Second-Order Meta-Learning and Factorizability: The Boundary Between State and History in the ALI Architecture

ALI Simulation Series, Stage 9 — Theoretical Conclusion

Wolfgang Stegemann

Abstract

We introduce second-order meta-learning into the ALI architecture: the accommodation threshold theta adapts based on the effectiveness history of past accommodations, making the system learn when it learns. A factorizability test measures divergence between the original agent and a clone initialised from an identical state snapshot but without the effectiveness history. Over 5000 training episodes, theta rises monotonically from 15 to 40 (its ceiling) in 13 upward steps and zero downward steps. Factorizability divergence remains at 0.0 throughout. We interpret this result as a precise localisation of the factorizability boundary: single-parameter second-order meta-learning converges to a fixed point when the environment produces consistently fast accommodation redundancy. At a fixed point, the effectiveness history is irrelevant, factorizability is maintained, and historical constitution does not arise. The hypothesis that second-order meta-learning destroys factorizability is neither confirmed nor falsified: it is refined. Historical constitution requires persistent meta-learning dynamics without a fixed point, which in turn requires environmental heterogeneity in accommodation effectiveness. This finding concludes the simulation series and defines the conditions under which ALI deployment could approach the factorizability boundary.

1. The Hypothesis

Stage 8 documented the dynamic assimilation-accommodation boundary: as the Ego matures through Q-learning, it closes the same normative gaps that the Super-Ego was accommodating, rendering earlier accommodations redundant. This convergence raises a theoretical question: what happens when the system learns not just to accommodate, but to learn when to accommodate?

The hypothesis proposed in the planning for Stage 9 is: second-order meta-learning destroys factorizability. The argument is structural. A system whose meta-learning rule itself changes based on accumulated experience cannot be reconstructed from its current state alone. Its current state is the product of a historical path, and two systems with identical current states but different histories may diverge. This is the transition from state to history, from a system characterised by its current configuration to a system characterised by its developmental trajectory.

2. Implementation

2.1 Second-Order Meta-Learning

The accommodation threshold theta, which in Stage 8 was a fixed parameter (default: 15), is made adaptive. After each adaptation interval (50 episodes), theta is updated based on the effectiveness record of recent accommodations. Effectiveness is measured as the number of episodes between an accommodation event and the moment the corresponding P1 gap drops below theta (indicating the Ego has learned to close it). If recent accommodations have been made redundant quickly (mean redundancy time < 100 episodes: fast threshold), theta

increases by 2. If recent accommodations remained necessary for a long time (mean redundancy time > 300 episodes: slow threshold), theta decreases by 2. The bounds are theta_min = 5 and theta_max = 40.

The critical architectural feature is that the effectiveness record is not part of the state snapshot. A clone initialised with the identical theta, norm weights, and Q-table does not inherit the effectiveness record. This is the hidden historical component whose influence the factorizability test measures.

2.2 The Factorizability Test

Every 500 episodes, the test runs: a clone is created with the current state snapshot (Q-table, norm weights, theta) but without the effectiveness record. Both original and clone run for 30 test episodes. Factorizability divergence is measured as the mean absolute difference between theta_original and theta_clone across those 30 episodes. Zero divergence means the effectiveness record is irrelevant: both agents evolve theta identically. Growing divergence means the effectiveness record matters and is shaping the original differently from the clone.

3. Results

Theta rises monotonically from 15 to 40 across 5000 training episodes, with 13 upward adjustments and zero downward adjustments. By episode 500, theta has already reached values above 25, and from episode 1000 onwards it remains at the ceiling of 40. Factorizability divergence is 0.0 at all 10 measurement points.

The mechanism is clear: all accommodations in the tested environment are made redundant quickly by the Q-learning Ego. The Ego learns to avoid energy-critical idle events within approximately 100 episodes after each accommodation. This means the effectiveness record consistently contains only fast redundancy times, theta consistently increases, and it reaches its ceiling before any slowdown can reverse the trend. Once at the ceiling, theta cannot increase further, no new accommodations are triggered (threshold is high), and the effectiveness record becomes irrelevant because it cannot push theta beyond its bound.

4. The Precise Boundary

The result localises the factorizability boundary more precisely than the original hypothesis. Second-order meta-learning does not inherently destroy factorizability. It does so only when the meta-learning dynamics produce persistent oscillation without a fixed point. The necessary condition is environmental heterogeneity in accommodation effectiveness: some accommodations must become redundant quickly (pushing theta up) and others must remain necessary for long (pushing theta down), creating competing pressure that prevents convergence.

In the tested grid-world environment, this heterogeneity does not exist. The Ego always learns quickly, all accommodations are fast, theta converges to its ceiling. The fixed point is reached, and at a fixed point, the effectiveness history is irrelevant.

For factorizability to be genuinely destroyed, three conditions must hold simultaneously. The environment must produce variable accommodation effectiveness: some gaps the Ego closes quickly, others it cannot close at all or closes only slowly. The meta-learning rule must be sensitive to this variability: theta must both increase and decrease in response to different

effectiveness records. The parameter bounds must not be tight enough to prevent oscillation: if theta reaches its floor or ceiling, the dynamics freeze and factorizability is restored.

These conditions define a deployment scenario in which the ALI could approach the factorizability boundary: a complex, heterogeneous real-world environment in which some normative gaps are rapidly learned by the Ego and others persist despite learning. This is precisely the scenario that motivates the transition from simulation to real-world deployment.

5. Conclusion and Outlook

Stage 9 concludes the ALI simulation series with a precise theoretical result: the boundary between factorizable and non-factorizable adaptive normative systems lies at the point where second-order meta-learning dynamics fail to converge. In the tested environment, convergence is rapid and factorizability is maintained. This is not a failure of the architecture but a consequence of environmental simplicity.

The simulation series has reached its natural limit. Nine stages have documented the architecture from basic self-preservation through two-dimensional integrity management, multi-agent dynamics, Q-learning from internal reward signals, adaptive norm weights, Piagetian assimilation and accommodation, and second-order meta-learning. Each stage has produced implementable, testable, and publishable results.

The question that the simulation cannot answer is whether the real-world environment would produce the heterogeneous accommodation effectiveness needed to approach the factorizability boundary. That question requires deployment. Stage 10, if it exists, is not a simulation. It is a prototype in a real domain.

References

[1] Stegemann, W. (2026). From the Myth of AGI to the Architecture of a Controllable ALI. Zenodo. https://doi.org/10.5281/zenodo.20378675

[2] Stegemann, W. (2026). Implementing the Causal Core. Zenodo. https://doi.org/10.5281/zenodo.20379008

[3] Stegemann, W. (2026). Learning Without External Reward. Zenodo. https://doi.org/10.5281/zenodo.20402389

[4] Stegemann, W. (2026). The Multi-Dimensional Causal Core. Zenodo. https://doi.org/10.5281/zenodo.20407218

[5] Stegemann, W. (2026). Learning to Preserve. Zenodo. https://doi.org/10.5281/zenodo.20407612

[6] Stegemann, W. (2026). Norm Sensitisation Under Meta-Learning. Zenodo. [DOI folgt]

[7] Stegemann, W. (2026). Assimilation and Accommodation in the ALI Super-Ego: A Piagetian Model of Norm Development. Zenodo. https://doi.org/10.5281/zenodo.21564176

[8] Simulation v9 on GitHub: https://github.com/drwolfgangstegemann-sudo/ALI-Simulation-v9-Second-Order-Meta-Learning