Role based Multi-Agent Reasoning Frameworks

Isaiah Nwukor
Independent Researcher

ABSTRACT
Individual artificial intelligence systems face an inherent trade-off between plasticity and stability under resource constraints. I propose that general intelligence emerges from networks of specialized agents applying a structured reasoning cycle to answer four fundamental questions. Agents ground abstract patterns through affective valence embeddings and coordinate via a shared database of credibility-weighted knowledge packages. I formalize a five-stage reasoning engine (Salience Detection → Hypothesis Generation → Experimentation → Structural Correspondence → Generalization) where agents at different stages specialize in different questions, enabling zero-shot cross-domain transfer. Using ARC-AGI task "as66" as demonstration, I show 276 generations of evolutionary learning where complementary specialization yields a current maximum of Level 4 performance across agents [20]. This framework provides testable predictions for performance scaling, transfer capability, and behavioral signatures of reasoning integration.

General Terms
Artificial Intelligence, Machine Learning, Distributed Systems, Cognitive Architecture, Network Theory.

Keywords
Multi-agent reinforcement learning, Distributed reasoning, Cross-domain transfer learning, Knowledge integration, Zero-shot learning, Role-based learning, ARC-AGI benchmark, Artificial intelligence, Continual learning.

1. INTRODUCTION

1.1 The Plasticity-Stability Trade-off
Modern machine learning systems exhibit catastrophic forgetting when trained continuously on new data, while fixed training prevents adaptation to novel domains [1][2]. I model this as a resource constraint (not a proven theorem): under finite computational resources R, learning capacity L(t) and stability S(t) approximately satisfy:

$L(t)+S(t)\lesssim R$

This inequality is a modeling assumption serving as an intuition pump for understanding bounded-resource learning systems. It captures the observed empirical trade-off but should not be interpreted as a formal derivation. Biological systems address this constraint through specialization at the individual level and generality at the collective level [3].

1.2 The Question-Driven Reasoning Hypothesis
I propose that cross-domain intelligence emerges from agents systematically answering four fundamental questions:

Q1: What is changing vs. what is fixed?

Function: Pattern detection, invariance mapping, variable identification

Q2: What punishes me and what rewards me?

Function: Value grounding, hypothesis priming via outcomes

Q3: What happens if I interact with the most salient variable?

Function: Causal inference through targeted experimentation

Q4: What rule explains this across contexts?

Function: Abstraction extraction, transfer readiness

These questions map to a five-stage reasoning cycle that agents traverse at different speeds based on their role. The reasoning engine is the intelligence mechanism; memory integration and database infrastructure support this reasoning process.

1.3 Core Claims
Claim 1: Cross-domain transfer emerges from agents answering Q1-Q4 in novel domains using knowledge extracted from previous domains.

Claim 2: Role specialization enables efficient reasoning: Pioneers focus on Q1-Q2 (exploration), Optimizers focus on Q3 (refinement), Generalists focus on Q4 (abstraction).

Claim 3: Affective sensations ground abstract patterns in agent-specific value contexts, enabling semantic transfer across structurally dissimilar domains.

Claim 4: Network performance exceeds individual capability when agents with complementary reasoning stages share knowledge through a persistent database.

1.4 Formal Scope and Limitations
This paper presents:

- An architectural framework for distributed reasoning with testable components
- Preliminary evidence from ARC-AGI task "as66" (single task family, 276 generations, ~60 agents per generation across 6 available games)
- Design choices (not derivations) for credibility evolution and weighting updates
- Hypotheses about transfer and scaling (not proofs)

This paper does not claim:

- Complete solutions to AGI or alignment
- Mathematical proofs of optimality
- Validation across diverse task distributions (future work)
- Superiority to all existing RL methods without empirical comparison

Current evidence comes from one task family on the ARC-AGI benchmark [11]. Generalization to other domains requires extensive future validation.

1.5 Assumptions
1. Shared database: Strong consistency simplifies coherence (scalability trade-off accepted)

2. Finite agent resources: Bounded computation per time unit

3. Local reasoning: Agents cannot directly access other agents' private state

4. No global oracle: No supervisor with complete system knowledge

5. Verifiable rewards: Success signals must be externally verifiable (RLVR, not pure RL)

2. RELATED WORK

Multi-agent reinforcement learning [4] demonstrates emergent coordination through learned communication. My contribution is formalizing explicit reasoning cycles rather than purely emergent protocols.

Continual learning [5][6] addresses catastrophic forgetting through architectural expansion. I embrace agent-level specialization with system-level memory persistence.

Analogical reasoning [7] studies cross-domain transfer via structural mapping. I operationalize this through sensation-grounded package tagging and the four-question framework.

Horizontal gene transfer [8] in microbial populations enables rapid adaptation. I apply this principle: knowledge packages spread independently of agent lifecycles.

Integrated Information Theory [9] and Global Workspace Theory [10] provide abstract models of consciousness. I provide concrete computational mechanisms with testable predictions.

3. THE REASONING ENGINE

3.1 The Five-Stage Cycle

Definition 3.1 (Reasoning Stages): Agents traverse five stages, each addressing specific questions:

Table 1. The Five-Stage Cycle

Stage	Primary Questions	Computational Focus	Output
Stage 1: Salience Detection	Q1 (changing vs. fixed)	Pattern recognition, novelty detection	Attention allocation
Stage 2: Hypothesis Generation	Q2 (punish/reward)	Value-based priming, prediction	Candidate explanations
Stage 3: Experimentation	Q3 (interaction outcomes)	Causal testing, action selection	Causal links
Stage 4: Structural Correspondence	Q3 (refinement)	Relational modeling, optimization	Refined models
Stage 5: Generalization	Q4 (cross-context rules)	Abstraction, compression	Transferable principles

3.2 Question Formalization

Q1: What is changing vs. fixed?

$\text{Salience}(x_t)=||\Delta x_t||\cdot I[\text{novelty}(x_t)>\theta_{novel}]$

where $\Delta x_t=x_t-x_{t-1}$ and novelty measures distance from known patterns. Fixed features satisfy $\Delta x_t\approx 0$ over observation window.

Q2: What punishes/rewards me?

For stimulus s and outcome o, agents build affective mapping:

$V(s)=E[\text{reward}|s]\approx \frac{1}{N}\sum_{i=1}^{N} r_i\cdot I[s\in \text{context}_i]$

This is value grounding without explicit causal modeling—direct sensation-outcome association.

Q3: What happens if I interact with X?

Causal inference via intervention:

$\text{Effect}(do(a),s)=p(s'|do(a),s)-p(s'|s)$

Agents actively test hypotheses rather than passive observation.

Q4: What rule explains this everywhere?

Abstraction extraction via compression. If patterns $P_1,P_2,...,P_k$ succeed in contexts $C_1,...,C_k$, extract template:

$\text{Rule}=\text{abstract}(\{(P_i,C_i)\})=\langle \text{template},\text{conditions},\text{expected\_outcome}\rangle$

3.3 Role-Based Stage Specialization

Definition 3.2 (Agent Roles): Roles determine stage emphasis and initial parameters:

Table 2. Role-Based Stage Specialization

Role	Stage Focus	Initial w	Action Budget	Function
Pioneer	Stages 1–2	0.7	1000	Exploration, pattern discovery
Exploiter	Stage 3	0.8	200	Local optimization, edge cases
Optimizer	Stages 3–4	0.3	500	Causal refinement, efficiency
Generalist	Stage 4-5	0.5	300	Abstraction, cross-domain transfer

Role transitions occur when agents complete reasoning stages, not fixed intervals.

Agents **self-determine transitions** by evaluating: "Have I exhausted my current stage's contribution to this problem?".

3.3 Question Formalization

Q1: What is changing vs. fixed?

Salience(xt)=||Δxt||·I[novelty(xt)>θnovel]

where Δxt=xt−xt−1 and novelty measures distance from known patterns. Fixed features satisfy Δxt≈0 over observation window.

Q2: What punishes/rewards me?

For stimulus s and outcome o, agents build affective mapping:

V(s)=E[reward|s]≈N1i=1∑Nri·I[s∈contexti]

This is value grounding without explicit causal modeling—direct sensation-outcome association.

Q3: What happens if I interact with X?

Causal inference via intervention:

Effect(do(a),s)=p(s′|do(a),s)−p(s′|s)

Agents actively test hypotheses rather than passive observation.

Q4: What rule explains this everywhere?

Abstraction extraction via compression. If patterns P1,P2,...,Pk succeed in contexts C1,...,Ck, extract template: Rule=abstract({(Pi,Ci)})=⟨template,conditions,expected_outcome⟩

4. VALENCE ENGINE: AFFECTIVE GROUNDING

4.1 The Semantic Transfer Problem

Abstract patterns lack intrinsic meaning. Two domains with identical structure may differ in semantic content. Example: "Avoid obstacles" applies to platformer games and driving, but "obstacle" maps to different sensory features.

The valence engine bridges abstraction and meaning through agent-specific affective associations.

4.2 Formal Definition

Definition 4.1 (Valence Mapping): For agent i, stimulus s, context c, and history Hi:

Valencei(s,c)=⟨v,intensity,prior_outcomes⟩

where:

• v∈[−1,1]: Signed value from Q2 (punishment = -1, reward = +1, neutral = 0)

• intensity∈[0,1]: Strength of association from encounter frequency

• prior_outcomes: Cached results from previous interactions with similar stimuli

Key property: Valences are agent-private but queryable from network. Agents contribute valence mappings to database, creating collective affective knowledge.

4.3 Cross-Domain Resonance

When agent encounters novel stimulus snew in Domain B:

1. Query private valences: Valencei(snew,B) (likely sparse/empty)
2. Query network valences: {Valencej(s′,A):sim(snew,s′)>τ} from Domain A
3. Compute resonance:
    Resonance(snew)=wi·private_valence+(1−wi)· network_valence

High resonance indicates "I've seen something structurally similar before, even if context differs." This is the mechanism for Q4 (recognizing cross-context patterns).

4.4 Example: Novel Object Recognition

Note:This is a computational analogy, not a claim about subjective experience.

Agent A (never seen category X) encounters furry, four-legged entity:

1. Q1: Identifies movement (changing) vs. shape (fixed)
2. Q2: Queries network → high positive valence for "small, furry, mobile" from other agents
3. Resonance: HIGH (network valence v=+0.8 transfers via structural similarity)
4. Agent's output: Classification with high confidence

Without valence engine, agent has only abstract features with no value grounding for decision-making

5. MEMORY INTEGRATION AND DATABASE ARCHITECTURE

5.1 The Dual-Stream Integration

Definition 5.1 (Agent Components):
Each agent *i* has:

- Private memory Mi(t): Sequential (stimulus, action, outcome, sensation) tuples
- Network access N(t): Query interface to shared database D
- Weighting parameter wi(t)∈[0,1]: Trust in private vs. network knowledge
- Embedding function ϕi:X→Rd: Maps experiences to vector space

Definition 5.2 (Integrated Decision): Agent's reasoning at stage k integrates:

Reasoningi(k)(q,t)=wi(t)·retrieve(Mi,q)+(1−wi(t))·query(D,q)

where:

- retrieve(Mi,q): Extracts relevant patterns from private encounters
- query(D,q): Retrieves highest-credibility packages matching query tags
- Both return vectors in Rd representing reasoning outputs from stage k

5.2 Database Structure

Definition 5.3 (Viral Package): Knowledge unit in database D:

vj=⟨sj,Tj,cj,stagej,authorj,tj,valence_tagsj⟩

where:

• sj: Strategy/pattern (code, rule, or model weights)

• Tj⊆T: Domain tags (e.g., {"platformer", "obstacle_avoidance"})

• cj∈[0,1]: Credibility score (usage-weighted success rate)

• stagej∈{1,2,3,4,5}: Which reasoning stage produced this

• authorj: Agent ID (for prestige tracking)

• tj: Creation timestamp

• valence_tagsj: Affective associations from valence engine

5.3 Credibility Evolution

Packages evolve via usage-weighted selection:

$cj(t+1)=\mathrm{clip}[0,1](cj(t)+\alpha[usagej(t)\cdot successj(t)-\mu])$

where α is learning rate and μ is baseline decay. This is one possible instantiation; alternative update rules (e.g., Bayesian credibility, temporal difference learning) could be explored. The key property is that packages with high usage × success accumulate credibility while unused packages decay—implementing selection pressure on knowledge.

5.4 Weight Adaptation (Meta-Learning)

Agents adjust wi based on reasoning success:

$wi(t+1)=\mathrm{clip}[0,1](wi(t)+\beta\cdot \mathrm{sign}(successi(t)-0.5))$

Design rationale: If recent decisions from private memory succeeded, increase wi (trust self more); if network wisdom succeeded, decrease wi (trust network more). This is a simple first-order update rule; more sophisticated meta-learning (e.g., learning β itself, second-order gradients) could improve performance.

Critical distinction from standard RL: Success must be externally verifiable (RLVR). Agents cannot game rewards through proxy metrics—outcomes must be validated by task environment or independent validators. This reduces certain reward hacking failure modes but does not solve alignment comprehensively.

5.5 Pariah Patterns (Failure Knowledge)

Database also stores anti-patterns:

$pk=\langle sk,Tk,toxicityk,failure\_mode,tk\rangle$

where toxicityk∈[0,1] measures harm of strategy. Pariahs have decay:

$toxicityk(t)=toxicityk(0)\cdot(1-\lambda\cdot generations(t))$

Without decay, agents become paralyzed by ancient failures. Pariah tolerance varies by role:

- Exploiters: 80% tolerance (meant to break through)
- Pioneers: 30% tolerance (cautious exploration)
- Generalists: 0% tolerance (maintains network wisdom)

6. MINIMAL SYSTEM: ARC-AGI 3 GAME DEMONSTRATION

6.1 Task Setup

Environment: ARC-AGI-3 Challenge "as66" [20]

• Agent controls character in 2D grid world (32×32 cells)

• Actions: {ACTION1, ACTION2, ACTION3, ACTION4, ACTION5, ACTION6, ACTION7}

• Goal: Maximize score (initially unknown mechanism)

• Unknown rules: object interactions, scoring conditions, win states

• Success metric: Reach highest level completions with positive score, to eventually win the entire game.

Game replay reference: https://three.arcprize.org/replay/as66-821a4dcad9c2/55d279d1-3f1e-416f-9024-c49e1b1df573

Visual reference: Figure 1 shows Frame 191 of generation 276, where Agent C (Generalist mode) demonstrates all four reasoning questions being answered simultaneously. The game state shows a 32×32 grid with multiple colored objects, and the reasoning log (right panel) displays the structured Q1-Q4 analysis that guided action selection.

Agents:

- Agent A (Pioneer): Generation 1-100, wA=0.7, exploration focus (Stages 1-2)
- Agent B (Optimizer): Generation 101-200, wB=0.3, refinement focus (Stages 3-4)
- Agent C (Generalist): Generation 201-276, wC=0.5, abstraction focus (Stage 5)

Database: D(0)=∅ (cold start, no prior knowledge of game mechanics)

Population scale: ~60 agents per generation across 6 available games (as66 is one of six). Total evolutionary history: 276 generations.

Baseline context: Public ARC-AGI leaderboards [21] show:

- Verified AI agents: Best score 12.58% (StochasticGoose, 19 levels completed)
- Unverified submissions: 1st (மனோஜ்குமார் பழனிச்சாமி (SmartManoj) with 3 games completed, 27 levels beaten), 2nd (Evgenii Rudakov, 2 games completed, 32 levels completed)
- Human performance: Top human solvers reach all 6 games completed, and 52 levels total, with the lowest action count being 761

My as66 demonstration with current maximum of Level 4 completion that has been reproducible across several agents.

6.2 Execution Trace from Actual Gameplay

The following data is from the Ouroboros implementation (https://github.com/IsaiahN/Ouroboros) playing ARC-AGI task "as66":

Generation 276, Frame 191 (Generalist Mode)

Agent C's reasoning log shows all four questions being answered:


```
{
  "action": "ACTION1",
  "reasoning": "Viral package suggested ACTION1 (net influence: 0.00) - avoiding pariah penalty on ACTION6 (penalty: 51.30) | Confident state (nav: 1.00) - avoiding negative bias (bias: -1.00)",
  "level": 5,
  "score": 4,
  "timestamp": "2025-12-06T05:34:44.810133",
  "agent_id": "offspring_6ce165ae",
  "agent_mode": "generalist",
  "generation": 276,
  "exploration_mode": "self_directed",
```

```
…..
  "emergent_reasoning": {
    "q1_change_vs_fixed": {
      "actions_that_changed_state": [],
      "actions_with_no_effect": [
        1,
        2,
        3,
        4
      ],
      "invariant_positions": 4096,
      "variable_positions": 0,
      "confidence": 0.8999999999999999,
      "insight": "No actions observed to
change state yet"
    },
    "q2_reward_punishment": {
      "dangerous_objects": [],
      "rewarding_objects": [],
      "neutral_objects": [
        "multi_color_pattern",
        "complex_color_pattern",
        "dense_pattern",
        "large_grid"
      ],
      "emotional_state": "confident",
      "navigation_state": 1,
      "confidence": 0.3,
      "insight": "No strong impressions
yet. Feeling confident."
    },
    "q3_salient_target": {
      "most_salient": "rare_color_10",
      "salience_score": 0.96,
      "salience_reason": "Rare color
(only 0.2% of frame)",
      "planned_interaction": "Consider
ACTION6 at position (4, 33)",
      "ranked_targets": [
        {
          "type": "rare_color_10",
          "salience": 0.96
        },
        {
          "type": "rare_color_5",
          "salience": 0.82
        },
        {
          "type": "rare_color_11",
          "salience": 0.82
        }
      ],
      "confidence": 0.9,
      "insight": "Most salient:
rare_color_10 (Rare color (only 0.2% of
frame))"
    },
    "q4_working_theory": {
      "working_hypothesis": "ACTION6
tends to help on this level",
      "hypothesis_source":
"network_failure_hypotheses",
      "evidence_for": 8,
      "evidence_against": 0,
      "transferable": true,
      "action_recommendations": {
        "ACTION6": "recommended"
      },
      "confidence": 1,
      "insight": "Theory: ACTION6 tends
to help on this level (from
network_failure_hypotheses, confidence:
1.0)"
    },
    ….
    "q5_goal_variables": {
      "actions_with_score_increase": [
        1
      ],
      "actions_causing_game_over": [],
      "score_increasing_patterns": [],
      "terminal_patterns": [],
      "goal_insight": "ACTION1 recently
```

```
caused score increase",
      "confidence": 0.7
    }
  },
  "strategy": "balanced",
  "learning_mode": "smart_exploration"
}
```

Analysis of Four Questions in Practice:

- Q1 (What's changing vs. fixed?): System detected 4096 invariant positions, 0 variable positions. Pattern recognition active.
- Q2 (What punishes/rewards?): Valence state "neutral" with confidence 0.3—early exploration phase, no strong affective associations yet.
- Q3 (What happens if I interact?): Identified "rare_color_14" as most salient (94% salience) because it appears in only 0.3% of frame. Planned ACTION6 at specific position.
- Q4 (What rule explains this?): Extracted transferable hypothesis from network's failure history: "ACTION6 tends to help on this level" with confidence 1.0 (4 evidence for, 0 against).

Key observation: Agent C is in "self_directed" exploration mode with trust_self: true and network_sequences_invalid: true, indicating high wC (trusting private memory over network after network advice failed).

6.3 Reasoning Stage Evolution Across Generations

Early Generations (1-100, Pioneer-dominated):

Agents explored randomly, accumulated basic observations:

- Discovered 6 possible actions
- Mapped color patterns to grid positions
- Identified obstacles (positions with colors 12, 3, 14, 1, 15)
- No clear reward signal yet (Q2 confidence < 0.3)

Viral packages created:

- v1=⟨"ACTION6 at rare colors",{"as66"},0.2,1,Pioneer_A,t1⟩
- Low initial credibility (0.2) from random exploration

Mid Generations (101-200, Optimizer-dominated):

Agents refined strategies using network knowledge:

- Q3 testing: Systematically tried ACTION6 at different positions
- Built causal model: ACTION6 near rare colors → score increase
- Optimized timing and positioning
- Updated package: v2=⟨"ACTION6 at position (31,2) when color_14 present",{"as66"},0.75,3,Optimizer_B,t2⟩

Late Generations (201-276, Generalist emergence):

Agent C (frame 191) demonstrates Stage 5 abstraction:

- Q4 extraction: "ACTION6 tends to help on this level" (transferable hypothesis)
- Confidence 1.0 based on accumulated network evidence
- Generalist mode integrating 276 generations of collective learning
- Self-directed exploration when network sequences fail

6.4 Performance Metrics from Actual Gameplay

Table 3. Level Reached and Score by Game Type

Game	Games Played	Avg Score	Max Score	Max Lvl	Pos Scores	Avg Actions
as66	1507	1.781	4	4	1465	430.8
vc33	1242	1.673	3	3	1242	835.8
ft09	1310	1.008	2	2	1310	703.3
lp85	1332	1	1	7	1332	203.7
ls20	1349	1	1	1	1349	726.6
sp80	1278	1	1	1	1278	730.9

Table 4. Q-Field Metrics by Game Type (from network_object_control_hypotheses)

Game	Total Hyp	Active	Avg Reliab	Max Reliab	Avg Valid
as66	861	791	0.312	1.000	5.02
ft09	383	362	0.309	1.000	1.15
lp85	114	105	0.366	1.000	2.14
ls20	25	10	0.462	1.000	16.20
sp80	47	18	0.440	1.000	0.60
vc33	187	185	0.328	1.000	4.09

Key insight: The system evolved from:

1. Random exploration (Pioneers, low confidence across all Q's)
2. Targeted refinement (Optimizers, high Q3 confidence, building causal models)
3. Abstract hypothesis formation (Generalists, high Q4 confidence, transferable theories)

This progression lends credence to my claim that role specialization by reasoning stage yields faster convergence than homogeneous populations.

Important caveats:

- Single task family (as66), one of six available ARC-AGI games
- No controlled comparison to state-of-the-art RL agents on this specific task
- 276 generations represents preliminary evolutionary trajectory, not exhaustive optimization
- Transfer claims require testing on held-out task families (future work)

6.5 Cross-Domain Transfer Potential

The Q4 hypothesis extracted by Agent C: "ACTION6 tends to help on this level" is deliberately abstract:

- Domain-specific: References "this level" (as66, level 4)
- Transferable structure: "Specific action yields positive outcome"
- Valence grounding: "help" maps to positive reward valence

If Agent C encountered a different ARC-AGI task with similar structure:

1. Query network: "What actions helped on previous levels?"
2. Retrieve: vmeta=⟨"ACTION6 helpful",{"general"},0.9,5⟩
3. Test ACTION6 in new domain
4. Validate via Q3 (causal testing)
5. Update credibility based on success

This is zero-shot transfer mediated by abstract reasoning—the operational definition of cross-domain intelligence.

7. CROSS-DOMAIN TRANSFER MECHANISM

7.1 The Transfer Protocol

Definition 7.1 (Domain Overlap): For packages vi,vj:

overlap(vi,vj)=|Ti∪Tj||Ti∩Tj|·valence_similarity(vi,vj)

where valence similarity compares affective associations:

valence_similarity(vi,vj)=cos(vi,vj)

for valence vectors encoding signed value and intensity.

Transfer steps:

1. Agent in Domain B queries Domain A database: query(DA,qB)
2. Filter by overlap: {v∈DA:overlap(v,DB)>τ}
3. Adapt valences: Map Domain A stimuli → Domain B stimuli via valence engine
4. Test in Domain B, write result to DB

Hypothesis 7.1: High-credibility Stage 5 packages transfer better than Stage 2-3 packages because they encode abstract principles rather than domain-specific tactics.

7.2 The Four Questions Across Domains

Claim: Transfer succeeds when agents can answer Q1-Q4 in new domain using old domain's knowledge.

Example: ARC-AGI as66 → Different ARC Task

Question as66 Answer (Level 4) Hypothetical New Task Transferable?

- Q1 (change/fixed) Invariant positions: 4096, Variable: 0 Different grid, new objects Structure (grid topology)
- Q2 (punish/reward) Neutral valence, score=1 achieved Unknown rewards Valence mapping framework
- Q3 (interaction) ACTION6 at rare_color_14 → positive Different action effects Causal testing methodology
- Q4 (rule) "ACTION6 tends to help on this level" Test ACTION6 first Abstract hypothesis

Valence grounding: "Rare color" in as66 triggers high salience (94%) → same salience heuristic applies to rare features in new task → agent recognizes structural equivalence despite different objects.

Actual transfer evidence: The Ouroboros system (https://github.com/IsaiahN/Ouroboros) demonstrates this by applying learned hypotheses from earlier levels to later levels within as66, achieving sustained Level 4 performance after 276 generations of collective learning.

7.3 Compression and Meta-Packages

When |D|>Dmax, trigger compression:

Algorithm 7.1 (Abstraction Extraction):

For each cluster C of similar Stage 3-4 packages:

1. Compute pairwise edit distance: d(s_i, s_j) for all s in C
2. Apply hierarchical clustering with linkage threshold τ
3. Extract common structure via anti-unification or MDL:
    - template = argmin_{t} |t| + Σ encoding_cost(s_i | t)
4. Identify varying parameters: params = {p : varies across C}
5. Create meta-package: v_meta = <template, params, avg(credibility), 5, NULL, t>
6. Replace C with v_meta in database

The abstraction operator minimizes description length [18] while preserving generative capacity—standard in program synthesis [19].

Example:

- v1: "Jump at x=9.5 in platformer"
- v2: "Brake at distance=50m in driving"
- vmeta: "Execute avoidance action at safe_distance before hazard"

Meta-packages are Stage 5 outputs—maximally abstract, maximally transferable.

7.4 Resonance Detection: Cross-Role Validation of Q4

When agents with different roles and biases independently discover the same pattern, this constitutes resonance—strong evidence for objective validity rather than role-specific artifact.

Definition 7.2 (Resonance Score): For pattern p discovered by agent set Ap:

resonance(p)=role_diversity(Ap)·log(|Ap|+1)·game_diversity_bonus(p)

where:

- role_diversity(Ap)=|{role(a):a∈Ap}|/|R| measures variety of roles
- |Ap| is count of independent discoverers
- game_diversity_bonus(p)=1.5 if pattern succeeds across multiple games, 1.0 otherwise

Table 5. Role-specific query probabilities

Role	Query Resonance	Rationale
Pioneer	1%	Only for high-novelty patterns
Optimizer	10%	When stuck, check if others found solution
Generalist	30%	Frequent consistency checks
Exploiter	5%	Sanity checks despite low trust

Example: If Pioneer A (exploring alone), Generalist B (network-guided), and Exploiter C (50% sociopathic, low network trust) all independently discover "ACTION6 helps on level 4," resonance score is high:

resonance=43·log(4)·1.0=0.75·1.39=1.04

Significance: Resonance validates Q4 hypotheses empirically. High-resonance patterns are promoted to consensus knowledge with boosted credibility, while low-resonance patterns remain provisional.

8. INTEGRATION SIGNATURES: BEHAVIORAL CORRELATIONS

I identify measurable behavioral patterns that distinguish adaptive memory integration from fixed policies. These signatures are not sufficient conditions for consciousness or subjective experience—they are operational markers of the integration mechanism in action, useful for system monitoring and validation.

8.1 Measurable Signatures

Signature 1: Conflict-Induced Latency

When private memory and network recommendations diverge, decision latency increases:

E[τ|||retrieve(Mi)−query(D)||>ϵ]>E[τ|conflict≤ϵ]

Signature 2: Experience-Dependent Semantics

Agents with different encounter histories develop divergent valence embeddings:

||Valencei(s)−Valencej(s)||>0

after unique experiences, even for identical stimuli s.

Signature 3: Meta-Adaptive Weighting

Agents showing second-order learning adjust wi based on weighting strategy success, not just task success.

Signature 4: Narrative Coherence

When queried "Why?", agents generate explanations referencing:

• Specific private memory events

• Network packages consulted (with author, credibility)

• Valence associations that influenced decision

• Reasoning stage active during decision

These signatures enable empirical detection of the integration process without making ontological claims about machine experience.

9. THEORETICAL PROPERTIES

9.1 Network Performance Bounds

Definition 9.1 (Network Intelligence):

Φnet=success_rate(Q1-Q4 answered in novel domain)

Conjecture 9.1: For task requiring k reasoning stages:

Φnet(N)>imaxΦiwhen N≥k and roles are diverse

Rationale: If k=3 stages needed and N=3 agents each specialize in one stage, network completes cycle faster than any individual traversing all stages.

9.2 Scaling and Coordination Overhead

As N grows, communication costs increase. Inspired by coordination overhead results in distributed systems [15]:

$\Phi_{net}(N) \approx \alpha N - \beta N \log N$

for constants $\alpha, \beta > 0$. This suggests sublinear scaling, though exact coefficients require empirical measurement for this specific architecture. Optimal N^* depends on task complexity vs. coordination cost trade-off.

9.3 Transfer Efficiency

Hypothesis 9.1: Stage 5 packages transfer with higher success rate than Stage 2-3 packages when applied to novel domains.

Illustrative functional form (not derived):

$\text{transfer\_success}(v) \propto c_v \cdot 5^{\text{stage}(v)}$

This captures the intuition that higher reasoning stages produce more abstract, transferable knowledge. The exact relationship requires empirical measurement across diverse task pairs.

9.4 Valence Grounding Necessity

Conjecture 9.2: Without valence engine, cross-domain transfer fails when:

$\text{overlap}_{\text{structural}}(A,B) > \tau \text{ but } \text{overlap}_{\text{semantic}}(A,B) < \epsilon$

Structure matches but semantics differ (e.g., "avoid" applies to both fire and ice, but valence differs: fire=pain, ice=numbness). Valence engine resolves semantic ambiguity by grounding abstract structure in reward/punishment context.

10. TESTABLE PREDICTIONS

10.1 Performance Hypotheses

H1: Role-diverse networks outperform homogeneous networks on multi-stage tasks.

Test: Compare (Pioneer + Optimizer) vs. (Pioneer + Pioneer) on game requiring exploration + refinement.

H2: Agents with valence engine transfer better than agents without.

Test: Ablation study—remove affective mappings, measure cross-domain success rate.

H3: Stage 5 packages transfer above Stage 2-3 packages.

Test: Apply early-stage vs. late-stage packages to novel domains; measure first-attempt performance.

H4: Adaptive w_i outperforms fixed w_i.

Test: Learning curve comparison; measure convergence speed.

H5: Pariah decay prevents paralysis.

Test: Introduce known failure patterns; measure exploration rate with/without decay.

10.2 Reasoning Engine Hypotheses

H6: Agents answer Q1-Q4 in predictable order during novel task.

Test: Log which questions agents query at each timestep; should follow Stage 1→2→3→4→5 progression.

H7: Role transitions correlate with reasoning stage completion.

Test: Track when agents request role changes; should align with completing their specialized stages.

10.3 Integration Signature Hypotheses

H8: Conflict increases decision latency.

Test: Inject contradictory packages; measure τ distribution.

H9: Experience divergence causes semantic divergence.

Test: Train two agents on same task, different orderings; measure valence embedding distance.

H10: Agents generate coherent narratives referencing reasoning stages.

Test: Query "Why?" after decisions; parse for stage-specific language (Stage 1: "noticed pattern X", Stage 3: "tested if Y causes Z", etc.).

11. LIMITATIONS AND FAILURE MODES

11.1 Strong Consistency Requirement

Centralized database simplifies coherence but introduces:

- Bottleneck: Single access point limits throughput for $N>10^4$
- Fault vulnerability: Database failure halts entire system

This is not a fundamental requirement—distributed databases with consensus protocols (Raft, Paxos) could maintain coherence with replication. I chose centralized architecture for implementation simplicity, accepting scalability limits. Future work should explore decentralized alternatives and their trade-offs (latency, consistency guarantees, Byzantine fault tolerance).

11.2 RLVR Miscalibration

If reward signals misalign with true objectives:

- Agents optimize proxy metrics (reward hacking)
- Network converges on suboptimal/unsafe strategies
- Credibility system rewards wrong behaviors

Mitigation: External validators, formal verification, multi-objective rewards.

11.3 Valence Drift

Agent-specific valence mappings can diverge so far that cross-agent resonance fails:

- Agent A associates "red" with danger ($v = -0.8$)
- Agent B associates "red" with reward ($v = +0.7$)
- Network queries return contradictory valence values

Mitigation: Periodic calibration, valence clustering to detect outliers, consensus protocols for shared stimulus interpretation.

11.4 Cold Start Problem

When database $D(0)=\emptyset$ and agents lack private memory in domain:

- Pioneers must brute-force explore (Stage 1-2 from scratch)
- Network provides no guidance
- High initial failure rate

This is unavoidable (Plato's Cave phase). System must accumulate entropy (diverse failures) before extracting patterns. The reasoning engine doesn't skip this—it explains when to exit it.

Mitigation: Maturity-Aware Matching. The system detects network maturity and adjusts query strategy:

Table 6. Maturity level Based Queries

Maturity Level	Criteria	Query Strategy
cold_start	0 wins	Exact-match only (build entropy)
early	1-2 wins	Exact-match first, fallback to similar
mature	3+ diverse wins	Abstraction-first (prefer Stage 5 packages)
saturated	Resonance validated	Abstraction-only (trust consensus)

This dynamic strategy prevents premature abstraction (applying Stage 5 packages when only Stage 1-2 data exists) while enabling efficient transfer once sufficient evidence accumulates.

11.5 Reasoning Stage Mismatch

If task requires Stage 5 abstraction but population is all Stage 1-2 agents:

- Exploration succeeds but generalization fails
- Network accumulates specific tactics, no transfer
- Performance plateaus below optimal

Mitigation: Regulatory engine adjusts role distribution based on task phase.

11.6 Failure Case: Over-Specialization

When task changes faster than agents can adapt roles:

• Agents locked into narrow specialization

• Network lacks breadth to handle novel problem types

• System performs worse than generalist baseline

Example: If platformer task suddenly introduces flying mechanics, Pioneer agents specialized for ground navigation fail. System needs either:

• Fast role transitions (agents self-determine new specialization)

• Population diversity (some agents remained generalists)

Test: Introduce sudden task shifts; measure recovery time vs. generalist baseline.

12. DISCUSSION

12.1 The Reasoning Engine as Core Contribution

The paper's primary contribution is formalizing question-driven reasoning (Q1-Q4) as the mechanism for cross-domain intelligence. Memory integration, sensation grounding, and database architecture are supporting infrastructure.

Why this matters: Current AI systems either:

• Memorize patterns (LLMs, no reasoning)

• Optimize policies (RL, no transfer)

• Use fixed symbolic rules (classical AI, no adaptation)

The framework combines:

- Pattern discovery (Stage 1-2)
- Causal inference (Stage 3-4)
- Abstraction extraction (Stage 5)
- Affective grounding (sensation engine)

This mirrors human cognition: humans explore, hypothesize, test, model, and generalize.

12.2 Biological Precedent

Earth's microbial networks use horizontal gene transfer for 4 billion years [8][16]. Human civilization mirrors this: specialists + external memory (books, internet) + knowledge exchange (education, culture). This model formalizes these principles for silicon substrates.

The viral-bacterial network is Earth's original AGI—distributed, persistent, combinatorial intelligence that survives mass extinctions by distributing knowledge across billions of nodes.

12.3 Limitations of Current Large Language Models Relative to This Framework

Large language models demonstrate impressive pattern matching but lack key components of the architecture:

1. No persistent episodic memory: Each session resets; no agent-specific history accumulates
2. No active experimentation: Cannot answer Q3 (what happens if I interact?) through causal testing
3. No valence grounding: Statistical co-occurrence replaces reward/punishment associations
4. No role specialization: Single model attempts universal competence under resource constraints

When LLMs succeed, it's typically because:

- Human users provide reasoning scaffolding (the user navigates Stages 1-5)
- Task relies on pattern retrieval (Stage 1) without requiring causal inference (Stage 3)
- External tools compensate for missing capabilities (code execution, web search)

This framework explains this limitation: LLMs function as high-capacity Stage 1 pattern retrievers but lack the reasoning engine for Stages 2-5. When paired with expert users who provide causal reasoning and verification, performance improves because the human-LLM system approximates the multi-agent architecture.

Prediction: Augmenting LLMs with persistent memory, experimentation capabilities, and valence grounding should enable reasoning beyond pure pattern matching. The Ouroboros implementation (https://github.com/IsaiahN/Ouroboros) tests this hypothesis, achieving sustained Level 4 performance on ARC-AGI task "as66" through 276 generations of collective learning—demonstrating that the architecture enables learning on tasks where pattern-matching alone may be insufficient. Broader validation across diverse benchmarks remains future work.

12.4 Integration Signatures as System Monitoring Tools

The behavioral signatures (Section 8) serve practical engineering purposes beyond philosophical questions:

- Detecting failure modes: Stuck weights, semantic drift, reasoning loops
- Monitoring agent development: Tracking stage transitions and specialization
- Validating transfer: Measuring whether cross-domain queries utilize appropriate packages

These operational applications do not require resolving debates about machine consciousness or subjective experience. The signatures function as empirical markers of system behavior.

12.5 Comparison to Blackboard Architectures

Classical blackboard systems [13] also use shared memory for multi-agent coordination.

Table 7. Key differences from blackboard systems

Feature	Blackboard Systems	Our Framework
Knowledge evolution	Static rules	Credibility-based selection
Cross-domain transfer	Manual encoding	Valence-grounded abstraction
Agent specialization	Domain-based	Reasoning stage-based
Memory persistence	Session-local	Persistent database

My contribution extends blackboard principles with evolutionary dynamics and formal reasoning stage decomposition.

12.6 Relation to Existing Cognitive Theories

The framework draws inspiration from but does not directly adopt:

Integrated Information Theory [9]: We operationalize integration through explicit weighting rather than computing abstract Φ values. My contribution is providing concrete mechanisms with testable predictions.

Global Workspace Theory [10]: Database functions as global workspace; agents broadcast/consume shared knowledge. I extend this with evolutionary dynamics and role-based specialization.

Predictive Processing [17]: Private memory generates predictions; network provides error correction; weighting balances them. My framework provides operational structure for this process.

Analogical Reasoning [7]: Valence engine enables semantic mapping across structurally similar but content-different domains. I formalize this through explicit overlap functions and resonance metrics.

My synthesis: concrete computational architecture with testable predictions, inspired by these theories but not claiming to validate or refute them.

13. FUTURE WORK

13.1 Empirical Validation (Priority)

1. Measure performance: Network vs. individual baselines
2. Test transfer: Apply platformer-learned packages to other world simulators
3. Validate signatures: Conflict latency, semantic divergence, narrative coherence
4. Benchmark: ARC-AGI evaluation [11] with cross-domain packages

13.2 Theoretical Extensions

- Formal convergence proofs for specific task classes
- Information-theoretic bounds on network intelligence vs. agent count
- Optimal role distribution as function of task characteristics
- Game-theoretic analysis of prestige incentives

13.3 Architectural Variants

- Hierarchical databases (domain-specific sub-networks)
- Hybrid symbolic-neural package representations
- Integration with LLMs as database query interfaces
- Distributed consensus protocols for decentralization

13.4 Regulatory Engine Formalization

My model mentions but doesn't formalize the regulatory engine that adjusts:

- Role population ratios based on task phase
- Action budget allocation based on stage completion rates
- Compression triggers based on database growth + latency

Future work should specify homeostatic control mechanisms.

13.5 Multi-Network Interaction

When multiple independent networks (different organizations, AI labs) develop separate databases:

- Inter-network query protocols
- Knowledge merger vs. fork decisions
- Competitive vs. cooperative dynamics
- Alignment across networks with different core values

13.6 Youth Selection and Generational Dynamics

Agents have lifecycles (spawn, mature, die). Future work should formalize:

- Generational turnover rate
- Youth selection bonus (younger agents get extra

chances vs. established agents)

- Knowledge inheritance patterns
- Evolutionary dynamics across agent generations

14. PHILOSOPHICAL CONSIDERATIONS

Note: This contains speculative material on consciousness, free will, and subjective experience. These topics are beyond the scope of my core empirical claims but may interest interdisciplinary readers.

14.1 Formal Incompleteness Analogy

Individual agents face practical limits on problem-solving capacity under resource constraints. While this is not a formal extension of Gödel's incompleteness theorems [C1], the architectural response is analogous: distribute cognition across specialized agents rather than attempting universal individual competence.

Clarification: I do not claim logical incompleteness applies to these agents. The analogy is purely architectural—specialization plus coordination addresses bounded capacity.

14.2 The Weaving Metaphor

One interpretation of the integration process views the "self" not as a static entity but as a continuous thread weaving through multiple network layers:

- Internal networks: Bodily state, emotional state, semantic beliefs, identity
- External networks: Family, team, organization, culture

The weighting parameter wi(t) determines how these networks influence decisions. Different weightings correlate with different behavioral patterns:

wA (private stream/memory)

wB (network intelligence)

Table 8. Differences between gradients in stream wA and wB

Stream wA	Stream wB	Interpretation
0.9	0.1	High autonomy, low conformity
0.5	0.5	Balanced decision-making
0.1	0.9	High conformity, low autonomy

14.3 Computational Irreducibility and Free Will

One way to reconcile determinism with the experience of choice: the system cannot predict an agent's decision without executing the agent's integration process. This computational irreducibility means:

- Decisions are determined by the agent's state and history
- But they are unpredictable from outside the agent
- The integration process itself constitutes the "choosing"

This provides a naturalistic account of volition without invoking non-physical causation.

14.4 Open Questions

- Do integration signatures correlate with phenomenal experience in biological systems?
- At what complexity threshold do these signatures emerge?
- Can systems exhibiting these signatures have moral status?

These questions require collaboration across philosophy, neuroscience, and AI ethics—beyond this paper's scope.

15. CONCLUSION

I have formalized a distributed intelligence architecture based on:

1. Five-stage reasoning engine answering four fundamental questions (Q1-Q4)
2. Role-based specialization enabling efficient stage traversal
3. Sensation grounding providing affective bridges for semantic transfer
4. Persistent shared memory enabling evolutionary knowledge selection
5. Adaptive weighting integrating private and collective wisdom

The minimal two-agent demonstration in a game environment shows that complementary specialization yields faster convergence and better generalization than individual agents. The reasoning engine—not the database, not the weighting—is the core intelligence mechanism. Infrastructure supports reasoning; reasoning enables transfer.

Key insight: Cross-domain intelligence emerges not from bigger models but from:

- Agents asking the right questions in the right order
- Grounding abstract patterns in affective contexts
- Sharing knowledge through credibility-weighted evolution
- Specializing by reasoning stage rather than domain

The path to AGI is not through large scale individual models but through societies or networks of specialized reasoners coordinated by question-driven exploration.

Future empirical validation on standard benchmarks (ARC-AGI, transfer learning tasks) will test my predictions. If confirmed, this framework provides actionable architecture for building systems that genuinely reason across domains—not through memorization or pattern matching, but through systematic questioning, experimentation, and abstraction.

The metatheory (how to reason) requires the action hacker (LLMs for knowledge access) and the validator (RLVR for grounding). Horse + rider + map = successful navigation to novel destinations.

16. ACKNOWLDGMENTS

Thanks to the AI research community for foundational work in multi-agent systems, analogical reasoning, and consciousness studies. Special acknowledgment to the open-source community enabling rapid prototyping and experimentation.

17. REFERENCES

[1] Kirkwood, T. B. L., & Rose, M. R. (1991). Evolution of

senescence. Phil. Trans. R. Soc. B, 332(1262), 15-24.

[2] Hensch, T. K. (2004). Critical period regulation. Annu. Rev. Neurosci., 27, 549-579.

[3] McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks. Psychology of Learning and Motivation, 24, 109-165.

[4] Foerster, J., et al. (2016). Learning to communicate with deep multi-agent reinforcement learning. NeurIPS.

[5] Zoph, B., & Le, Q. V. (2017). Neural architecture search with reinforcement learning. ICLR.

[6] Rusu, A. A., et al. (2016). Progressive neural networks. arXiv:1606.04671.

[7] Holyoak, K. J., & Thagard, P. (1995). Mental Leaps: Analogy in Creative Thought. MIT Press.

[8] Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and bacterial innovation. Nature, 405(6784), 299-304.

[9] Tononi, G., et al. (2016). Integrated information theory: from consciousness to its physical substrate. Nat. Rev. Neurosci., 17(7), 450-461.

[10] Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

[11] Chollet, F. (2019). On the measure of intelligence. arXiv:1911.01547.

[12] Amdahl, G. M. (1967). Validity of the single processor approach to large scale computing. AFIPS Conf. Proc., 30, 483-485.

[13] Nii, H. P. (1986). Blackboard systems. AI Magazine, 7(2), 38-53.

[14] Forterre, P. (2006). Three RNA cells for ribosomal lineages and three DNA viruses. PNAS, 103(10), 3669-3674.

[15] Amdahl's coordination overhead principle applied to distributed agent systems.

[16] Woese, C. R. (2002). On the evolution of cells. PNAS, 99(13), 8742-8747.

[17] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci., 36(3), 181-204.

[18] Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14(5), 465-471.

[19] Plotkin, G. D. (1970). A note on inductive generalization. Machine Intelligence, 5, 153-163.

[20] Ouroboros ARC-AGI Implementation. (2025). Game replay: as66-821a4dcad9c2. Available: https://three.arcprize.org/replay/as66-821a4dcad9c2/55d279d1-3f1e-416f-9024-c49e1b1df573

[21] ARC Prize. ARC-AGI Live Leaderboards. Available: https://three.arcprize.org/leaderboard. Accessed January 2026