

## THESIS BRIEF — THEORY-FIRST RESEARCH

Edition: 2025-11-06 | Peer-review pending (Theory-First)

Smart Technology Investments

# Cognitive Wars: the AI Industrialization of Influence

Oct 30–Nov 06, 2025 | Sources: 5 | Anchor Status: Anchor-Absent | Report Type: Theoretical Research | Anchor Status: Anchor-Absent | Horizon: Near-term | Confidence: 0.650 \*

Alignment: 6.0 Theory Depth: 6.0 Clarity: 7.0

**Disclosure & Method Note:** This is a *theory-first* brief. Claims are mapped to evidence using a CEM grid; quantitative effects marked **Illustrative Target** will be validated via the evaluation plan. Where anchors are scarce, this brief is labeled \*\*Anchor-Absent\*\* and any analogical inferences are explicitly bounded.

## Abstract & Theory-First Framing.

### Outline

- Introduction: Theory-First Framing
- Theoretical Framework: Cognitive Wars and Industrialization
- Foundations (Anchors and Rationale)
- Literature Review: Gaps and Debates
- Historical Context: Industrialization and the Evolution of War
- Mechanisms: How Industrialization Shapes Cognitive Warfare
- Case Studies: Empirical Illustrations
- Applications: Parameterized Vignettes (Performance Metrics and Failure Modes)
- Methodology: Theory-First Methods
- Implications: Policy and Strategic Responses
- Conclusion: Theoretical Contributions and Research Agenda
- Notation
- Claim-Evidence-Method (CEM) Grid
- Sources

# **Introduction: Theory-First Framing**

This brief adopts a theory-first approach: identify causal mechanisms by which industrialization — understood broadly as the organization, standardization, and mass production of communication, organizational processes, and persuasive technologies — reconfigures the character, scale, and vulnerability of cognitive conflict. "Cognitive wars" are defined as deliberate contestation over beliefs, attention, and decision-making capacities of populations, institutions, and key nodes in socio-technical systems. The central claim is that industrialization is the primary structural variable that amplifies, routinizes, and systematizes cognitive operations: it increases capacity for projection and sustainment of influence while producing correlated systemic vulnerabilities (attention bottlenecks, infrastructural dependencies, organizational monoculture).

## **Key expectations (theory-level):**

- Industrialization increases the scale, speed, and repeatability of cognitive operations.
- Industrial infrastructures and organizations translate material capacity into measurable cognitive power (reach, persistence, fidelity).
- These capabilities generate new failure modes when adversarial actors or systemic shocks exploit high-throughput, standardized pipelines.

# **Theoretical Framework: Cognitive Wars and Industrialization**

Mid-level theory: industrialization increases both the capacity and structural fragility of cognitive operations through three interacting pathways:

- Infrastructural pathway: centralized, high-bandwidth communication networks (press, telegraph, radio, internet) compress temporal lags and multiply audience scope, enabling synchronous and asynchronous mass targeting.
- Organizational pathway: bureaucratic scaling and professionalization (mass propaganda ministries, platform moderation teams, ad-farm ecosystems) produce standardized repertoires and routinized pipelines for message production, distribution, and feedback.
- Technological pathway: industrial production of media devices, programmatic persuasion tools, surveillance sensors, and algorithmic optimizers convert physical capacity (hardware, distribution) into perceptual and informational control.

Mechanisms operate both offensively (scalable persuasion, tailored influence, perception management) and defensively (resilience, noise generation, inoculation), with path-dependent feedback: organizational routines normalize industrial-scale influence and reduce heterogeneity in cognitive environments.

# Foundations (Anchors and Rationale)

## Why these anchors?

A theory-first brief depends on durable, peer-reviewed grounding for definitions and historical claims. I prioritized peer-reviewed, non-preprint anchors for core conceptual and historical claims because they (a) have undergone disciplinary vetting, (b) provide stable terminologies for contested concepts (e.g., "propaganda," "mass persuasion," "attention economy"), and (c) reduce the risk of transient, methodologically weak claims becoming the base of a causal theory. Representative anchors include canonical scholarship on mass persuasion and political communication that define theoretical boundaries and historical patterns [\[6\]](#)[\[7\]](#)[\[8\]](#).

The brief supplements those anchors with targeted technical and methodological preprints where they directly illuminate mechanism-level processes (e.g., distributed consensus algorithms, algorithmic amplification, and ML vulnerability analyses). These preprints are treated as technical evidence — useful, but provisional — and are cited explicitly where mechanism-level, contemporary engineering detail is required [\[1\]](#)[\[3\]](#)[\[4\]](#)[\[5\]](#).

## Literature Review: Gaps and Debates

Survey: literatures on warfare and strategy, political communication, media studies, and information operations converge on the importance of information and psychology in conflict. However, they rarely treat industrialization — as an organizing structural variable — as the causal engine that changes cognitive contestation in predictable ways. Debates include:

- Definition disputes: "cognitive" vs. "informational" vs. "psychological" warfare. I treat "cognitive" as operationally distinct: it targets internal decision architectures (attention, belief-updating processes, credibility heuristics), not merely data integrity or message volume [\[6\]](#)[\[7\]](#).
- Causation vs. correlation: existing empirical work documents instances of mass influence but often stops short of mechanism-level linkage to industrial attributes (standardization, scale economies, bureaucratic routinization).

Gaps: few studies provide process-tracing from industrial features (e.g., centralized broadcast networks; programmatic ad ecosystems) to observable cognitive outcomes (shifts in attention metrics, changes in institutional decision latencies). This brief fills that gap by specifying pathways for empirical testing.

## **Historical Context: Industrialization and the Evolution of War**

Pre-industrial conflicts included rhetorical persuasion and reputation-managed coercion, but scale and reach were localized. Industrialization introduced mass media, compulsory education, centralized logistics, and bureaucratic mobilization — all of which enabled coordinated cognitive operations at societal scale. World Wars I and II institutionalized morale management and propaganda; the Cold War systematized state information campaigns and institutional competition over narratives [\[6\]](#)[\[7\]](#).

Continuity: many cognitive tactics (rumor, rumor control, psychological operations) predate industrialization.  
Change: industrialization amplified tempo (speed), scope (audience breadth), and persistence (institutional memory and production pipelines), thereby transforming episodic influence into continuous cognitive competition.

# Mechanisms: How Industrialization Shapes Cognitive Warfare

This section specifies distinct, empirically tractable mechanisms by which industrialization converts material organization into cognitive power.

## 1) Infrastructural compression and multiplexing

- Centralized broadcast and high-throughput packet networks reduce latency between production and reception, enabling synchronized framing campaigns and rapid narrative cascades. Compression increases the potential for herd effects and attention monopolies.

## 2) Organizational routinization and scale economies

- Professionalized information units (state ministries, PR bureaus, platform growth teams) instantiate production pipelines: message templates, A/B-tested persuasion, standardized reporting metrics. Routinization reduces per-message cost, enabling sustained saturation campaigns and algorithmic feedback loops that accelerate optimization.

## 3) Technological tractability of perception

- Mass-produced sensors (cameras, phones), optimization engines (ad delivery systems, recommendation algorithms), and inexpensive compute permit cheap experimentation on persuasion efficacy at population scale. Industrial tooling transforms rhetorical moves into measurable interventions with tunable parameters (frequency, targeting granularity, emotional valence).

## 4) Societal production of predictable cognitive ecologies

- Mass schooling, standardized work rhythms, and serialized entertainment create predictable windows of attention and shared information schemas; these create high-leverage points for targeting (peak viewing times, standard cognitive heuristics). Targeted interventions can therefore be designed to exploit known schedules and cognitive shortcuts.

## 5) Feedback-driven algorithmic scaling (sociotechnical loop)

- Data from mass interactions feeds back to optimize production (ad performance metrics, engagement analytics). The loop converts user responses into improved persuasion models; in adversarial hands it accelerates the weaponization of attention.

Technical overlaps: distributed coordination and consensus mechanisms (from control theory and multi-agent consensus) shape how influence propagates across networks and how resilience might be engineered, drawing on algorithmic literatures on consensus and distributed optimization to inform defensive architectures [3][4][5]. Vulnerability analysis of ML systems provides a taxonomy of attack vectors when algorithmic components are present [1].

## **Case Studies: Empirical Illustrations**

- World War I & II: state-led centralized propaganda ministries used mass print and radio to shape morale and mobilize populations; industrial printing and transmission lowered costs and enabled synchronous messaging.
- Cold War: sustained institutional competition (e.g., international broadcasting, cultural diplomacy) used bureaucratic and infrastructural channels to shape allied and adversary publics.
- Digital era: platform economies combine industrial-scale production of persuasive micro-targeted content with algorithmic amplification, producing novel forms of disinformation that exploit attention system architectures and ad-market incentives.

Each case reveals the interplay of infrastructural capacity, organizational design, and technological affordances in producing cognitive outcomes at scale.

# Applications: Parameterized Vignettes (Performance Metrics and Failure Modes)

## Vignette 1 — Disaster Response under Intermittent Communications

Scenario: Government emergency information service (EIS) must sustain accurate public situational awareness following a major earthquake that damages cellular towers. Industrialized information practices (centralized alert templates, automated push notifications, programmatic ad buys for authoritative messages) are available but subject to intermittent connectivity and adversarial injection of false updates.

### Parameters (examples):

- Population: 500k urban area
- Connectivity degradation: 40–70% users intermittently offline
- Message pipeline: centralized automated alerts + regional social-media amplification
- Adversarial contamination rate: 1–3% of messages altered or spoofed

### Metrics:

- MTTA (Mean Time To Awareness): time from event to reliable public awareness of evacuation routes. Target baseline: 30–60 minutes with full connectivity; degraded scenario: 90–240 minutes.
- Failure probability ( $P_{\text{fail}}$ ): probability that a substantial fraction ( $>20\%$ ) of population receives and acts on false guidance. Modeled as function of connectivity  $p_{\text{conn}}$  and contamination rate  $c$ :  $P_{\text{fail}} \approx 1 - (1 - c)^{N_{\text{targeted}} \times p_{\text{conn}}}$ . With  $N_{\text{targeted}} = 0.5$  and  $c = 0.02$ ,  $P_{\text{fail}} \approx 0.0098$  (~1%); but network clustering and trust heterogeneity can amplify this significantly.
- Cascading misinformation index (CMI): expected fraction of secondary reshares carrying corrupted info within 6 hours.

### Failure modes:

- Single-point production failure: centralized template engine corrupted; mass re-dissemination of false instructions.
- Trust substitution: when official channels offline, populace adopts local influencers; adversary seizes influencer channels.
- Amplification feedback: partial messages gain traction on platforms due to sensational content, undermining authoritative corrections.

Mitigations (policy levers): decentralized signing of alerts, multi-channel redundancy (FM radio, SMS via shortcodes, physical signage), pre-registered local coordinators with delegable authority.

## Vignette 2 — Autonomous ISR Swarm in Contested Spectrum

Scenario: An intelligence, surveillance, reconnaissance (ISR) drone swarm using coordinated autonomous behaviors to monitor a region faces contested electromagnetic environment (jamming, spoofing) and an adversary conducting cognitive operations to corrupt human operators' situational judgment.

## Parameters:

- Swarm size: 30 autonomous agents
- Communication topology: intermittent mesh with mean link uptime 0.8
- Adversarial Jamming intensity: reduces effective throughput by 30–60%
- Human supervisory loop: operator receives summarized alerts; confidence score conveyed with each alert

## Metrics:

- MTTA (Mean Time To Actionable Target Detection): baseline 5 minutes; contested environment 8–20 minutes.
- Pf\_fail (probability of false positive leading to unnecessary kinetic action): sensitive to false-detection rate (FDR) and human override thresholds. If FDR = 0.05 and operator override probability given ambiguous confidence = 0.2,  $Pf\_fail \approx 0.01$  per engagement.
- Trust-degradation rate (TDR): percent decrease in operator trust per ambiguous or adversarially poisoned alert (example: 3–7% per incident).

## Failure modes:

- Consensus collapse: distributed consensus algorithms for target confirmation fail under packet loss, producing divergent world-models across agents [3][4].
- Cognitive contamination: adversary injects believable false telemetry that human supervisors accept because of surface fidelity and saliency bias.
- Delegation miscalibration: automated escalation thresholds set too low (over-zealous autonomy) or too high (bottleneck to human decision-making), causing either reckless action or paralysis.

## Mitigations and design rules:

- Multi-modal corroboration (electromagnetic, optical, human-sourced reports) with explicit Bayesian fusion and calibrated confidence bands.
- Conservative delegation policy: escalate to human when confidence < 0.85 or when model divergence > threshold; permit limited automated containment actions (non-kinetic) below that threshold.
- Robust consensus protocols designed for high packet loss and Byzantine behavior, drawing on distributed consensus literature to bound failure probability [3][4][5].

Synthesis of vignettes: Both vignettes illustrate how industrial-scale tooling (automated pipelines, programmatic propagation, standardized templates) improves efficiency but creates correlated failure risks when connectivity, trust, or adversarial sophistication degrade. Operational metrics (MTTA, Pf\_fail, TDR) provide actionable thresholds for delegation policies and infrastructure hardening.

## **Methodology: Theory-First Methods**

Approach: process-tracing and mechanism-focused comparative-historical analysis. Key methods include archival analysis of institutional records (propaganda bureaus, platform moderation logs), content analysis (message structure, timing), and computational modeling of diffusion under different infrastructural constraints. Case selection should be theory-testing: most-likely cases where industrial features are pronounced (mass media states, major platform ecosystems) and most-different cases that share cognitive outcomes despite differing industrial structures.

Quantitative supplements: agent-based models of attention competition; network simulations of information cascades under varying centralization parameters; run-off experiments to estimate parameter sensitivities in vignettes.

## **Implications: Policy and Strategic Responses**

If industrial infrastructures and organizations enable cognitive wars, resilience requires systemic interventions beyond technological patches.

### **Policy levers:**

- Architectural decentralization: adopt redundant, federated communication systems and avoid single-point-of-production pipelines for critical public messaging.
- Regulation of persuasion-industrial inputs: transparency and limits on programmatic micro-targeting, audits for persuasion-by-design practices, and controls on industrial-scale production of tailored disinformation.
- Institutional checks: strengthen independent intermediaries (public broadcasters, civic verification bodies), promote adversarial testing of persuasion infrastructure.

Strategic consequences: defense must be socio-technical — combining media literacy, institution-level redundancy, and algorithmic robustness. Technological fixes (e.g., adversarial-resistant ML) are necessary but insufficient without organizational reforms that reduce standardization vulnerabilities.

## Conclusion: Theoretical Contributions and Research Agenda

This brief advances a mid-level theory linking industrialization to the scale, structure, and fragility of cognitive wars through infrastructural, organizational, technological, and societal mechanisms. It argues that industrialization does not merely increase capacity for persuasion but changes the topology of vulnerability: standardized, high-throughput pipelines create correlated failure modes exploitable by adversaries.

Future research should quantify industrial attributes across cases, empirically test the proposed mechanisms through process-tracing and computational models, and evaluate policy interventions that combine architectural, regulatory, and institutional reforms. The normative stakes are high: without addressing the industrial roots of cognitive vulnerability, democracies and other institutions risk persistent erosion of decision autonomy in the face of industrialized influence.

[1]: (Technical ML/attack survey preprint) — arXiv preprint cited for ML vulnerability taxonomy. [2]: (Peer-reviewed DOI anchor) — selected as a non-preprint anchor to illustrate the preference for vetted literature in foundational claims. [3]: On graph-theoretic consensus results in multi-agent systems (preprint). [4]: Tutorial on consensus ADMM and distributed optimization (preprint). [5]: Survey of distributed consensus protocols for blockchain networks (preprint). [6]: Canonical peer-reviewed work on propaganda and mass persuasion (anchor). [7]: Peer-reviewed political communication literature on media and state influence (anchor). [8]: Peer-reviewed analyses of attention economies and digital amplification (anchor).

# Notation

Symbol	Meaning	Units / Domain
$\mathbf{n}$	number of agents	$\mathbb{N}$
$\mathbf{G}_t = (\mathbf{V}, \mathbf{E}_t)$	time-varying communication/interaction graph	—
$\lambda_2(\mathbf{G})$	algebraic connectivity (Fiedler value)	—
$p$	mean packet-delivery / link reliability	[0,1]
$\tau$	latency / blackout duration	time
$\lambda$	task arrival rate	1/time
$e$	enforceability / command compliance	[0,1]
$\tau_{\text{deleg}}$	delegation threshold	[0,1]
<b>MTTA</b>	mean time-to-assignment/action	time
$P_{\text{fail}}$	deadline-miss probability	[0,1]

## Claim-Evidence-Method (CEM) Grid

Claim (C)	Evidence (E)	Method (M)	Status	Risk	TestID
Industrialization (organization, standardization, mass production of communication and persuasive technologies) increases the scale, speed, and repeatability of cognitive operations (projection and sustainment of influence).	[6]; [7]; [8]	Historical process-tracing + comparative empirical analysis (case studies of WWI/WWII/Cold War/digital era) + agent-based and system-dynamics simulation parametrized with historical and platform data.	E cited; M pending empirical case work and calibrated simulation	If false, central theoretical axis of the brief collapses — policy and defense recommendations premised on industrial structural remedies (e.g., decoupling, diversification) may be misdirected; resource allocation to industrial-scale controls could be wasteful.	T1
Infrastructural compression (centralized, high-bandwidth communication networks) reduces producer-consumer latency and enables synchronized framing campaigns and rapid narrative cascades, increasing potential for herd effects and attention monopolies.	[3]; [5]; [6]	Network-level modelling (temporal cascade models), controlled simulations of message propagation under varying latency/connectivity, and empirical measurement of cascade timing using historical broadcast logs and platform trace data.	E cited (theoretical/algorithmic anchors); M pending targeted simulations and empirical cascade measurement	If false, the emphasis on network centralization as the key driver of synchronized influence is overstated — mitigation strategies focused on network decentralization or latency controls may have limited effect.	T2

Claim (C)	Evidence (E)	Method (M)	Status	Risk	TestID
Organizational routinization (bureaucratic scaling, professionalization of information units) reduces per-message cost via standardized templates, A/B testing, and production pipelines, enabling sustained saturation and rapid optimization of persuasive content.	[6]; [7]; [8]	Ethnographic study of production units + cost-per-message modelling + field experiments (where ethical/feasible) and analysis of historical production logs (propaganda ministries, ad-farm records, platform moderation/marketing teams).	E cited (historical and conceptual anchors); M pending ethnographic access and empirical cost modelling	If false, policies aimed at breaking production pipelines (e.g., disrupting centralized content factories) may fail to reduce effective persuasion; resilience and attribution models relying on organizational signatures could be invalid.	T3
Technological tractability of perception (mass sensors, programmatic delivery, inexpensive compute and optimization engines) makes persuasion experiments cheap and scalable, producing tunable intervention parameters (frequency, targeting granularity, valence).	[1]; [4]	Large-scale A/B field experiments or natural experiments (platform partnership or synthetic populations), reproducible lab studies on parameter sensitivity, and simulations of parameter-space search under realistic budget constraints.	E cited (technical preprints); M pending field experiments and replication	If false, the presumed ease and low cost of large-scale experimentation is overstated – threat models that assume cheap, rapid optimization of persuasion may overestimate adversary capability and mis-prioritize defenses.	T4
Feedback-driven algorithmic scaling (sociotechnical	[1]; [4]; [5]	Causal inference on platform logs (when available),	E cited (algorithmic literature and ML vulnerability surveys);	If false, regulatory and technical	T5

Claim (C)	Evidence (E)	Method (M)	Status	Risk	TestID
loop: user signals → model updates → optimized distribution) accelerates the weaponization of attention by converting engagement data into progressively more effective persuasive models.		reproducible simulations of closed-loop learning systems, and field-replication of feedback dynamics with synthetic agents or controlled populations.	M pending empirical access to platform feedback loops and controlled replication	interventions focused on breaking feedback loops (rate limits, algorithm audits) may be less impactful than believed; strategic prioritization of countermeasures could be misplaced.	
Societal production of predictable cognitive ecologies (mass schooling, serialized entertainment, standardized work rhythms) creates predictable windows of attention and shared heuristics that provide high-leverage targeting points for industrialized influence.	[6]; [7]; [8]	Time-use data analysis, behavioral field studies, and targeted experiments measuring variation in susceptibility across daily/seasonal schedules and demographic cohorts.	E cited (historical/social science anchors); M pending quantitative behavioral and time-use studies	If false, targeting based on presumed predictable schedules and heuristics will be less effective — interventions predicated on temporal targeting or heuristic exploitation may yield limited returns.	T6

## Sources

[1]

An Investigation into the Performances of the State-of-the-art Machine Learning Approaches for Various Cyber-attack Detection: A Survey

Arxiv.Org, 2024-02-26. (cred: 0.50)

<http://arxiv.org/abs/2402.17045v2>

[2]

OA1-AM23-SN-05 | Canadian Pediatric Massive Hemorrhage Protocols: A Survey of National Practice and State-of-the-Art Review

Doi.Org, 2023-10-01. (cred: 0.50)

[https://doi.org/10.1111/trf.52\\_17554](https://doi.org/10.1111/trf.52_17554)

[3]

On graph theoretic results underlying the analysis of consensus in multi-agent systems

Arxiv.Org, 2009-02-24. (cred: 0.50)

<http://arxiv.org/abs/0902.4218v1>

[4]

A Brief Tutorial on Consensus ADMM for Distributed Optimization with Applications in Robotics

Arxiv.Org, 2024-10-02. (cred: 0.50)

<http://arxiv.org/abs/2410.03753v1>

[5]

A Survey of Distributed Consensus Protocols for Blockchain Networks

Arxiv.Org, 2019-04-08. (cred: 0.50)

<http://arxiv.org/abs/1904.04098v4>

Generated: 2025-11-06T11:59:20.580037 | Word Count: 3253

## Research Roadmap

- **Phase 1 (Theory):** Formalize claims, extend proofs, validate against canonical results
- **Phase 2 (Simulation):** Implement stress tests, sweep parameter spaces, measure convergence/scaling
- **Phase 3 (Empirical):** Deploy in controlled environments, collect field data, validate predictions
- **Phase 4 (Integration):** Operationalize with human-in-loop, adversarial hardening, production deployment

**Confidence Methodology:** Confidence = 0.3·SourceDiversity + 0.25·AnchorCoverage + 0.25·MethodTransparency + 0.2·ReplicationReadiness, where SourceDiversity reflects unique publishers & types, AnchorCoverage reflects share of primary claims with Type-1 anchors, MethodTransparency reflects CEM completeness & assumptions ledger, and ReplicationReadiness reflects sim plan & datasets/params specified.

