation, ultimacy—into their participants.

Governance must then choose suppression, containment, capture, or accommodation. Each has predictable failure modes. Suppression often backfires (T3), increasing commitment and martyrdom narratives. Containment requires institutional capacity and may fail if the movement captures institutions (P3). Capture by the movement can lead to authoritarian consolidation. Accommodation risks legitimizing dangerous dynamics, (**Fig.4**).



**Figure 4:** Political Implications of the Global Diffusion of Contemporary Evil Information

### 9.3 Practical Countermeasures: Indicative Policy Levers

Based on the model's indicator structure, countermeasures can target specific dynamics.

First, early-warning systems can use T indicators and governance response shifts as monitoring tools for fascist/authoritarian fields before consolidation and can deploy online protocols (Section 8.5) for real-time detection of allegiance formation.

Second, tailored containment strategies can design legal/constitutional containment depending on emergent versus constructed versus hybrid type. For emergent systems, the focus should be disrupting SE recursion through formation environments and ritual density. For constructed systems, the focus should be targeting P indicators such as elite bargains and propaganda capacity. For hybrid systems, the focus should be multi-level intervention addressing consciousness, power, and institutions.

Third, de-propagation and cognitive resilience efforts can align media, education, and platform policy with disrupting SE and P indicators rather than only suppressing surface content. They can strengthen cognitive autonomy through critical thinking education, ambiguity tolerance training, and developmental supports [28][29]. They can counter memetic hijacking with "pre-bunking" (inoculation against manipulation techniques) rather than "de-bunking" (which can trigger backfire) [30].

Fourth, counter-field formation can invest in pluralism-supporting participation and formation environments that raise the threshold for "evil information" to become Leviathan-class. It can support civil-society ecologies that encode democratic norms, epistemic humility, and vulnerability protection [31]. It can create ritual and symbolic density around democratic values through civic rituals and shared narratives of inclusion.

These levers operationalize the consciousness-memetics-allegiance pipeline: intervene at individual cognition (resilience), collective transmission (de-propagation), formation environments (counter-fields), and institutional response (tailored containment).

## 9.4 For Contemporary Systems: AI as Potential Leviathan-Class Infrastructure

A key extension is that Leviathan dynamics are not limited to ideology or religion. They can occur in socio-technical infrastructures that recruit attention, shape identity, and displace judgment.

AI systems increasingly function as participatory environments: they shape attention (recommender algorithms), decision-making routines (AI assistants), and epistemic authority (LLMs as information sources) [32][33]. This makes them plausible candidates for Leviathan-class analysis.

The diagnostic question becomes: Do AI-driven infrastructures produce threshold dynamics in which people and institutions cannot easily disengage, and governance must escalate?

Regarding indicators, T indicators include nonlinear adoption, institutional lock-in (T2), suppression backfire when regulation is attempted (T3), and governance escalation through AI safety summits, executive orders, and international treaties (T5). Indicators include affective enthusiasm among users (A1), identity fusion with AI tools through "my AI" and anthropomorphizing (A2), instrumental dependence (A3), and growing costly allegiance through job retraining costs and cognitive offloading (A5). SE indicators include ritual density (daily AI use), ethical recursion (AI-mediated decision norms), ultimacy encoding ("AI determines truth"), and boundary mechanisms (AI haves/have-nots). P indicators include institutional capture by AI firms (P3), elite bargains between Big Tech and governments (P4), and administrative enforcement through automated governance (P5).

The "Humanity-Dominated (Human-Guided) AI Development" (HD-AI) framework proposed by Meijer et al. (2026) [34] can be interpreted as a preventive containment strategy: treating cognitive autonomy and human consciousness development as first-order constraints on AI capability scaling, preventing threshold crossing into Leviathan-class entanglement.

Annex E develops this extension in detail.

# 10. Limitations and Future Research

## 10.1 Risk of Reification

"Leviathan" can be misread as a hidden entity rather than a pattern label. The specification approach aims to prevent this: the model stands or falls by indicators and disconfirmation criteria.

## 10.2 Measurement Challenges

Some indicators are easier to measure than others. Governance response (documented in laws, decrees, institutional records) and nonlinear growth (demographic data) are relatively easy. Power mechanics (requiring investigation of elite networks, propaganda infrastructure) are moderately

difficult. Formation regimes and symbolic-ethical recursion (requiring qualitative coding, ethnography, interview data) are most difficult.

Future work should develop standardized coding protocols with inter-rater reliability testing, datasets of comparative cases (positive and negative), and integration with computational social science methods (digital trace data, network analysis).

### 10.3 Multi-Field Ecologies

The model must handle plurality: competing sub-movements, schisms, absorptions, hybridization. Most "traditions" are field ecologies, not monolithic entities. Disaggregation is essential.

### 10.4 Integration with Developmental Frameworks

A promising operational direction involves integrating empirical measures of meaning-making capacity and ego development (Cook-Greuter, Kegan, Loevinger frameworks) as proxies for identity fusion susceptibility, tolerance of ambiguity, and resistance to totalizing allegiance. This is proposed as a research pathway, not assumed as proof.

### 10.5 Cross-Cultural Validity

The model has been developed primarily with Western historical and contemporary cases. Testing in non-Western contexts (East Asian, African, Latin American movements) is essential for validity claims.

## 11. Memetics: A Theory for Evil Information Spreading

### How Do You Meme?

A meme is a unit of cultural information that spreads from person to person through imitation, communication, or other forms of social learning [40][41]. The term was coined by evolutionary biologist Richard Dawkins in his 1976 book *The Selfish Gene*, where he proposed that cultural ideas, behaviors, and practices could replicate and evolve in a manner analogous to genes in biological evolution. Apart from Energy and Matter, Information is a third fundamental building block of the Universe [35]; [36].

Memes encompass a wide range of cultural elements including ideas and beliefs (religious doctrines, scientific theories, philosophical concepts), behaviors and practices (rituals, customs, fashion trends), skills and techniques (cooking methods, artistic styles, technological innovations), and symbols and language (words, phrases, icons, gestures). Key characteristics of memes include replication (memes spread through communication and imitation), variation (as memes are transmitted, they may be altered or adapted), and selection (some memes are more successful at spreading and persisting than others, depending on factors like memorability, emotional impact, or practical utility), [40], [41].

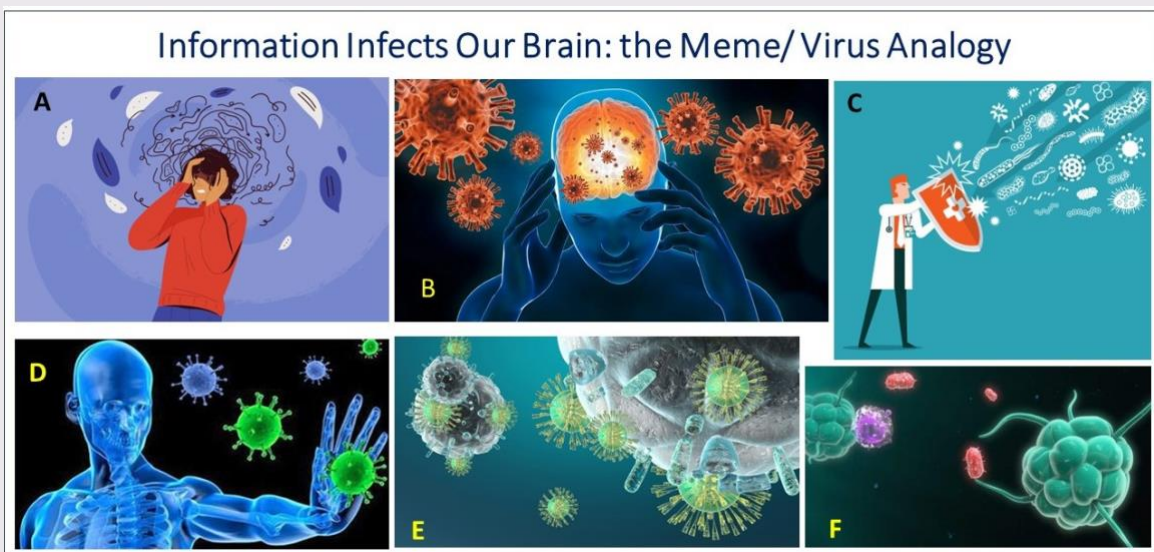
Examples of memes range from simple cultural elements like catchy tunes or popular phrases to complex systems like political ideologies or religious traditions. In recent years, the term "meme" has become popularly associated with internet memes: viral images, videos, or text that spread rapidly online, often with humorous or satirical content [37][38].

Memetics is closely related to the broader field of cultural evolution, which studies how human culture changes over time through processes analogous to biological evolution [39]. Memes can thus

be seen as cultural replicators that compete for attention and transmission within the "ecosystem" of human minds and social networks. This framework has been applied to understand phenomena such as the spread of religions, the evolution of languages, the dynamics of social movements, and the viral nature of online content [40][42;44].

The concept of memetics helps explain how cultural information propagates, transforms, and sometimes becomes "viral" in society, much like how genes propagate and evolve in the biological realm [43]. The meme model was found to be relevant for a better understanding of the manner of the spreading of evil information that is occurring in human societies [36;37;38][45].

The spreading of information, particularly misinformation and disinformation, can be analyzed through the lens of human consciousness and memetic theory, a framework that examines how cultural ideas, behaviors, and practices (memes) replicate and spread through society, much like genes in biological evolution [46]. This perspective offers valuable insights into why certain information—especially emotionally charged or "evil" content—can spread so rapidly and persistently, Fig.5 (see [40;41] ).



**Figure 5:** *A: Multiple Memes Can undergo mutation and selection to strong variants; B: Memes may survive through competition in a Darwinian mechanism and attain stable expression in our brain C: It is preferable to build up a versatile fire wall to prevent infection D: Personal attitudes can help*

Human consciousness serves as both the medium and the battleground for memetic transmission [45;45;47]. Several characteristics of human cognition make us particularly susceptible to viral information. Emotional resonance means that information triggering strong emotions (fear, anger, disgust, or tribal identification) captures attention and is more likely to be remembered and shared. Cognitive biases including confirmation bias, in-group favoritism, and pattern-seeking tendencies make humans receptive to information that confirms existing beliefs or simplifies complex realities. Social identity leads people to share and defend information that signals group membership and reinforces social bonds. Narrative processing means the human mind is wired for stories, and information packaged in compelling narratives spreads more effectively than raw data [48][49].

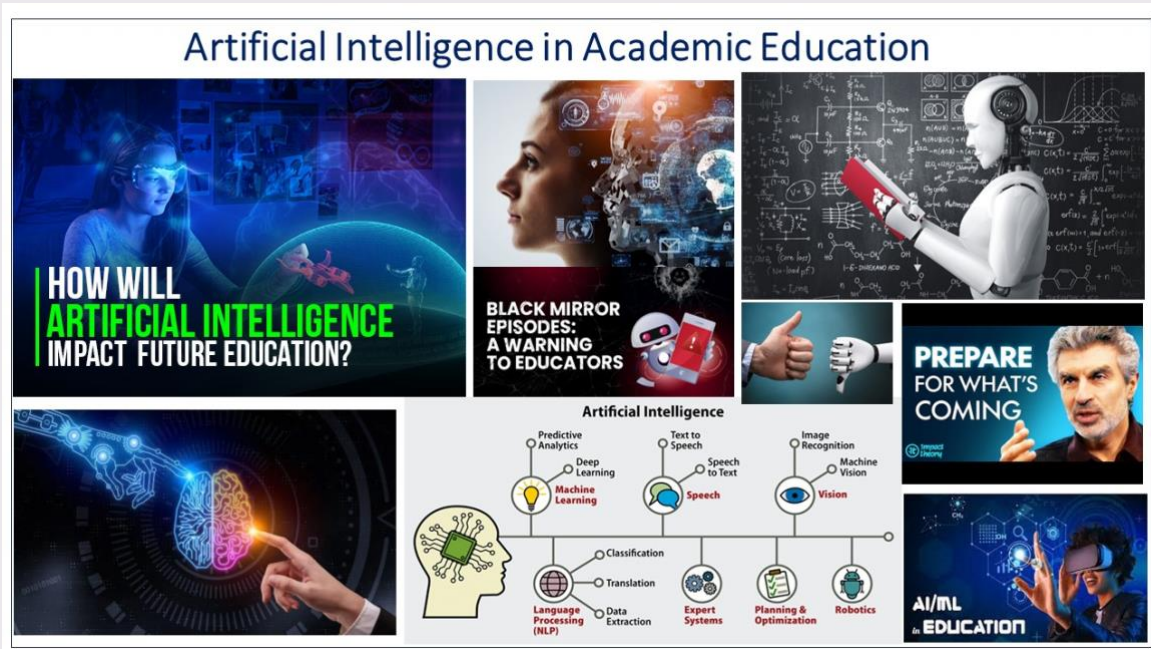


**Figure 6:** Examples of Political Memes, Displayed in Pictures and Meant to Influence or Mislead Voters. This Type of Meme May Invite a Host of Counter Measures, Installed in a Digital Immune System Behind the Scenes

These consciousness-level vulnerabilities interface with memetic dynamics to create powerful amplification loops. Variation occurs as information mutates as it spreads, becoming more emotionally potent or "sticky" through successive retellings, (Fig.6). Selection means memes that evoke stronger emotional responses, or better fit existing cognitive biases outcompete more nuanced or accurate information. Replication occurs as social media and digital networks provide unprecedented replication speed and reach. Recombination happens when memes combine and recombine, creating novel variants that may be even more virulent.[40;41]

These memetic mechanisms integrate directly with the Leviathan framework developed in this paper. The consciousness-memetics-allegiance pipeline (Section 2.3) describes how individual cognitive vulnerabilities scale into collective allegiance formation. The symbolic-ethical recursion indicators (SE1–SE5) capture the formation dynamics through which memetic content becomes encoded into identity and obligation. The suppression backfire phenomenon (T3) reflects the memetic principle that attempted suppression often increases transmission salience.

Understanding these memetic dynamics is essential for designing effective countermeasures (Section 9.3, Fig.6),[39; 40] Rather than simply suppressing "evil information" after it spreads—which often triggers backfire—interventions should target the consciousness-level vulnerabilities that make minds susceptible to memetic hijacking in the first place. This is the rationale for emphasizing cognitive resilience, pre-bunking, and counter-field formation alongside traditional content moderation and legal containment strategies. Even the chronic use of AI programs may lead to dependence and addition, since knowledge acquisition perse, can feel as extremely rewarding[39]. One major calamity is the rapidly growing overproduction of scientific papers, generated to easily with AI, by which current science is flooded with reports, often not even read by the authors. This may even corrupt the very art of academic activity: asking the right question, collecting adequate data and presenting them in a personal analysis, Fig.7).



**Figure 7:** The impact of AI in Various Modalities (middle below) in Education: Novel Chances for Efficient Knowledge Acquisition and Innovation but Beware of What is Coming!

## 12. Conclusion

The Leviathan Hypothesis (v3) proposes a testable claim: some movements and participatory systems cross a threshold into self-reinforcing allegiance such that external governance cannot manage them through ordinary means and must shift into containment, capture, or systemic reconfiguration.

The contribution is specification: operational definitions, indicator families (T, A, SE, P), typology (emergent/constructed/hybrid), falsifiability criteria, strict-mode scoring sheet, and an empirical protocol for real-time testing.

The model is intentionally contestable. It fails if comparable movements show the same indicator constellation without requiring "Leviathan" dynamics, if outcomes can be fully reduced to standard diffusion, incentives, or coercion, if symbolic-ethical recursion proves epiphenomenal, or if negative cases cannot be identified.

Contemporary urgency makes this more than an academic exercise. Authoritarian and fascist movements demonstrate the pattern across multiple democracies. Understanding Leviathan dynamics—and designing countermeasures that address consciousness, memetics, formation environments, and power mechanics—is not merely academic. It is a governance imperative.

The model's value lies not in metaphorical force, but in generating comparative research questions, enabling early-warning systems, and informing practical interventions across religion, ideology, social movements, and socio-technical infrastructures.

## 12. References

- [1] Levitsky, S., & Ziblatt, D. (2018). *How Democracies Die*. Crown.
- [2] Stanley, J. (2018). *How Fascism Works: The Politics of Us and Them*. Random House.
- [3] Sunstein, C. R. (2017). *Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
- [4] Tucker, J. A., et al. (2018). Social media, political polarization, and political disinformation. *The ANNALS of the American Academy of Political and Social Science*, 681(1), 44-77.
- [5] Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.
- [6] Centola, D. (2018). *How Behavior Spreads: The Science of Complex Contagions*. Princeton University Press.
- [7] Weber, M. (1922/1978). *Economy and Society*. University of California Press.
- [8] Durkheim, É. (1912/1995). *The Elementary Forms of Religious Life*. Free Press.
- [9] Le Bon, G. (1895/2002). *The Crowd: A Study of the Popular Mind*. Dover.
- [10] Arendt, H. (1951). *The Origins of Totalitarianism*. Harcourt.
- [11] Lifton, R. J. (1961/1989). *Thought Reform and the Psychology of Totalism*. University of North Carolina Press.
- [12] Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- [13] Blackmore, S. (1999). *The Meme Machine*. Oxford University Press.
- [14] Swann, W. B., Jr., et al. (2012). When group membership gets personal. *Psychological Review*, 119(3), 441-456.
- [15] Whitehouse, H., et al. (2017). The evolution of extreme cooperation via shared dysphoric experiences. *Scientific Reports*, 7, 44292.
- [16] McCauley, C., & Moskaleiko, S. (2017). Understanding political radicalization. *American Psychologist*, 72(3), 205-216.
- [17] Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.
- [18] Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online*. Data & Society Research Institute.
- [19] Muirhead, R., & Rosenblum, N. L. (2019). *A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy*. Princeton University Press.
- [20] Cook-Greuter, S. R. (2005). *Ego development: Nine levels of increasing embrace*. Unpublished manuscript.

- [21] Kegan, R. (1982). *The Evolving Self: Problem and Process in Human Development*. Harvard University Press.
- [22] Zuboff, S. (2019). *The Age of Surveillance Capitalism*. PublicAffairs.
- [23] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- [24] Mudde, C., & Kaltwasser, C. R. (2017). *Populism: A Very Short Introduction*. Oxford University Press.
- [25] Snyder, T. (2017). *On Tyranny: Twenty Lessons from the Twentieth Century*. Tim Duggan Books.
- [26] Applebaum, A. (2020). *Twilight of Democracy: The Seductive Lure of Authoritarianism*. Doubleday.
- [27] Dobson, R. (2025). Protocol for Coding Emergent Allegiance in Online Debates: Testing the Leviathan Hypothesis in Real Time. <https://www.academia.edu/146215283/>
- [28] Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *PNAS*, 116(7), 2521-2526.
- [29] Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5, 65.
- [30] Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2), 348-384.
- [31] Dobson, R. (2025). *Vulnerability, Pluralism, Universal Ethics for a Fractured World (Revised Edition)*. [https://www.academia.edu/145253478/Vulnerability Pluralism Universal Ethics for a Fractured World Revised Edition](https://www.academia.edu/145253478/Vulnerability_Pluralism_Universal_Ethics_for_a_Fractured_World_Revised_Edition)
- [32] Bender, E. M., et al. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of Fact 2021*, 610-623.
- [33] Weidinger, L., et al. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.
- [34] Meijer, D. K. F., de Leeuw, G., Keizer, P., & Dobson, R. (2026). Beyond the Current AI Hype: Towards a Human-Guided AI Evolution by Fostering Human Consciousness. In preparation
- [35] Meijer D K F (2012). The Information Universe. On the Missing Link in Concepts on the Architecture of Reality. *Syntropy Journal*, 1, pp 1-64. <http://www.sintropia.it/english/2012-eng-1-1.pdf>
- [36] Meijer D K F (2013). Information: What Do You Mean? *Syntropy Journal*, 2013 (3), pp 1-49. <http://www.lifeenergyscience.it/english/2013-eng-3-01.pdf>
- [37] Meijer DKF, (2015). The Universe as a Cyclic Organized Information System. An Essay on the Worldview of John Wheeler. *NeuroQuantology*, 1, pp 1-40, <http://www.neuroquantology.com/index.php/journal/article/view/798/693>

[38] Meijer D K F, (2017). The Processes of Science and Art Modeled by Toroidal Flow of Information. Open Journal of Philosophy, 8, 365-400. doi: [10.4236/ojpp.2018.84026](https://doi.org/10.4236/ojpp.2018.84026).

<https://www.scirp.org/journal/PaperInformation.aspx?PaperID=86591>

[39] Meijer D K F, 2024. Everything Is Said, but Nothing Has Been Told. On the Current State of Art of Science and Academic Education: Problems and Perspectives.

[https://www.researchgate.net/publication/377151629\\_Everything\\_Is\\_Said\\_but\\_Nothing\\_Has\\_Been\\_Told\\_On\\_the\\_Current\\_State\\_of\\_Art\\_of\\_Science\\_and\\_Academic\\_Education\\_Problems\\_and\\_Perspectives](https://www.researchgate.net/publication/377151629_Everything_Is_Said_but_Nothing_Has_Been_Told_On_the_Current_State_of_Art_of_Science_and_Academic_Education_Problems_and_Perspectives)

[40] Meijer D K F, 2024a. On the Internet Meme/Virus Analogy: Part 1. Can We Prevent Contagious Information that Infects Our Sub-Conscious? A Plea for a Versatile Immune System for the Internet in the Present AI – Era. **(21) (PDF) On the Internet Meme/Virus Analogy: Part 1. Can We Prevent Contagious Information that Infects Our Sub-Conscious? A Plea for a Versatile Immune System for the Internet in the Present AI -Era (researchgate.net)**

[41] Meijer, 2024b. On the Internet Meme/Virus Analogy, Part 2. From Meme to Medicine: Imaging Current Drug Design and Therapeutics.

[https://www.researchgate.net/publication/380792348\\_On\\_the\\_Internet\\_MemeVirus\\_Analogy\\_Part\\_2\\_From\\_Meme\\_to\\_Medicine\\_Imaging\\_Current\\_Drug\\_Design\\_and\\_Therapeutics](https://www.researchgate.net/publication/380792348_On_the_Internet_MemeVirus_Analogy_Part_2_From_Meme_to_Medicine_Imaging_Current_Drug_Design_and_Therapeutics)

[42] Meijer D K F, 2025 a. Universal Spectrum of Self-Transcendent Mystical Experiences as Transformative Psi- Phenomena, Part 1: The Relation with Universal Consciousness and Sonic Coherence.

[https://www.academia.edu/128936840/Universal\\_Spectrum\\_of\\_Self\\_Transcendent\\_Mystical\\_Experiences\\_as\\_Transformative\\_Psi\\_Phenomena\\_Part\\_1\\_The\\_Relation\\_with\\_Universal\\_Consciousness\\_and\\_Sonic\\_Coherence](https://www.academia.edu/128936840/Universal_Spectrum_of_Self_Transcendent_Mystical_Experiences_as_Transformative_Psi_Phenomena_Part_1_The_Relation_with_Universal_Consciousness_and_Sonic_Coherence)

[43] Meijer D.K.F, 2025b. Universal Spectrum of Self-Transcendent Mystical Experiences as Transformative Psi-Phenomena, Part 2: Potential Healing Role in the Future of Mankind and our Planetary Life. **(99+) Universal Spectrum of Self-Transcendent Mystical Experiences as Transformative Psi-Phenomena, Part 2: Potential Healing Role in the Future of Mankind and our Planetary Life**

[44] Meijer D K F, Kieft W, 2025. The Role of Humanity in a Self-Learning Universe: A Musical Space Journey to Novel Horizons in the Fabric of Reality. **(99+) The Role of Humanity in a Self-Learning Universe: A Musical Space Journey to Novel Horizons in the Fabric of Reality. An Essay for All People Interested in Life Sciences, Including Non-Scientists**

[45] Meijer D K F and M D Forghani, 2025 The Sonic and Conscious Universe: Transactional Wave Resonance Creates Coherence from Planck to Cosmic Scales, as a Basis for Human Intelligence and Unified Cosmo-Psychism. **(99+) The Sonic and Conscious Universe: Transactional Wave Resonance Creates Coherence from Planck to Cosmic Scales, as a Basis for Human Intelligence and Unified Cosmo-Psychism**

[46] Ott R and D K F Meijer, (2025). Scale-Invariant Unifying Resonant Fields of Physics, AI and Consciousness. **(99+) Scale-Invariant Unifying Resonant Fields of Physics, AI and Consciousness**

[47] Dobson, R. (2025). Vulnerability, Pluralism, Universal Ethics for a Fractured World (Revised Edition).

[48] Dobson, R. (2025). Before the Machine Decides.

[https://www.academia.edu/144940132/Before the Machine Decides](https://www.academia.edu/144940132/Before_the_Machine_Decides)

[49] Dobson, R. (2025). Layered Intelligence & Logic in Reality: Toward Transformative Leadership and Collaboration.

[https://www.academia.edu/127795672/Layered Intelligence and Logic in Reality Toward Transformative Leadership and Collaboration](https://www.academia.edu/127795672/Layered_Intelligence_and_Logic_in_Reality_Toward_Transformative_Leadership_and_Collaboration)

---

## Annexes

### Annex A: Strict-Mode Scoring Sheet

Purpose: Structured evaluation tool for applying the Leviathan Hypothesis in "strict mode." Reduces interpretive drift, prevents post-hoc fitting, enables comparative falsifiable analysis.

#### Case Specification

Case label: \_\_\_\_\_

Time window: \_\_\_\_\_

Region: \_\_\_\_\_

Unit of analysis: \_\_\_\_\_

Primary data sources: \_\_\_\_\_

#### A1. Threshold Indicators (T)

T1 Nonlinear growth: Score 0 / 1 / 2 (notes: \_\_\_\_)

T2 Local-node replication: Score 0 / 1 / 2

T3 Suppression backfire: Score 0 / 1 / 2

T4 Polarization: Score 0 / 1 / 2

T5 Governance escalation: Score 0 / 1 / 2

#### A2. Allegiance Profile (A)

A1 Affective allegiance: Score 0 / 1 / 2

A2 Symbolic-identity allegiance: Score 0 / 1 / 2

A3 Instrumental allegiance: Score 0 / 1 / 2

A4 Constrained allegiance: Score 0 / 1 / 2

A5 Costly allegiance: Score 0 / 1 / 2

#### A3. Symbolic-Ethical Recursion (SE)

SE1 Ritual density: Score 0 / 1 / 2

SE2 Ethical recursion: Score 0 / 1 / 2

SE3 Ultimacy encoding: Score 0 / 1 / 2

SE4 Boundary mechanisms: Score 0 / 1 / 2

SE5 Formation regime: Score 0 / 1 / 2

#### **A4. Power Mechanics (P)**

P1 Coercion capacity: Score 0 / 1 / 2  
P2 Propaganda capacity: Score 0 / 1 / 2  
P3 Institutional capture: Score 0 / 1 / 2  
P4 Elite bargains: Score 0 / 1 / 2  
P5 Administrative enforcement: Score 0 / 1 / 2

#### **A5. Governance Response Mode**

Select one: Ordinary management / Suppression attempt / Containment strategy / Capture-institutional fusion / Accommodation / Fragmentation management

#### **A6. Disconfirmation Notes**

What evidence would most strongly disconfirm Leviathan dynamics here?

#### **Interpreting Results**

Leviathan-class: Multiple threshold indicators plus meaningful costly allegiance plus symbolic-ethical recursion plus governance escalation, not reducible to power alone.

Borderline/hybrid: Some indicators present, but dependence on coercion or lack of durability remains unresolved.

Not Leviathan-class: Growth or intensity without threshold behavior, internalization, or governance disruption.

## **Annex B: Glossary**

**Allegiance:** Durable commitment combined with identity combined with obligation.

**Costly allegiance:** Commitment that persists under real risk or loss.

**Participation:** Repeated engagement that forms perception and action.

**Field (methodological):** A distributed system of reinforcement shaping behavior; not physics.

**Threshold:** Point where feedback loops dominate; non-linear dynamics emerge.

**Symbolic-ethical recursion:** Meaning leading to identity leading to obligation leading back to meaning loop.

**Formation regime:** Training practices that stabilize behavior under pressure.

**Governance escalation:** Shift from ordinary management to special measures.

**Containment:** Structural stabilization of a phenomenon via doctrine/law/institution.

**Capture:** Movement takes over institutions or becomes institutional infrastructure.

**Constructed:** Engineered via coercion/propaganda; top-down dominated.

**Emergent:** Bottom-up self-reinforcing replication through local nodes.

**Hybrid:** Feedback between emergence and construction.

**Leading indicators:** Early signals (identity charge, ultimacy framing, exit-and-return) that precede overt governance disruption.

**Lagging indicators:** Confirmatory signals (governance escalation, legal exceptionalism, institutional capture) that follow threshold crossing.

## **Annex C: Genesis of the Hypothesis (v1 → v2 → v3)**

The transition from v1 to v2 addressed circularity concerns, added unit-of-analysis specification, introduced the emergence-construction axis, centered governance response, and added falsifiability criteria and scoring sheet.

The transition from v2 to v3 addressed reviewer feedback (Meijer 2026) through early explanation of "Leviathan" term (Sections 1.2, Annex D), explicit focus on contemporary fascism as test case (Sections 2.2, 6.4, 8.2), integration of consciousness-memetics-allegiance pipeline (Section 2.3), practical countermeasures section (Section 9.3), empirical protocol for online testing (Section 8.5), memetics section on evil information spreading (Section 11), and leading versus lagging indicator distinction (Section 5.6).

## **Annex D: Symbolic Antecedents—Why "Leviathan"?**

The term "Leviathan" has three relevant historical uses. In ancient Near Eastern and biblical usage, it appears as an untamable sea creature representing chaotic forces beyond human control (Job 41:1-34: "Can you pull in Leviathan with a fishhook?... Any hope of subduing it is false"). In Hobbesian political theory, it describes the sovereign state as a constructed entity transcending individual wills, necessary to prevent the war of all against all (Hobbes 1651). In this paper, it serves as a structural label for movements or systems that exceed ordinary governance tools and force institutional escalation.

The model does not depend on biblical or Hobbesian metaphysics. The term could be replaced with "threshold allegiance system" or "ungovernable participatory dynamic." But "Leviathan" captures the encounter with emergent reality that resists external control—a phenomenon governance must confront but cannot easily tame.

This annex is optional reading. Readers concerned with strict historical sociology may omit it.

## **Annex E: Extension to Socio-Technical Systems—AI as Potential Leviathan-Class Infrastructure**

Section 9.4 introduced AI as a potential Leviathan-class infrastructure. This annex develops the analysis in greater detail.

AI systems increasingly operate as participatory environments that shape attention through recommender algorithms, decision-making routines through AI assistants, and epistemic authority through LLMs as information sources. This makes them plausible candidates for Leviathan-class analysis.

Practical implications for AI governance follow from the countermeasure framework in Section 9.3. For early-warning, monitor T indicators (adoption acceleration, institutional lock-in, governance escalation) before irreversible threshold crossing. For containment design, implement the HD-AI framework as preventive containment—cognitive autonomy as first-order constraint on capability scaling. For de-propagation, regulate algorithmic amplification (SE/P disruption) and require epistemic humility in AI systems ("I don't know" is better than confident hallucination). For counter-field formation, invest in education that strengthens critical AI literacy, reduces dependency, and maintains human decision-making authority.

## Disclaimer

The Leviathan Hypothesis is an academic diagnostic framework offered for scholarly analysis and debate. It is movement-agnostic, politically neutral, and carries no theological or conspiratorial implications. The term "Leviathan" is a mnemonic shorthand interchangeable with "threshold allegiance system." This framework is a testable proposition, not a settled conclusion; the authors invite disconfirmation and disclaim responsibility for bad-faith applications or misuse. References to contemporary movements are illustrative, not accusatory, and do not constitute claims about the legality, morality, or intentions of any specific group or individual. This work represents the authors' independent scholarly opinion and not the position of any government, institution, or third party.

---

### Correspondence:

Richard Dobson, [Richard@clarafutura-andorra.world](mailto:Richard@clarafutura-andorra.world)

Dirk K.F. Meijer, [meij6076@planet.nl](mailto:meij6076@planet.nl)

The Leviathan Hypothesis is developed under Clara Futura World and RINHUMAI as part of ongoing research on governance, vulnerability, and the protection of human autonomy in an age of emergent systems.