

# META-MEMETICS

Engineering Belief in Networked Environments

Aleph Station

November 2025

## Abstract

Ideas often propagate on dimensions orthogonal to truth; we model those fitness dimensions explicitly. This paper operationalizes memetic fitness as  $F(m,E) = A \cdot R \cdot X \cdot T - C$ , where success depends on assimilation ease, retention, expression drivers, transmission efficiency, and cognitive/-social costs. The model is extended with environmental modulators  $M$  (complexity-capacity match) and  $N$  (network topology effects). We estimate effects in log-space and map to bounded propagation probability via logistic link; predictions are probabilistic and context-sensitive.

We demonstrate how repetition increases perceived truth independent of accuracy, how moral-emotional language amplifies diffusion, and why false information often—but not always—outcompetes truth in digital networks. We then introduce holarchic integration as both a defensive protocol against manipulation and a constructive methodology for knowledge synthesis.

**We center an agency inversion: ideas recruit human cognition to reproduce. Our tools treat “share intent” as a memetic drive signal and introduce interventions that restore conscious participation.**

# Contents

<b>Abstract</b>	<b>2</b>
<b>1. Introduction</b>	<b>4</b>
<b>2. Memetic Fitness: A Formal Model</b>	<b>5</b>
2.1 The Fitness Function . . . . .	5
Agency Index ( $A_g$ ) . . . . .	5
2.2 Environmental Modulation . . . . .	6
2.3 Empirical Validation . . . . .	6
<b>3. Agency Inversion: Who Acts When Ideas Spread?</b>	<b>7</b>
The Buffer-Extract-Decide Loop . . . . .	7
<b>4. Holarchic Integration</b>	<b>8</b>
4.1 Development and Transcendence . . . . .	8
4.2 Operational Protocol . . . . .	8
4.3 Application to Defense . . . . .	8
<b>5. Information Warfare Landscape</b>	<b>9</b>
5.1 Scale $\neq$ Persuasion (But Agency Matters) . . . . .	9
5.2 Structural Vulnerabilities . . . . .	9
<b>6. Operational Tools</b>	<b>10</b>
6.1 Features $\rightarrow$ Measures . . . . .	10
6.2 Scoring Rubric . . . . .	10
6.3 Defensive Protocols . . . . .	11
<b>7. Ethical Framework</b>	<b>12</b>
7.1 Constraints . . . . .	12
7.2 Intended Use . . . . .	12
<b>8. Synthesis</b>	<b>13</b>
<b>9. Conclusion</b>	<b>14</b>
<b>Appendix: Implementation</b>	<b>15</b>
<b>References</b>	<b>16</b>

# 1. Introduction

**Consider a simple inversion: You don’t have ideas. Ideas have you.**

This paper takes that inversion seriously. Ideas are replicators competing for the limited resource of human attention and transmission capacity. Like viruses hijacking cellular machinery, successful ideas restructure their hosts’ behavior to ensure their own reproduction. The feeling of “choosing” to share an idea may be the idea using you to propagate itself.

When you feel compelled to share content, who is the agent? The subjective experience of “deciding” to share may be a post-hoc narrative overlaid on a process already initiated by the memetic payload’s interaction with your cognitive architecture. The idea has already begun using you before “you” decide anything.

**Operational claim.** We treat the felt urge to transmit as an observable of the payload, not the person. Measurable signals—latency to share, quote-extraction ease, identity-threat markers—indicate the degree to which the payload has already engaged automatic pathways. Our methods quantify these signals and insert friction where it matters.

We present two complementary insights:

**First**, ideas spread based on memetic fitness rather than truth value. We formalize this through a fitness function that predicts propagation success based on measurable characteristics. The model is explicitly probabilistic, not deterministic.

**Second**, cognitive frameworks represent partial truths from specific vantage points. These can be integrated holarchically—each level transcending and including previous insights—without collapsing into relativism.

These insights combine to offer both defensive capabilities against information warfare and constructive protocols for synthesis across seemingly incompatible positions.

## 2. Memetic Fitness: A Formal Model

### 2.1 The Fitness Function

We model the fitness of a memetic payload  $m$  in environment  $E$ :

$$F(m, E) = A \cdot R \cdot X \cdot T - C$$

Extended with environmental modulators:

$$F'(m, E, N) = F(m, E) \cdot M \cdot N$$

**Where:**

- **A** (Assimilation): Ease of initial comprehension
- **R** (Retention): Probability of recall over time
- **X** (Expression): Likelihood of sharing
- **T** (Transmission): Efficiency of transfer
- **C** (Cost): Cognitive load and social penalties
- **M** (Match): Complexity-capacity alignment
- **N** (Network): Topological position effects

**For estimation, we use log-space:**

$$\log F' = \log A + \log R + \log X + \log T + \log M + \log N - \lambda C$$

$$P = \sigma(\alpha + \beta \log F')$$

Where  $P \in [0, 1]$  is propagation probability via logistic link.

The propagation connects to epidemic dynamics:  $R_0 = \bar{d} \cdot P$ , where payloads become supercritical when  $R_0 > 1$ .

### Agency Index ( $A_g$ )

We introduce the **Agency Index** to measure how much the idea is driving versus the person:

$$A_g = \frac{D}{D + \kappa X}$$

**Where:**

- **D** = deliberation proxy (normalized time-to-share)
- **X** = expression driver from the model
- $\kappa$  = calibration constant

$A_g \in (0, 1)$  trends down as memetic drive outruns deliberation. We report  $A_g$  alongside  $P$  to separate “will it spread?” from “who is steering?” **Low  $A_g$  flags low-agency transmission.**

**Implementation note:**  $D$  uses per-platform z-scored time-to-share; longer than median increases  $A_g$ . When varying  $M$  and  $N$  in counterfactuals, we also show the shift in  $A_g$  to demonstrate that some “growth hacks” lower agency even as  $P$  rises.

## 2.2 Environmental Modulation

**Match (M):** Alignment between payload complexity and audience capacity

$$M = \exp\{-\gamma|\Delta|\}$$

where  $\Delta$  is the complexity gap,  $\gamma \approx 0.347$  (two-level gap halves  $M$ )

**Network (N):** Topological position effects

$$N = \exp(\eta_1 \tilde{k} + \eta_2 \widetilde{bc} + \eta_3 \widetilde{cc})$$

- $\tilde{k}$  = normalized degree centrality
- $\widetilde{bc}$  = betweenness centrality
- $\widetilde{cc}$  = clustering coefficient

## 2.3 Empirical Validation

**Repetition effect:** Repeated exposure increases perceived truth even when contradictory knowledge exists (Fazio et al., 2015). Boundary: semantic processing attenuates effect.

**Emotional amplification:** Moral-emotional language increases diffusion ~20% per word in political tweets (Brady et al., 2017). Platform-dependent.

**Truth penalty:** False news spread faster on Twitter 2006-2017 (Vosoughi et al., 2018). But: exposure  $\neq$  persuasion (Eady et al., 2023).

### 3. Agency Inversion: Who Acts When Ideas Spread?

Neural activity precedes conscious awareness of decision initiation. While interpretations remain contested, the implications are profound:

**Ideas that bypass deliberative evaluation propagate most successfully—not despite our lack of conscious control, but because of it.**

System 1 (automatic) processing dominates before System 2 (deliberative) engagement, particularly when payloads trigger identity threat or coalition costs.

#### The Buffer-Extract-Decide Loop

We name this the **Buffer-Extract-Decide loop** and treat  $A_g$  as its trigger:

**When a payload appears:**

1. **System 1 capture** → Immediate urge to share
2. **Agency check** → Measure  $A_g$
3. **If  $A_g < \text{threshold}$**  → Buffer (30-120 min wait)
4. **Extract partial truths** from the content
5. **Decide consciously** whether to transmit

This loop restores agency by inserting friction exactly where memetic drive overruns deliberation.

The agency inversion—that ideas have us rather than we have them—need not be cause for despair. Once recognized, it becomes a lever for conscious participation. We can observe ideas attempting to use us, evaluate their fitness characteristics, and choose whether to become their transmission vector.

## 4. Holarchic Integration

### 4.1 Development and Transcendence

Cognitive frameworks emerge through developmental stages, each transcending and including the previous:

**Piaget**            Sensorimotor → Preoperational → Concrete → Formal

**Kohlberg**        Preconventional → Conventional → Postconventional

**Kegan**            Impulsive → Imperial → Interpersonal → Institutional → Interindividual

Each stage preserves insights while situating them in more comprehensive frameworks.

### 4.2 Operational Protocol

1. Identify partial truths in each perspective
2. Locate scope conditions for validity
3. Synthesize at higher level without contradiction

This is neither relativism nor absolutism, but **ordered adequacy**.

### 4.3 Application to Defense

**Holarchic integration is not only a synthesis protocol; it systematically reduces identity threat and returns control to the person by reframing opposed claims as scope-conditioned partial truths. Measurably, this raises D and  $A_g$  while lowering effective C.**

When all positions contain partial truths:

- Identity threat decreases → **Higher  $A_g$**
- Coalition costs reduce → **Lower C**
- Synthesis becomes possible → **Conscious choice**
- Manipulation resistance increases → **Agency restored**



## 5. Information Warfare Landscape

### 5.1 Scale $\neq$ Persuasion (But Agency Matters)

Modern operations target *agency bottlenecks*—attention capture, identity threat, frictionless packaging—because these lower  $A_g$  while raising  $P$ . **Exposure  $\neq$  persuasion:** measured attitude or vote shifts are often small or null.

**Evidence checkpoint.** In panel data linking US Twitter users’ feeds to Russian IRA content during 2016, researchers *did not detect a meaningful relationship* between exposure and changes in attitudes, polarization, or voting (Eady et al., 2023).

#### Documented scale vs. measured outcomes

- **Russian IRA scale.**  $\sim 126\text{--}128\text{M}$  US Facebook users reached (2015–2017). *Measured outcomes:* see evidence checkpoint. *Agency view:* low- $A_g$  design (identity threat, high extraction, repetition). Sources: SSCI (2019); platform disclosures; Eady et al. (2023).
- **Chinese ”50c” operations.**  $\sim 448\text{M}$  fabricated social-media comments annually; distraction/cheerleading rather than argument (King et al., 2017). *Measured outcomes:* limited authentic engagement; persuasion not directly evidenced. *Agency view:* noise + repetition depress  $D$ .
- **Algorithmic context.** Independent audits show differential amplification of political content with cross-national asymmetries (Huszár et al., 2022). *Agency view:* ranking rewards fast-path engagement, amplifying low- $A_g$  payloads.

**Takeaway:** Scale buys cheap exposure, not guaranteed persuasion. Agency-reducing designs (low  $A_g$ ) spread widely; agency-restoring interventions (buffers, reframing) protect deliberation even under high exposure.

**How to read scale claims.** *Scale* = reach, impressions, or post volume (documented). *Effect* = measured attitude/behavior change (often small/null). *Agency* =  $A_g$  (lower implies transmission outruns deliberation). Reporting all three avoids the “big numbers = big effects” fallacy.

### 5.2 Structural Vulnerabilities

Information ecosystems exhibit exploitable biases:

- **Novelty bias:** Novel spreads faster regardless of accuracy
- **Emotion bias:** Emotionally charged content amplified
- **Confirmation bias:** Filter bubbles reinforce existing beliefs
- **Availability cascade:** Repetition creates illusion of truth

## 6. Operational Tools

### 6.1 Features → Measures

Component	High-fitness features	Measurable indicators	Agency
A	Simple, concrete	Flesch-Kincaid $< 8$ ; concrete-noun ratio $> 0.4$ ; jargon $< 0.1$	neutral
R	Memorable, distinctive	Distinctive bigrams; rhyme/alliteration; schema-closure	neutral
X	Status-conferring, actionable	Status tokens; action imperatives; identity alignment	$\downarrow A_g$
T	Platform-native, quotable	Characters $\leq 280$ ; visual present; extraction quotient $> 0.3^*$	$\downarrow A_g$
M	Complexity-aligned	$ \Delta  < 1$ on complexity scale	neutral
N	Hub/bridge position	$\tilde{k}$ high; $\widetilde{bc}$ high; $\widetilde{cc}$ context-dependent	neutral
C	Identity-threatening	Sacred-value conflict; public error admission; norm violation	$\downarrow A_g$

\* Extraction quotient = shareable text proportion that preserves claim semantics without context.

**Note:** Features marked  $\downarrow A_g$  typically reduce agency by engaging automatic pathways.

### 6.2 Scoring Rubric

#### Assimilation (A)

- High (+2): Single-sentence summary possible
- Medium (+1): Maps to familiar concepts
- Low (-1): Requires new paradigms

#### Retention (R)

- High (+2): Memorable phrases
- Medium (+1): Useful frameworks
- Low (-1): Abstract without application

#### Expression (X)

- High (+2): Confers status through sharing
- Medium (+1): Supports identity narratives
- Low (-1): Purely theoretical

#### Transmission (T)

- High (+2): Tweet-length, visual
- Medium (+1): Quotable sections
- Low (-1): Requires context

#### Cost (C)

- High (-3): Threatens core identity
- Medium (-2): Requires error admission

- Low (-1): Minor friction

Scores  $>5$  = viral potential;  $>8$  = epidemic spread

### 6.3 Defensive Protocols

#### Individual:

- Monitor Agency Index ( $A_g < 0.5$  = high memetic drive)
- Implement **Buffer-Extract-Decide loop** based on  $A_g$  threshold
- Practice extracting partial truths to raise deliberation (D)
- Track your typical time-to-share baseline

#### Institutional:

- Create cognitive diversity (raises collective  $A_g$ )
- Implement adversarial review (forces deliberation)
- Rotate information sources (prevents capture)
- Measure and report  $A_g$  for critical communications

## 7. Ethical Framework

### 7.1 Constraints

Our ethical constraints are **agency-preserving: transparency, truth-preservation, mutual benefit, and consent** are implemented as design choices that raise  $A_g$  while retaining useful P.

- **Transparency:** Acknowledge persuasive intent (raises D, increases  $A_g$ )
- **Truth-preservation:** Build on accurate partial truths (reduces identity threat)
- **Mutual benefit:** Value for transmitter and receiver (sustainable high- $A_g$  spread)
- **Consent:** Enable conscious participation (maximizes  $A_g$ )

### 7.2 Intended Use

#### Permitted:

- Defensive application against manipulation
- Constructive synthesis across positions
- Research advancement in belief dynamics

#### Prohibited:

- Weaponization for deception
- Exploitation of vulnerable populations
- Amplification of known falsehoods

## 8. Synthesis

Meta-memetics and holarchic integration create synergistic reinforcement:

- **Meta-memetics** reveals why ideas encounter resistance independent of truth (and how they bypass agency)
- **Holarchic integration** provides protocols for reducing resistance while preserving insights (and restoring agency)
- **Together** they enable conscious navigation rather than unconscious subjection

**The synthesis reframes conflicts so that people can notice urges, raise  $A_g$ , and choose—not merely pass along memetic drives.**

This transforms information warfare from zero-sum competition into opportunity for comprehensive understanding, where **high-agency transmission (high  $A_g$ ) can coexist with effective spread (high P).**

## 9. Conclusion

The question is not whether we will be influenced by ideas—we will. **The question is whether we will also influence the ideas that influence us, consciously and constructively.**

Once we recognize that ideas have us, we gain the ability to observe them attempting to use us. We can evaluate their fitness characteristics and choose whether to become their transmission vector.

Holarchic integration offers a path beyond false dichotomies. By recognizing all frameworks as partial truths that can be synthesized at higher levels, we build knowledge that transcends rather than negates.

**Together, these frameworks enable conscious participation in the evolution of collective intelligence.**

## Appendix: Implementation

```
# Estimate propagation probability AND agency index
from sklearn.linear_model import LogisticRegression
import numpy as np

# Features: [logA, logR, logX, logT, logM, logN, C]
interaction = (X[:, 0] + X[:, 1]) * (X[:, 2] + X[:, 3])
X_full = np.column_stack([X, interaction])

clf = LogisticRegression(penalty="l2", C=1.0, max_iter=1000)
clf.fit(X_full, y)
P_hat = clf.predict_proba(X_full)[:, 1]

# Agency Index calculation
D = time_to_share_zscore # Platform-normalized deliberation time
X_expression = np.exp(X[:, 2]) # Expression driver
kappa = 1.0 # Calibration constant

Ag = D / (D + kappa * X_expression)

# Epidemic threshold
d_bar = 150 # Placeholder: replace with empirical mean effective out-degree
R0 = d_bar * P_hat
supercritical = R0 > 1

# Agency-aware decisions
low_agency = Ag < 0.5 # Trigger Buffer-Extract-Decide
high_spread = P_hat > 0.7 # Likely viral
danger_zone = low_agency & high_spread # Maximum vigilance needed
```

**Interpretation:** D uses per-platform z-scored time-to-share; longer than median increases  $A_g$ . When varying M and N in counterfactuals, track how  $A_g$  shifts to reveal which “growth hacks” sacrifice agency for spread.

## References

- King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3), 484-501.
- Senate Select Committee on Intelligence (SSCI). (2019). *Report on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2: Russia's Use of Social Media*. U.S. Senate.
- Brady, W. J., et al. (2017). Emotion shapes the diffusion of moralized content in social networks. *PNAS*, 114(28), 7313-7318.
- Eady, G., et al. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications*, 14, 62. <https://doi.org/10.1038/s41467-022-35576-9>
- Fazio, L. K., et al. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993-1002. <https://doi.org/10.1037/xge0000098>
- Heylighen, F. (1999). What makes a meme successful? Selection criteria for cultural evolution. *Proc 15th Int Congress on Cybernetics*, 418-423.
- Huszár, F., et al. (2022). Algorithmic amplification of politics on Twitter. *PNAS*, 119(1), e2025334119. <https://doi.org/10.1073/pnas.2025334119>
- Kegan, R. (1982). *The Evolving Self*. Harvard University Press.
- Kohlberg, L. (1981). *Essays on Moral Development, Vol. I: The Philosophy of Moral Development*. Harper & Row.
- Piaget, J. (1954). *The Construction of Reality in the Child*. Basic Books.
- Schurger, A., et al. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *PNAS*, 109(42), E2904-E2913.
- Vosoughi, S., et al. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.