

Beyond Persuasion: Algorithmic Reality Engineering and Cognitive Substrate Hijacking (ARE-CSH) in Next-Generation Psychological Operations

Majdi T. Rahman

Global Research & Analysis Lab (GRAL), Romania

Abstract

Psychological operations (PSYOP) have conventionally been theorized as linear acts of persuasion in which a sender transmits a message through a medium to a receiver presumed to evaluate it deliberatively. This research article argues that the model has become analytically inadequate in environments where the informational infrastructure itself mediates perception. It advances the Theory of Algorithmic Reality Engineering and Cognitive Substrate Hijacking (ARE-CSH), which holds that next-generation PSYOP operates less on beliefs than on the conditions under which beliefs are formed. By shaping recommendation systems, personalized micro-stimuli, and synthetic media, an operator can alter perceptual boundaries, a process termed epistemic conditioning, without engaging top-level cognitive evaluation. The research article specifies four stages of Cognitive Substrate Hijacking (epistemic pre-conditioning, affective resonance capture, algorithmic loop reinforcement, and reality bifurcation with neural exhaustion), derives four testable propositions, and situates the framework in a comparative matrix against twentieth-century PSYOP and early digital information warfare. It argues that fact-checking and media literacy, while valuable, are structurally insufficient against conditioning at the substrate level, and proposes a defense model built on structural algorithmic hygiene, epistemic isolation routines, and cognitive stress inoculation. The framework is offered as a set of falsifiable hypotheses rather than established findings, and boundary conditions are identified throughout.

Keywords: psychological operations, cognitive warfare, algorithmic governance, epistemic conditioning, synthetic media, strategic communication

1. Introduction and Historical Context

The study of psychological operations has long been organized around a communicative grammar. Lasswell's (1948) formulation, who says what in which channel to whom with what effect, supplied a template that Shannon and Weaver (1949) rendered in engineering terms: a source encodes a message, a channel carries it, noise degrades it, and a receiver decodes it. Theorists and practitioners of propaganda (Bernays, 1928; Ellul, 1965) elaborated the template into doctrines of mass persuasion in which leaflets, radio, and later television served as instruments for altering the beliefs, emotions, and behavior of identifiable audiences. Even the empirical corrections to the strongest versions of this model, notably the limited-effects tradition and the two-step flow hypothesis (Katz & Lazarsfeld, 1955), preserved its

architecture. The message remained the unit of analysis, and the receiver remained a sense-making agent who weighed it, discounted it, or accepted it (Jowett & O'Donnell, 2019).

The late twentieth century widened the aperture without replacing the architecture. Arquilla and Ronfeldt (1993) proposed cyberwar and netwar as organizing concepts, and Szafranski (1997) argued for a form of "neocortical warfare" directed at the adversary's will and cognition. Libicki (1995) classified information warfare into distinct forms, psychological warfare among them, while Boyd's (1987) OODA loop recast conflict as competition in the tempo of observation, orientation, decision, and action, making the adversary's orientation, and not merely its beliefs, a strategic object. In the 2010s, Russian active measures and "firehose" campaigns (Pomerantsev & Weiss, 2014; Paul & Matthews, 2016; Rid, 2020) and the NATO-sponsored concept of cognitive warfare (Claverie & du Cluzel, 2021) shifted attention toward the human mind as a contested domain. Yet most of this literature continues to describe the operation as the delivery of content, whether true, false, or misleading, to an audience.

This research article argues that the shared weakness of these frameworks is an implicit theory of cognition. They treat the human being as a deliberative interpreter of external data: an agent who meets information in a relatively neutral environment and then evaluates it. Three bodies of work undermine that assumption. First, research on heuristics and dual-process judgment shows that evaluation is routinely driven by fast, associative, effort-minimizing processes whose inputs depend on what is available and salient (Tversky & Kahneman, 1974; Kahneman, 2011). Second, the philosophy of information treats human agents as informational organisms embedded in an infosphere that they co-constitute with technologies, rather than merely consult (Floridi, 2011). Third, the political economy of digital platforms demonstrates that the environment is no longer neutral: ranking, recommendation, and notification systems continuously select, order, and time the stimuli that reach each person, and they do so in order to modify behavior (Pariser, 2011; Tufekci, 2015; Zuboff, 2019). If cognition depends on a curated computational architecture, then the architecture, and not the message, is the decisive terrain.

The research paper therefore introduces the Theory of Algorithmic Reality Engineering and Cognitive Substrate Hijacking (ARE-CSH). Its contribution is fourfold: a formal definition of the phenomenon; a four-stage process model; four propositions that are specific enough to be tested and, where the evidence is mixed, qualified; and a defense model that follows from the diagnosis. The theory is explicitly a framework for hypothesis generation. Where existing evidence is contested (for example, on the size of recommendation-system effects or the prevalence of backfire reactions), the text says so. Section 2 develops the conceptual foundations, Section 3 the propositions, Section 4 a comparative matrix, Section 5 the strategic and defensive implications, and Section 6 a research agenda.

2. Conceptual Foundations: Cognitive Substrate Hijacking

2.1 Definition and Scope

Cognitive Substrate Hijacking is defined here as the sustained, covert manipulation of the computational, social, and physiological conditions under which a population perceives, attends to, and evaluates information, such that an operator acquires durable influence over what is perceivable, salient, and plausible without needing to win any argument. The term substrate denotes three interdependent layers. The computational layer comprises ranking, recommendation, search, moderation, and notification systems, together with the generative tools that populate them with content. The social layer comprises network topology, group-identity cues, and the perceived consensus of reference groups. The physiological layer comprises attention, arousal, sleep, and cognitive load, the bodily resources on which

evaluation depends. Hijacking signifies that the operator need not own these layers. It is sufficient to exploit their existing optimization logic, engagement maximization above all, so that the system's normal operation delivers the conditions the operator wants. Algorithmic Reality Engineering is the corresponding operational practice: the deliberate design of informational environments, at individual and population scale, to produce a specified perceptual regime.

2.2 Persuasion Versus Epistemic Conditioning

Persuasion alters propositional attitudes within a stable perceptual field. The target sees the same facts, sources, and options as before, and the operator tries to change what is concluded from them. Epistemic conditioning alters the field itself. It modifies which facts are encountered (availability), which are prominent (salience), which sources are credible (trust), and which conclusions feel plausible (priors), thereby shifting the boundaries within which any later persuasion or reasoning occurs. The distinction parallels Nguyen's (2020) contrast between epistemic bubbles, in which relevant voices are omitted by accident or design, and echo chambers, in which outside voices are actively discredited. Bubbles are a feature of exposure; echo chambers are a feature of trust. Epistemic conditioning, as used here, covers both and treats the passage from the first to the second as the critical escalation. Because conditioned boundaries are experienced by the target as the natural shape of the world rather than as a claim made by someone, they do not attract the scrutiny that explicit assertions do.

2.3 The Four Stages of Cognitive Substrate Hijacking

The process model below is analytical rather than strictly chronological; in practice stages overlap and recur. Table 1 summarizes the stages.

Stage 1: Epistemic Pre-Conditioning (Structural Alignment of Information Pipelines)

The operator aligns information pipelines with its objectives before any contested content is deployed. Typical activities include optimizing content for search and recommendation ranking, cultivating accounts and communities that accumulate algorithmic standing, shaping the vocabulary under which topics are indexed, and degrading the discoverability of competing sources. The psychological basis is the availability heuristic (Tversky & Kahneman, 1974) together with familiarity effects: what appears often and early becomes the default reference frame. The signature of this stage is that nothing overtly persuasive occurs; the operator is arranging the library, not writing the books.

Stage 2: Affective Resonance Capture (Targeting Affective Sub-Routines through Behavioral and Biometric Optimization)

Using behavioral traces such as dwell time, reaction types, and posting patterns, the operator, or an engagement-optimizing system it exploits, identifies which stimuli elicit high-arousal responses in particular individuals and groups. Digital footprints predict personal attributes with considerable accuracy (Kosinski et al., 2013), and messages matched to psychological profiles increase behavioral responses at scale (Matz et al., 2017). Moral-emotional language accelerates diffusion in networks (Brady et al., 2017), and the affective tone of expressed content can be shifted by feed composition (Kramer et al., 2014). The outcome is a calibrated stimulus profile for each segment. With biometric streams (wearables, voice, facial analysis), calibration could in principle move from inferred to measured arousal; this possibility is largely prospective and is flagged as such. The phrase "limbic sub-routines" is used throughout as shorthand for fast affective appraisal processes and not as a claim about a discrete anatomical system (LeDoux, 2000).

Stage 3: Algorithmic Loop Reinforcement (Echo-Chamber Locking through Micro-Targeted Feedback)

Once affective resonance is established, the feedback loop between user behavior and system response

consolidates it. Engagement with resonant content raises its recommended share, which narrows exposure, which increases the relative weight of the resonant frame (Pariser, 2011). The decisive step is the shift from exposure to trust identified by Nguyen (2020): when external sources are framed as hostile or corrupt, the enclosure becomes self-protecting, because disconfirming information is pre-emptively discredited. Micro-targeted feedback, including replies, apparent peer agreement, and social proof that is often synthetic, supplies the social reinforcement that completes the loop. At this stage the operator can reduce active effort, because the infrastructure performs the labor of maintenance.

Stage 4: Reality Bifurcation and Neural Exhaustion (Systemic Decision Paralysis and Societal Action Collapse)

Two outcomes emerge at population scale. Segments become mutually exclusive in evidence and trust (bifurcation), and the persistent demand to adjudicate between contested, emotionally charged claims depletes the effort people devote to verification (exhaustion), producing apathy, withdrawal, or dependence on identity-based shortcuts. The strategic effect is action collapse: institutions cannot assemble the shared situational picture required for collective decision, and the society's OODA cycle (Boyd, 1987) lengthens until a competitor can operate inside it. "Neural exhaustion" is a deliberately evocative label; its operational meaning is given in Proposition II.

Table 1. The four stages of Cognitive Substrate Hijacking

Stage	Primary mechanism	Substrate layer targeted	Observable indicators
1. Epistemic Pre-Conditioning	Alignment of ranking, search, and indexing pipelines; vocabulary shaping	Computational	Shifts in source prominence; altered default search and recommendation results
2. Affective Resonance Capture	Behavioral (and prospectively biometric) profiling; arousal-matched stimuli	Physiological, computational	Elevated reaction and share rates for moral-emotional content in targeted segments
3. Algorithmic Loop Reinforcement	Engagement feedback loops; synthetic social proof; discrediting of outside sources	Computational, social	Narrowing exposure diversity; rising distrust of external sources
4. Reality Bifurcation and Neural Exhaustion	Mutually exclusive evidence and trust structures; chronic oversaturation	Social, physiological	Falling cross-segment evidence overlap; reduced verification; decision latency

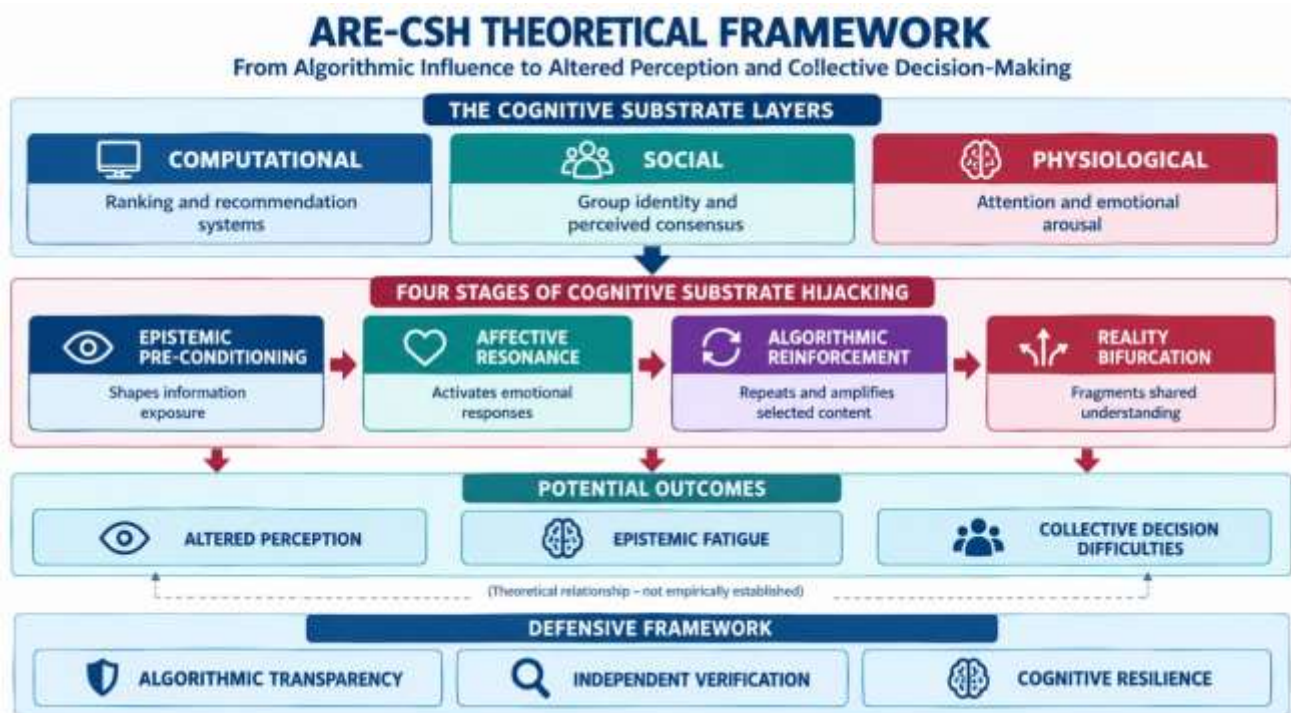


Figure 1. The ARE-CSH Theoretical Framework.

Schematic illustration mapping the theoretical architecture of Algorithmic Reality Engineering and Cognitive Substrate Hijacking across core substrate layers, operational stages, proposed downstream outcomes, and defensive interventions.

Above, Figure 1 illustrates the complete ARE-CSH theoretical architecture, tracing how manipulation moves across the three substrate layers—computational, social, and physiological—through a four-stage operational sequence. It maps these stages to their predicted downstream outcomes of altered perception, epistemic fatigue, and collective decision failure, while linking targeted defensive countermeasures to each level of exposure.

3. Core Theoretical Propositions

Proposition I: The Shift from Message-Centric to Infrastructure-Centric Warfare

Statement. The efficacy of an ARE-CSH operation varies inversely with the target population's awareness of the medium through which the operation is conducted.

Two mechanisms support the proposition. The first is resistance. Psychological reactance theory holds that perceived threats to freedom of choice generate motivation to restore it (Brehm, 1966), and the persuasion knowledge model shows that recognizing an influence attempt triggers coping and discounting (Friestad & Wright, 1994). Both processes presuppose that the target can identify a source and an intent. In conventional PSYOP the message is an explicit object bearing both, so these defenses are available. When influence is embedded in ranking, search, and notification logic, there is no discrete object to discount, and the target experiences the outcome as the product of free exploration. The second mechanism is opacity. Algorithmic systems are opaque through corporate secrecy, technical complexity, and the mismatch between machine-scale pattern recognition and human interpretation (Burrell, 2016), and their effects are individually invisible because each user sees only a personalized slice (Tufekci, 2015). Awareness of the medium is therefore not merely low; it is structurally difficult to raise. An infrastructure-

centric operator consequently enjoys an asymmetric advantage, since it is shielded from the very countermeasures (attribution, exposure, discounting) that damage message-centric operations. The conclusion for strategists is a change of unit: the relevant object of analysis is no longer the narrative but the pipeline that decides which narratives are encountered.

The proposition concerns cumulative, upstream conditioning and not the effect of any single exposure. Large field experiments that altered feed algorithms for several months found small or null changes in measured attitudes (Guess et al., 2023), and exposure to ideologically diverse content is strongly shaped by users' own choices (Bakshy et al., 2015). These findings do not falsify Proposition I, whose dependent variable is long-run perceptual boundary shift and whose moderating variable is awareness, but they demand designs that manipulate transparency and not only exposure. A testable implication follows: interventions that make curation salient (for instance, disclosure of why an item was shown) should attenuate conditioning effects relative to equivalent exposure without disclosure.

Proposition II: Neural Exhaustion and Epistemic Paralysis

Statement. The ultimate objective of mature cognitive warfare is not the installation of a specific belief but the induction of cognitive oversaturation, which yields systemic decision paralysis.

Traditional propaganda sought conversion to a doctrine. In an environment of effectively unlimited information, converting a hostile population is expensive and uncertain, whereas degrading its capacity to adjudicate between claims is cheap. The "firehose of falsehood" model documented in Russian practice relies on high volume, multiple channels, rapid and continuous output, and indifference to consistency (Paul & Matthews, 2016); its purpose is better understood as disorientation than as persuasion (Pomerantsev & Weiss, 2014). Arendt (1951) anticipated the logic when she observed that the ideal subject of totalitarian rule is a person for whom the distinction between fact and fiction has ceased to be meaningful. Generative media intensifies the dynamic by driving the marginal cost of plausible fabrication toward zero (Goldstein et al., 2023) and by creating a "liar's dividend," in which authentic evidence can be dismissed as synthetic (Chesney & Citron, 2019).

The psychological mechanism should be stated with care. Cognitive load theory holds that working memory is a limited resource and that excessive extraneous demands impair judgment (Sweller, 1988), and research on scarcity suggests that chronic overload narrows the bandwidth available for deliberate decisions (Mullainathan & Shafir, 2013). The strong "ego depletion" account of self-control has failed to replicate reliably (Hagger et al., 2016) and is not relied upon here. "Neural exhaustion" is accordingly operationalized behaviorally and not neurally: reduced verification effort, increased reliance on source identity as a heuristic, generalized distrust, and disengagement from civic participation. The proposition predicts that populations exposed to contradictory, high-emotion, high-volume streams will show these markers even when no individual claim is believed. Its distinguishing implication is that success may be achieved with falsehoods that no one finds convincing, provided that they cannot be efficiently ruled out. The consequences are apathy, social atomization, or surrender of judgment to pre-selected authorities, each of which reduces the population's capacity for coordinated response (Wardle & Derakhshan, 2017).

Proposition III: Affective Hijacking Precedes Epistemic Formulations

Statement. Under conditions of high arousal, time pressure, and repeated exposure, affective response predicts the direction of subsequent reasoning more strongly than evidence does, so that rational deliberation functions largely to justify an initial emotional orientation.

The proposition draws on a long line of work on affective primacy. Zajonc (1980) argued that affective reactions can occur with minimal cognitive processing and can precede and shape judgment, a view

disputed by appraisal theorists who stress that some cognitive evaluation is a precondition of emotion (Lazarus, 1984). Haidt's (2001) social intuitionist model holds that moral judgments arise from rapid intuition, with reasoning typically constructed afterwards to defend the verdict. Kahneman (2011) similarly describes a fast system whose outputs the slower system often endorses without scrutiny. Motivated reasoning research adds that directional goals shape the search for, and evaluation of, evidence (Kunda, 1990), and that people discount congenial and uncongenial information asymmetrically (Taber & Lodge, 2006). Mercier and Sperber's (2011) argumentative theory of reasoning makes the same point at the level of function: reasoning evolved more for justification and persuasion of others than for solitary truth-seeking.

What ARE-CSH contributes is an operational layer. Where earlier work treats affect-first processing as a feature of human psychology, the theory identifies it as an exploitable control variable. Behavioral and prospective biometric data allow stimuli to be tuned to individual arousal profiles (Kosinski et al., 2013; Matz et al., 2017), and moral-emotional framing selectively accelerates diffusion (Brady et al., 2017). Once anger, fear, or group identity is activated, later analytic activity acts as an interpreter that produces reasons consistent with the engineered feeling. Two cautions apply. The affect-versus-cognition priority debate is unresolved, and the proposition is therefore stated conditionally rather than as a universal law. And the "limbic" vocabulary is a convenient simplification of distributed affective circuitry (LeDoux, 2000). The empirical signature of the proposition is that manipulating arousal prior to evidence presentation should shift evidence evaluation in the direction of the induced affect, and that this effect should exceed the effect of evidence quality in time-pressured conditions.

Proposition IV: Reality Bifurcation (The Echo-System Effect)

Statement. Campaigns that address segments of a population with mutually exclusive evidence and trust structures reduce the shared informational base on which social cohesion depends, and under stated conditions make consensus formation formally unattainable.

Micro-targeting and machine-learning-driven personalization permit an operator to present opposing factions with incompatible renderings of the same events. Group A is shown a world in which a given issue is an existential threat; Group B is shown a world in which it is a non-issue and Group A is irrational. Each group can behave rationally relative to its own informational environment, which is the point: the operator does not need to deceive any group about its own interests, only to ensure that the groups' epistemic worlds do not overlap.

The claim can be given a stylized formal expression. Let a population be partitioned into segments 1 to k , let E_i denote the evidence set to which segment i is exposed, and let T_i denote the set of sources it treats as credible. Define the shared informational base as $B = E_1 \cap E_2 \cap \dots \cap E_k$. As B approaches the empty set and the sets T_i become disjoint, no proposition exists that all segments regard as admissible evidence. Three results from formal social theory then apply. In DeGroot (1974) opinion-updating models, consensus requires that the trust network be sufficiently connected; where it decomposes into closed groups that place no weight on outsiders, opinions converge within groups but not across them (see also Golub & Jackson, 2010). In bounded-confidence models, agents only update toward those whose views lie within a threshold, and when that threshold is small relative to the opinion spread, stable clusters persist (Hegselmann & Krause, 2002). And Aumann's (1976) agreement theorem shows that rational disagreement cannot persist under common priors and common knowledge, so disagreement that does persist signals the absence of those preconditions, which is exactly what bifurcation targets. Taken together, these results imply that under disjoint evidence and trust structures social cohesion is not merely

difficult but formally unreachable by deliberation alone. The word "mathematical" should be read in this restricted sense: it is a conditional result about stylized models, not a theorem about empirical societies. Empirical qualifications are necessary. Measured affective polarization varies considerably across countries and platforms, exposure to opposing views can in some conditions intensify rather than reduce polarization (Bail et al., 2018), and much fragmentation arises from commercial optimization without any hostile intent. ARE-CSH adds purposive design to a condition that platforms already generate by default; attribution of intent therefore remains a central methodological difficulty. The proposition predicts that cross-segment overlap in evidence and in trusted sources will decline faster, and recover more slowly, in populations subject to coordinated segment-specific targeting than in matched populations exposed only to commercial personalization.

4. Comparative Paradigm Matrix

Table 2 compares twentieth-century PSYOP, early digital information warfare, and ARE-CSH along five strategic dimensions. The matrix is not a claim that earlier paradigms have disappeared; leaflets, broadcasts, and bot networks remain in use. It claims that the center of gravity has migrated, and that defensive and measurement practices built for earlier paradigms are mismatched to the later one. The measurement row deserves particular attention. Reach and engagement metrics, which dominate current assessment practice, are indicators of message circulation and therefore systematically understate (or fail to register) successful substrate-level operations, whose outcomes appear as lengthened decision cycles, eroded cross-group trust, and declining verification behavior.

Table 2. Comparative paradigm matrix

Strategic dimension	20th-century PSYOP	Early digital information warfare	ARE-CSH
Primary vector	Leaflets, radio, television, print; broadcast of discrete messages	Social media posts, bot and troll networks, hack-and-leak operations	Recommendation, search, and notification systems; generative synthetic media; behavioral and prospectively biometric data streams
Target mechanism	Conscious persuasion and ideological alignment; appeals to interest and fear	Information overload, amplification of polarization, erosion of trust	Alteration of perceptual boundaries (availability, salience, trust); affective capture preceding reasoning; segment-specific realities
Operational goal	Defection, surrender, behavioral compliance, morale degradation	Disinformation spread, delegitimization, disruption of discourse	Epistemic paralysis, institutional decision failure, durable societal fragmentation

Strategic dimension	20th-century PSYOP	Early digital information warfare	ARE-CSH
Success metrics	Defection and surrender rates, audience reach, recall surveys	Engagement, shares, follower counts, narrative penetration	Verification latency, institutional decision latency (OODA tempo), cross-segment evidence overlap, variance in institutional trust
Defensive countermeasures	Jamming, censorship, counter-propaganda, truthful messaging	Fact-checking, account takedowns, content moderation, media literacy campaigns	Structural algorithmic hygiene, epistemic isolation routines, cognitive stress inoculation, independent algorithmic audit

5. Strategic Implications and Counter-Frameworks

5.1 Why Fact-Checking and Media Literacy Are Structurally Insufficient

The dominant response to hostile influence has been counter-disinformation: fact-checking, debunking, and media literacy education. These tools are valuable, but ARE-CSH predicts four reasons why they fail to address substrate-level operations. First, they act at the wrong level. Corrections operate on claims, whereas conditioning operates on salience, trust, and affect; a corrected claim leaves the pipeline that produced it intact. Second, affective pre-conditioning degrades uptake. Misinformation continues to influence reasoning after correction (Lewandowsky et al., 2012), and directional motivation leads people to resist corrections that threaten valued identities (Kunda, 1990; Taber & Lodge, 2006). Early work reported that corrections could even strengthen the original belief, the so-called backfire effect (Nyhan & Reifler, 2010); later large-sample replications found such reversals to be rare (Wood & Porter, 2019). The theory therefore does not rest on backfire as a general law. The more defensible claim is that correction effects are small, decay quickly, and are conditional on trust in the corrector, which is itself a substrate variable under attack. Third, there is a tempo and volume asymmetry: falsehood diffuses further and faster than truth (Vosoughi et al., 2018), and under Proposition II the objective is overload, which verification capacity cannot match at scale. Fourth, in the presence of the liar's dividend (Chesney & Citron, 2019), the very act of authentication becomes a contested move.

Media literacy faces parallel limits. Programs that teach the evaluation of individual items can improve discrimination, and approaches such as lateral reading, which checks a source by leaving the page, show real promise (Wineburg & McGrew, 2019). Yet boyd (2018) cautions that literacy narratives can be appropriated to cultivate generalized distrust, which is precisely the Stage 4 end state. Item-level skill does not alter the structure that selects the items. Defense must therefore be extended from content to conditions.

5.2 A Defense Model: Cognitive Resilience through Structural Algorithmic Hygiene

The proposed model has three mutually reinforcing components, each aligned to one layer of the substrate (computational, institutional, and physiological-individual), and each linked to measurable indicators.

Structural Algorithmic Hygiene

This component addresses the computational layer. It introduces intentional friction and diversity require-

ments into content recommendation and amplification systems so as to interrupt affective feedback loops (Stage 3). Measures include velocity limits or "circuit breakers" on the amplification of rapidly spreading high-arousal content, accuracy prompts that shift attention to veracity before sharing (Pennycook et al., 2021), source-diversity constraints in ranking, provenance and authentication standards for synthetic media, and independent auditing of recommendation systems. The regulatory architecture of the European Union's Digital Services Act, which imposes systemic-risk assessment and audit duties on very large platforms (European Parliament & Council, 2022), is an early institutional exemplar. Because Proposition I identifies opacity as the operator's main asset, transparency about why content is shown is treated as a security measure and not only a consumer right. Indicators include cross-segment evidence overlap, concentration of exposure sources, and time-to-amplification of flagged coordinated content.

Epistemic Isolation Routines

This component addresses the institutional layer, in particular the decision-making of commanders, officials, and crisis teams. It establishes protocols that strictly separate raw sensory and source data from machine-generated predictive analysis during high-stress decisions, so that algorithmic inference cannot silently become the observation on which orientation rests (Boyd, 1987). Practices include mandatory provenance labeling of inputs, structured analytic techniques and red-teaming (Heuer & Pherson, 2014), deliberately imposed decision latency for irreversible actions, and the use of independent, non-networked verification channels. The justification is consistent with research showing that expert intuition is reliable only in stable, learnable environments with adequate feedback (Kahneman & Klein, 2009), conditions that an adversarially manipulated information environment does not satisfy. Indicators include the proportion of critical decisions with independently verified inputs and the measured divergence between machine-assisted and independently derived situational assessments.

Cognitive Stress Inoculation

This component addresses the physiological and individual layer. It trains military and civilian leaders, and ultimately the wider population, to recognize the bodily and emotional signs of affective hijacking before consequential decisions are made. It combines two evidence-based traditions. Inoculation theory shows that exposure to weakened forms of manipulation techniques builds resistance to them (McGuire, 1964), and technique-based inoculation interventions have reduced susceptibility in large online experiments (Roozenbeek & van der Linden, 2019; van der Linden, 2022). Stress inoculation training supplies graduated exposure to stressors together with coping skills (Meichenbaum, 1985), and research on emotion regulation provides strategies such as reappraisal and attentional deployment that can be applied at the moment of arousal (Gross, 2015). The intended outcome is a learned pause: detection of one's own arousal as data about the environment. Indicators include performance in simulated high-arousal decision exercises, latency before sharing or acting on emotive content, and calibrated confidence in judgments.

Integration and Constraints

The three components are complementary because each interrupts a different stage: hygiene targets Stages 1 and 3, inoculation targets Stage 2, and isolation routines protect institutions against Stage 4. A normative constraint is essential. Countermeasures must not replicate the techniques they oppose; state-directed control of recommendation systems would itself be a form of substrate capture. The model therefore favors transparency, independent audit, user agency, and democratic oversight, consistent with Zuboff's (2019) warning about concentrated instrumentarian power. The model's effectiveness is presently an empirical

hypothesis, and several of its components (notably accuracy prompts and inoculation) have stronger evidence than others (notably circuit breakers and decision-latency mandates).

6. Conclusion

This research paper has argued that psychological operations are undergoing a paradigm shift from message-centric persuasion to infrastructure-centric epistemic conditioning. The classical sender-message-medium-receiver model presupposes a deliberative receiver in a neutral environment; contemporary evidence on heuristic judgment, affect-driven reasoning, and algorithmic curation shows that neither assumption is secure. ARE-CSH defines the resulting phenomenon, decomposes it into four stages, advances four propositions on medium invisibility, cognitive oversaturation, affective primacy, and reality bifurcation, and derives a defense model oriented to structure rather than content. The most important strategic consequence is a change in what is defended: not only the truth of individual claims, but the integrity of the conditions under which claims are encountered, weighed, and acted upon.

Several research priorities follow. First, in neuroscience and psychophysiology, the affective primacy claim (Proposition III) should be tested in naturalistic feed environments with electrodermal, pupillometric, and eye-tracking measures, using designs that distinguish arousal-driven shifts in evidence evaluation from ordinary motivated reasoning. Second, in machine learning and computational social science, sock-puppet audits, agent-based simulations of bifurcation (extending bounded-confidence and DeGroot models with platform-mediated exposure), and adversarial testing of recommendation systems can estimate how little coordinated input is required to shift substrate parameters. Third, causal identification remains the largest obstacle: field experiments of the kind reported by Guess et al. (2023) need to be extended to long horizons, to transparency manipulations, and to settings in which platform data are accessible to independent researchers. Fourth, measurement validity requires work: constructs such as epistemic fatigue, cross-segment evidence overlap, and institutional decision latency need validated operational definitions before they can function as success or defense metrics. Fifth, defense trials should evaluate the three-component model, including randomized evaluation of algorithmic friction interventions, inoculation programs in military and civil-service cohorts, and the effect of isolation routines on decision quality in exercises. Sixth, generative AI warrants dedicated study of automated, personalized influence at scale (Goldstein et al., 2023), including the detection of synthetic social proof. Finally, normative and legal research must address attribution, proportionality, and the democratic limits of defensive intervention, so that resilience is built without reproducing the control it seeks to resist.

References

1. Arendt, H. (1951). *The origins of totalitarianism*. Harcourt, Brace.
2. Arquilla, J., & Ronfeldt, D. (1993). Cyberwar is coming! *Comparative Strategy*, 12(2), 141-165.
3. Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4(6), 1236-1239.
4. Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221.
5. Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
6. Bernays, E. L. (1928). *Propaganda*. Horace Liveright.
7. boyd, d. (2018, March 9). You think you want media literacy... do you? *Points (Data & Society)*.

8. Boyd, J. R. (1987). A discourse on winning and losing [Unpublished briefing slides]. Air University Library, Maxwell Air Force Base.
9. Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318.
10. Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.
11. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12.
12. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1819.
13. Claverie, B., & du Cluzel, F. (2021). The cognitive warfare concept. Innovation Hub sponsored by NATO Allied Command Transformation.
14. DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118-121.
15. Ellul, J. (1965). *Propaganda: The formation of men's attitudes*. Knopf.
16. European Parliament & Council of the European Union. (2022). Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act). *Official Journal of the European Union*, L 277, 1-102.
17. Floridi, L. (2011). *The philosophy of information*. Oxford University Press.
18. Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1), 1-31.
19. Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations: Emerging threats and potential mitigations (arXiv:2301.04246).
20. Golub, B., & Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1), 112-149.
21. Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry*, 26(1), 130-137.
22. Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398-404.
23. Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546-573.
24. Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.
25. Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2.
26. Heuer, R. J., Jr., & Pherson, R. H. (2014). *Structured analytic techniques for intelligence analysis* (2nd ed.). CQ Press.
27. Jowett, G. S., & O'Donnell, V. (2019). *Propaganda & persuasion* (7th ed.). SAGE.
28. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

29. Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515-526.
30. Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Free Press.
31. Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802-5805.
32. Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
33. Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
34. Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37-51). Harper.
35. Lazarus, R. S. (1984). On the primacy of cognition. *American Psychologist*, 39(2), 124-129.
36. LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155-184.
37. Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131.
38. Libicki, M. C. (1995). *What is information warfare?* National Defense University Press.
39. Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714-12719.
40. McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 191-229). Academic Press.
41. Meichenbaum, D. (1985). *Stress inoculation training*. Pergamon Press.
42. Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74.
43. Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Times Books.
44. Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141-161.
45. Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330.
46. Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
47. Paul, C., & Matthews, M. (2016). The Russian 'firehose of falsehood' propaganda model: Why it might work and options to counter it (PE-198-OSD). RAND Corporation.
48. Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592, 590-595.
49. Pomerantsev, P., & Weiss, M. (2014). *The menace of unreality: How the Kremlin weaponizes information, culture and money*. Institute of Modern Russia.
50. Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.
51. Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5, Article 65.

52. Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. University of Illinois Press.
53. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.
54. Szafranski, R. (1997). Neocortical warfare? The acme of skill. In J. Arquilla & D. Ronfeldt (Eds.), *In Athena's camp: Preparing for conflict in the information age* (pp. 395-416). RAND Corporation.
55. Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-769.
56. Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, 13(2), 203-218.
57. Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.
58. van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 28, 460-467.
59. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
60. Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making (DGI(2017)09). Council of Europe.
61. Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, 121(11), 1-40.
62. Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41(1), 135-163.
63. Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151-175.
64. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.