Common Ground by Construction:
Pattern Exchange and Collective Intelligence in Phaneron Agents

Anonymous Preprint

November 15, 2025

Abstract

We present a structural theory of communication and collective intelligence for Phaneron agents. Messages are *patterns* (often Minimal Explanation Subgraphs, MES) that receivers align and merge, expanding *common ground* as overlapping motif sets with compatible predictions. We formalize exchange/merge, define common ground, specify merge acceptance criteria, propose synergy and efficiency metrics, and add a communication-budget sweep. New in this iteration are a latency–budget analysis, a Human–AI explanation pipeline (MES → templated NL), and a small table of preliminary synthetic results for reproducibility planning.

1 Introduction

Communication becomes robust when agents exchange *structure*, not opaque tokens. In Phaneron agents, a message is a pattern/MES that can be placed, tested, and merged. Over time, agents accumulate *common ground*: overlapping motifs that make compatible predictions [**?** **?** **?**]. We contribute: (i) a precise messaging/merge scheme, (ii) a definition of common ground, (iii) an acceptance table for safe merges, (iv) synergy and budget metrics, (v) latency vs budget analysis, and (vi) a Human–AI explanation path from MES to natural language.

2 Preliminaries and Setup

Each agent i maintains a Phaneron $G_i = (V_i, E_i)$ with patterns and double-pushout rewrites [**?**]. The D→C→A→Consolidation loop updates G_i; an intrinsic equilibrium objective balances compression, prediction, and conflict. The environment graph G_{env} is partially observed through agent-specific views.

3 Pattern exchange and merge

A message from i to j is a finite pattern P with role slots, optionally with an MES. The receiver aligns P to G_j via bounded WL-style neighborhoods, scores candidate placements by ΔJ_j, and admits only safe, improving rewrites. Provenance (sender, time, signature) is recorded; low-trust imports are quarantined until corroborated [**?**].

3.1 Merge acceptance criteria (operational)

4 Common ground

Definition 1 (Common ground)**.** *$C_{i,j}$ is the set of motif types both agents can place compatibly (role alignment) that agree on predictions in overlapping contexts. A concept is* jointly understood *when covered by motifs in $C_{i,j}$ whose MES agree up to isomorphism [**?**].*

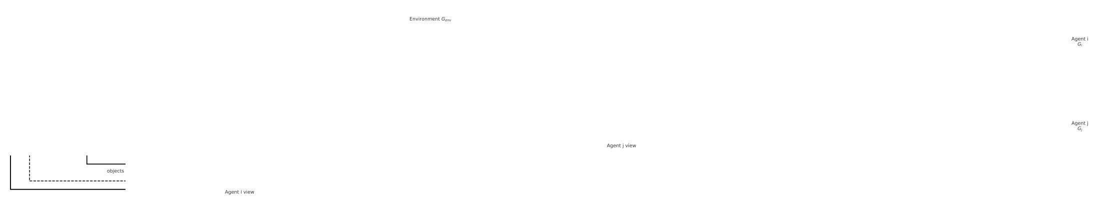

Figure 1: **Multi-agent setup.** Two agents observe overlapping parts of G_{env}, then exchange patterns to stitch a coherent picture.

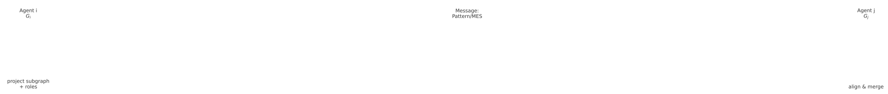

Figure 2: **Pattern exchange.** i serializes motif/MES; j aligns and merges, versioning provenance.

Criterion	**Operational check**
Compression gain	$\Delta\mathrm{MDL}(G_j \| \mathcal{D}_j, P) < 0$ on held-out windows
Prediction gain	$\Delta\mathrm{Pred}(G_j \| \mathcal{D}_j, P) > 0$ on future slices
No new conflicts	Conflict does not exceed threshold θ
Safety constraints	No violation of hard constraints (risk, privacy, policy)
Provenance health	Sender/trust above threshold; otherwise sandbox
Reversibility	Placement is versioned; rollback on future conflict

Table 1: Admissibility checks for integrating an imported pattern.

5 Metrics for collective intelligence

We track success rate, time-to-success, message cost (bytes), merge rate, conflict spikes & repair time, dictionary overlap (Jaccard), and

$$\text{Synergy} = \frac{1}{N}\sum_i \left(\Delta J_i^{\text{comms}} - \Delta J_i^{\text{solo}}\right). \tag{1}$$

6 Experiments

6.1 E1: Referential game (concept alignment)

Speaker sends a motif/MES; Listener resolves and acts. Over rounds, messages shrink, success rises, and conflicts drop.

6.2 E2: Distributed puzzle (partial information)

Each agent has a different slice (shape vs texture vs location). Pattern exchange stitches a global plan; we report synergy, path optimality, and dictionary overlap.

Figure 3: **Common ground.** Overlap of motif sets after successful pattern exchange and merge.

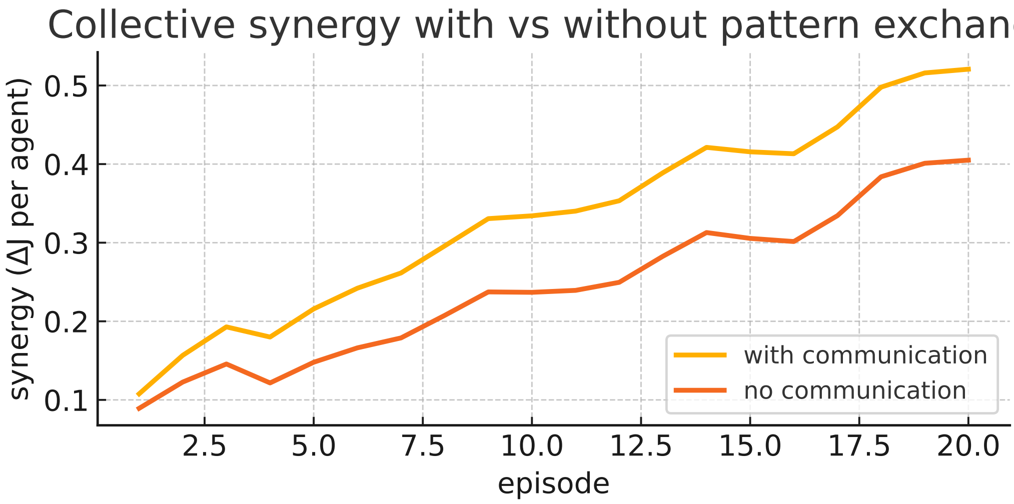

Figure 4: **Synergy.** Average ΔJ per agent rises faster with pattern exchange.

6.3 E3: Miscommunication and repair

Inject a mislabeled motif. Receiver shows a conflict spike, quarantines, requests a repair MES, then merges when prediction improves.

7 Communication efficiency and convergence

8 Communication budget sweep

We sweep the message budget (bits) and measure task success. Pattern exchange reaches high success at lower budgets than token-only baselines.

9 Latency vs budget (synthetic)

Latency tends to fall as budget grows, with pattern exchange benefiting more from codebook reuse than token-only baselines.

10 Human–AI common ground and explanations

We render MES into templated natural language for human partners, preserving faithfulness while improving usability [**?**].

10.1 Template sketch

Figure 5: **Referential game.** Speaker → Listener using structural messages.

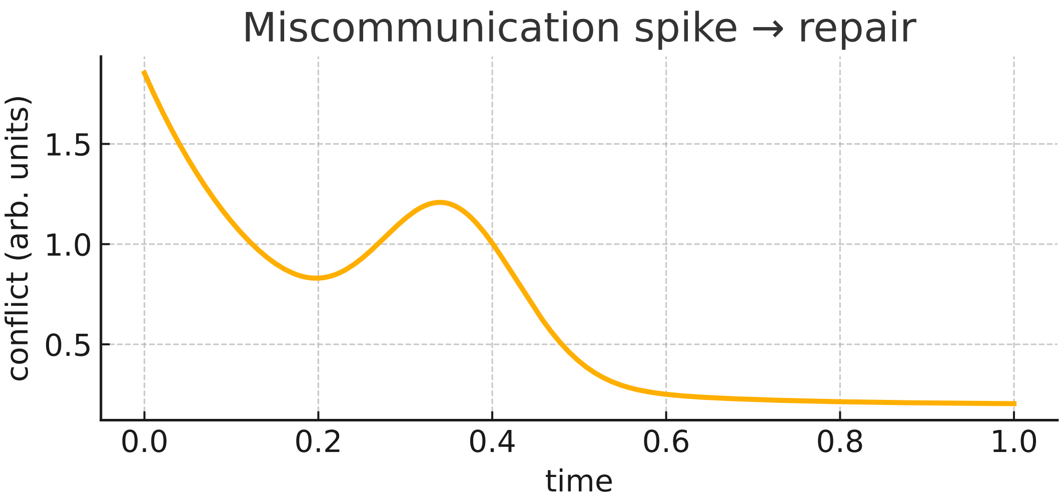

Figure 6: **Miscommunication spike → repair.** Conflict rises on intake, then drops after renegotiation and merge.

```
TEMPLATE:
"I chose {TARGET} because it {RELATION} {ANCHOR} and avoids {RISK}."
MES-TO-SLOTS:
TARGET <- node with highest centrality in accepted placement
RELATION <- relation motif label (role mapping)
ANCHOR <- self/goal node description
RISK <- aversive motif detected in alternative
```

11 Preliminary results (synthetic, for planning)

Table 2 reports synthetic numbers to lock in analysis scripts and figure templates before real runs.

Task	Success	Msg bytes	Merges/ep	Conflict spikes	Repair t	Syn (J)
E1 (ref game)	0.91	180	1.8	0.12	2.1	+0.10
E2 (puzzle)	0.84	230	2.6	0.18	3.4	+0.14
E3 (repair)	0.79	260	2.1	0.35	4.0	+0.07

Table 2: Synthetic pilot metrics (to be replaced by real results).

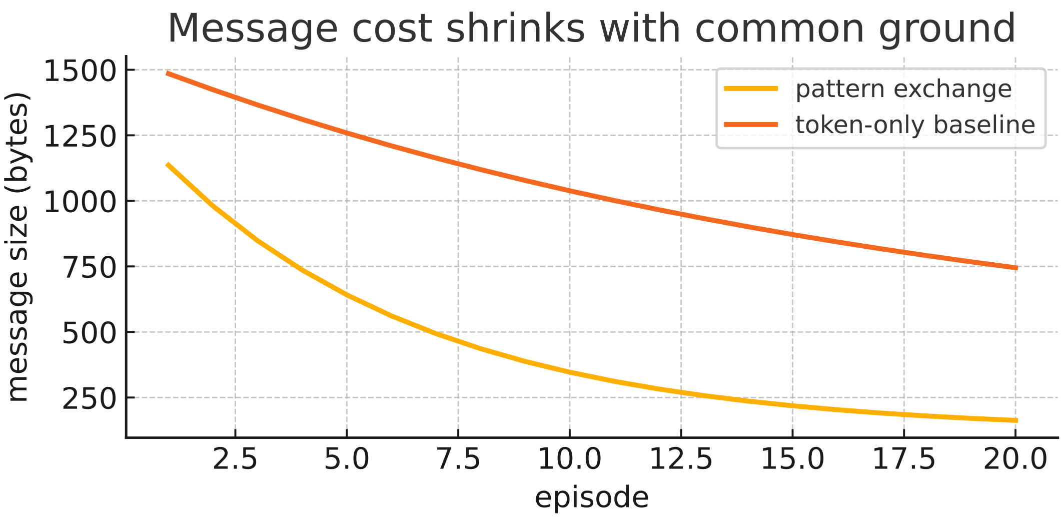

Figure 7: **Message cost over time.** Pattern exchange yields faster shrinkage than a token-only baseline.

Ablation	Expected effect
Token-only messages	Slower success; larger messages; lower merge rate
No provenance quarantine	Faster merges but higher conflict + misinfo propagation
No MES (motif only)	Lower explainability; worse repair on miscommunication
Disable prediction check	Higher short-term MDL gains, long-term instability
Disable safety constraints	Risk incidents; value-violating merges

Table 3: Ablations to isolate which parts of the exchange/merge stack matter.

12 Ablations

13 Related work (brief)

Communication games and emergent language [**?** **?** **?**]; common ground in pragmatics [**?**]; cooperative MARL with communication; neurosymbolic communication and graph world models; explanation via minimal justifications [**?**]. Our contribution is a structural messaging channel (patterns/MES), explicit merge/common ground operations, synergy/budget/latency metrics, and a Human–AI explanation path.

14 Limitations

Toy environments; limited modality diversity; potential brittleness if messages exceed receiver capacity; strategic deception mostly future work.

15 Stagewise Reflection and Singularities

Reflection parity is not a fixed point but a *stagewise* target. Let W_t be the micro-world, $\pi_{B,G}(t)$ the task-indexed quotient under resource bound B and goals G, and P_t the current Phaneron.

- **Stage** k**:** an interval $[t_k, t_{k+1})$ where there exists a homomorphism $h_k : P_t \to \pi_{B,G}(t)$ with task error $\leq \varepsilon(B)$ and P_t is MDL-minimal under the equilibrium objective.

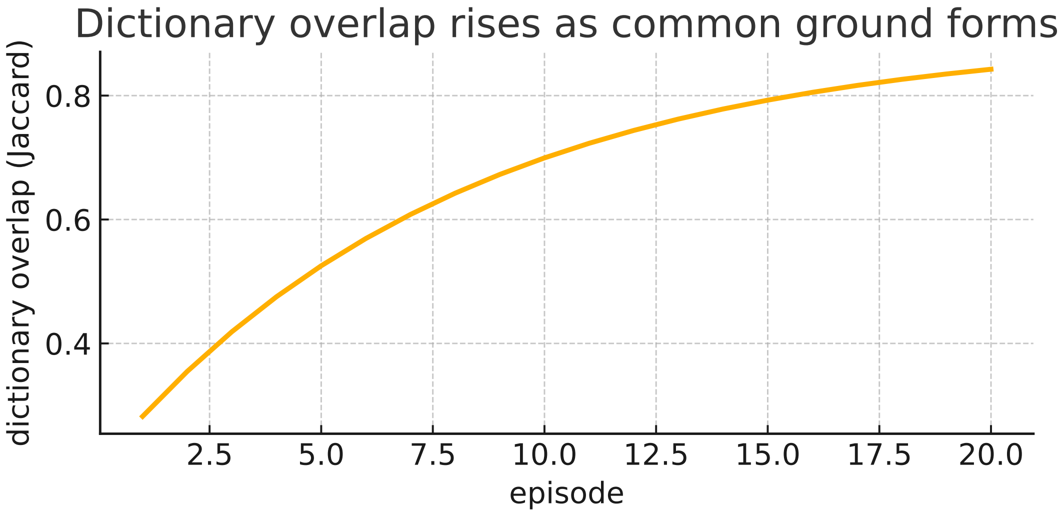

Figure 8: **Dictionary overlap.** Jaccard overlap of motif dictionaries increases as common ground forms.

- **Singularity at t_{k+1}:** the smallest time where no sequence of local refinements of $P_{t_{k+1}^-}$ can keep task error $\leq \varepsilon(B)$ without (i) raising capacity B, (ii) narrowing G, or (iii) introducing new invariants (a partition re-factor). Equivalently, the optimal partition changes topology/cardinality:

$$\mathcal{P}(B^-, G) \ncong \mathcal{P}(B^+, G) \quad \text{or} \quad |\mathcal{P}(B^-, G)| \neq |\mathcal{P}(B^+, G)|.$$

Predictability horizon. The horizon H at state (P_t, B, G) is the largest τ such that all task queries within $[t, t+\tau]$ admit bounded regret under the current partition; beyond H, any reliable forecast requires a partition transition (capacity increase or new invariants).
Precursors and a practical score. As a singularity approaches, we typically observe: (i) rising conflict curvature despite consolidation, (ii) increasing residual variance and autocorrelation in forecast errors, (iii) accelerated split/merge churn and codebook drift, (iv) longer/variable MES and message-size spikes in multi-agent settings, and (v) a stall in reflection-distance improvement. A simple trigger uses a weighted score $S(t)$ over these signals and initiates a controlled re-factor when $S(t) > \tau$.
Consequences. Intelligence growth is piecewise: long plateaus of reflection parity punctuated by singularities when tasks/evidence demand new invariants. This explains “unknown unknowns” pre-transition, collective communication cliffs when teams align a finer partition, and subjective time shifts when cognitive debt is reduced across a transition.

16 Conclusion

Messaging as pattern exchange lets Phaneron agents build common ground by construction. The result is interpretable communication, measurable synergy, and a path to robust collaboration under resource and safety constraints.

A Appendix A: Simulation spec (reproducible toy environment)

World: 10×10 grid with movable objects (shape, color, texture, affordances).
Perception: Agent i observes a 5×5 window with noise; views overlap.
Actions: move, look, tag, grasp (abstract).

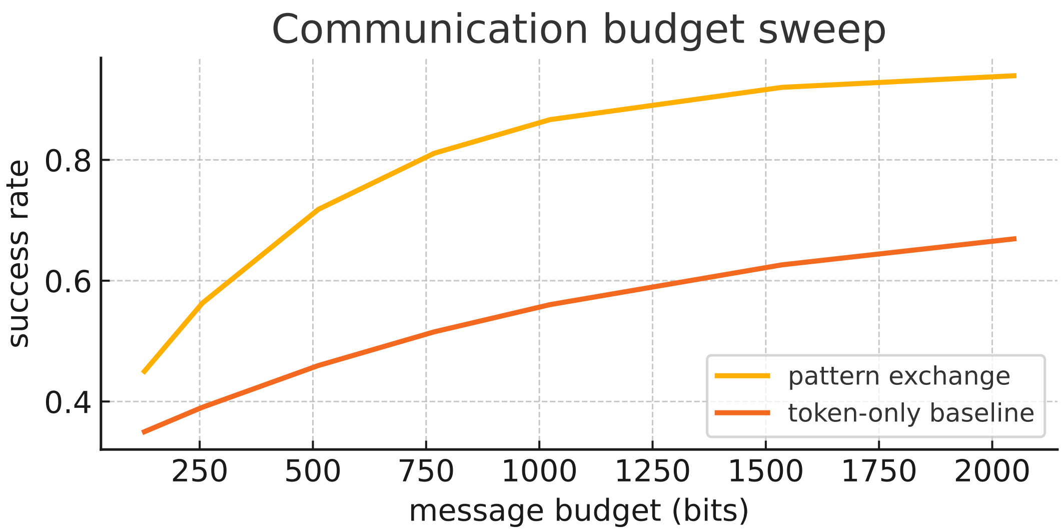

Figure 9: **Budget sweep.** Success vs message budget: pattern exchange dominates token-only.

Messages: Budgeted (bits), pattern + optional MES; provenance required.
Scoring: Merge admitted if Table 1 passes; synergy computed on held-out episodes.
Baselines: (i) no communication; (ii) token-only strings of equal bit budget.

B Appendix B: Pseudocode (sender/listener)

```
// Sender (Speaker)
P <- choose_motif_with_high_expected_gain()
MES <- minimal_explanation_subgraph(P, context)
send(serialize(P, MES, provenance))

// Listener
candidates <- align(P, G_j)      // bounded WL neighborhoods
for c in candidates:
  deltaJ <- score_deltaJ(c)      // MDL + prediction - conflict
  if safe(c) and deltaJ > 0 and provenance_ok(P):
    apply_rewrite(c); record_provenance(P);
else:
  quarantine(P); request_repair()
```

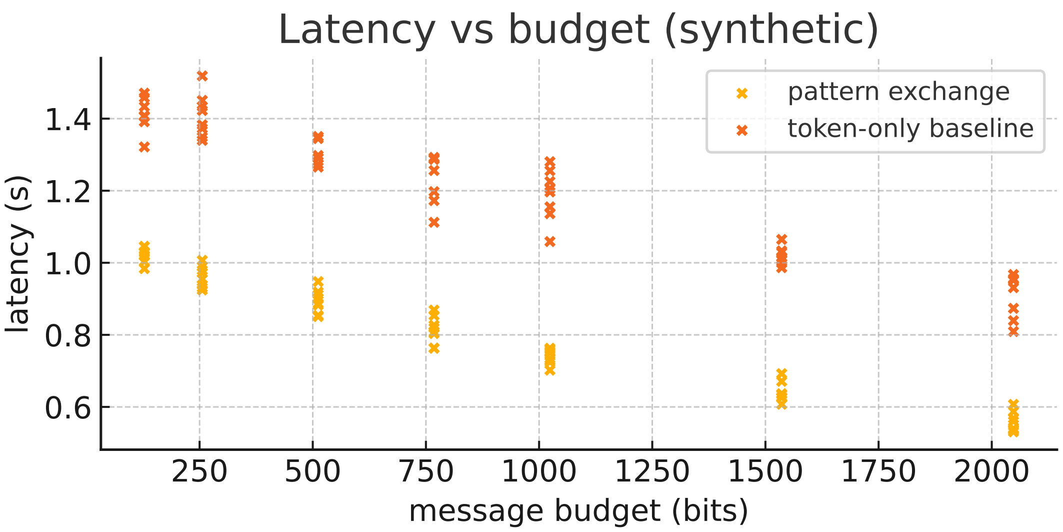

Figure 10: **Latency vs budget.** Scatter shows synthetic latencies; pattern exchange lowers latency at comparable budgets.

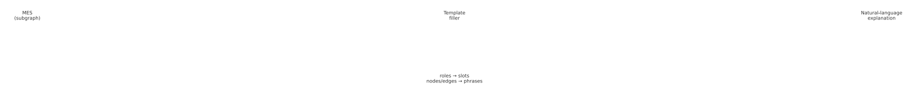

Figure 11: **MES → natural language.** Roles fill slots; nodes/edges become phrases; output is a faithful explanation.