

Persistent Parts: An IFS-Inspired Architecture for Lifelong Agentic Cognition

Taylor Burke

Abstract. Current LLM agents are commonly organized around a unitary objective, supplemented by memory, critics, planning, or temporary subagents. This position paper proposes a different organization: one persistent agentic identity containing multiple persistent, specialized cognitive *Parts* that can influence, but cannot independently actuate, the agent. The design draws inspiration from Jungian complexes, Internal Family Systems (IFS), Minsky’s *Society of Mind*, Global Workspace Theory, and predictive-processing accounts. Emotion is used only as an inferred salience signal: unusually significant events may create or reactivate narrow, intentionally biased Parts. Activated Parts bid for influence, make predictions, and revise local beliefs from later outcomes, while external action remains centralized in a single executive identity. We hypothesize that this architecture can preserve important minority concerns, support long-horizon relational learning, increase solution diversity, and make internal motivational conflict more inspectable than either unitary agents or disposable subagent swarms.

1. Motivation

Long-running agents face a problem that short-lived task agents largely avoid: experience must shape future behavior without allowing every important event to globally rewrite the system. Conventional memory retrieval can preserve facts, but facts and motivation are not identical. A stored statement such as “ambiguous instructions caused failure” does not necessarily create a process that persistently notices ambiguity, overweights its risk, and argues for caution months later.

Human cognition appears to make extensive use of this kind of specialization. Significant experiences can produce enduring expectations and behavioral tendencies that activate selectively when similar situations recur. These tendencies are useful partly because they are not perfectly balanced. A cautious perspective notices dangers that an impartial reasoner may discount; an optimistic perspective may search solution spaces that a cautious reasoner abandons.

We propose deliberately reproducing this property in an artificial cognitive architecture while retaining a single external agentic identity.

2. Cognitive Lineage

The proposal is inspired by several theories that describe cognition as differentiated rather than homogeneous.

Jungian psychology describes *complexes*: semi-autonomous structures organized around emotionally significant memories and expectations that can become disproportionately influential in relevant situations [1]. Internal Family Systems makes a related idea operational by describing persistent *Parts* with specialized protective beliefs and motives, coordinated by a broader Self rather than eliminated by it [2].

Minsky’s *Society of Mind* contributes the principle that intelligence can arise from many specialized processes rather than a single general-purpose mechanism [3]. Global Workspace Theory contributes competition for access to globally consequential cognition: many local processes may operate in parallel while only some gain executive influence [4]. Predictive-processing accounts add the idea that useful internal models should make expectations, encounter prediction error, and revise themselves [5].

These theories are not treated as equivalent, nor as literal computational descriptions of the brain. Together, however, they motivate a design space based on persistent specialization, selective activation, competition for influence, prediction-driven updating, and centralized actuation.

3. Architecture

The system contains one persistent **identity** with exclusive authority over external action. Tool use, messages, file changes, robotic control, and other effects remain behind this boundary.

Within that identity are persistent **Parts**. A Part is not a complete autonomous agent. It is a narrow specialist organized around a concern and an intentionally skewed belief about the world. Its purpose

is to preserve a perspective the broader agent might otherwise smooth away.

Part creation begins with appraisal of event salience. The system may infer that an event involves danger, betrayal, responsibility, novelty, attachment, unusual success, or another state that would ordinarily carry strong emotional significance. No claim about machine phenomenology is required. Emotion is used here as a computational primitive: a compressed estimate that an event may deserve disproportionate influence over future cognition.

A highly salient event can create a new Part or reactivate an existing one. For example, repeated failures involving ambiguous instructions might produce a Part that disproportionately expects ambiguity to conceal risk. Repeated success from unconventional strategies might produce another that overweights unlikely approaches.

When triggered, Parts bid for influence. They can recommend actions, surface concerns, and predict likely outcomes if their advice is followed or ignored. Multiple Parts may disagree. The persistent executive identity considers these perspectives together with current goals, memory, and world state, then chooses what to do.

Parts never act independently. This is the primary distinction from a swarm. The architecture contains one agent with differentiated cognition, not a society of externally acting agents.

After an outcome becomes observable, relevant Parts can compare what happened against their earlier predictions and revise their local beliefs. Accurate predictions may strengthen a Part’s confidence or trigger conditions; repeated errors may weaken or generalize them. Parts that no longer provide useful information can become inactive or archived. This enables strong local adaptation without globally rewriting the entire agent around every experience.

4. Why Persistent Parts?

Temporary role-based subagents can provide diverse viewpoints, but their specialization is usually assigned at inference time and disappears after the task. A persistent Part instead accumulates a history of why it cares. Its prior experiences alter what it notices, what it predicts, and what regions of the solution space it searches.

This may provide two advantages. First, it can support **relational and longitudinal learning**: concerns can remain behaviorally meaningful across long intervals without becoming universal rules. Second, deliberately biased Parts may support **creative search**. A balanced model tends toward reasonable high-probability solutions; specialists with different priors can search unusual alternatives while the executive remains responsible for deciding whether those alternatives are appropriate.

The design also separates **cognitive freedom from agentic authority**. Internal specialists can be cautious, optimistic, suspicious, imaginative, or wrong without each receiving an independent path to external action. This creates an inspectable boundary between having a

thought and becoming the actor.

5. Research Questions and Evaluation

The proposal makes several falsifiable predictions. Compared with an otherwise equivalent agent using conventional memory or disposable subagents, a Parts-based agent should:

- retain important minority concerns across longer time horizons;
- develop more stable relational expectations from repeated interaction;
- produce more diverse candidate solutions on ambiguous or difficult tasks;
- preserve explicit conflict among competing objectives rather than allowing one objective to silently dominate cognition; and
- expose failure modes such as pathological fixation or excessive reinforcement in an inspectable form.

A minimal evaluation could compare three systems in the same long-horizon simulated environment: (1) a conventional agent with memory, (2) an agent that instantiates temporary role-based subagents, and (3) an agent with persistent Parts. The environment should contain repeated relationships, conflicting objectives, rare failures, and opportunities where unconventional strategies sometimes succeed.

Useful measurements include prediction accuracy over time, retention and appropriate generalization of prior concerns, diversity and quality of proposed solutions, performance under conflicting goals, calibration of Part confidence, and frequency of harmful fixation. An especially important negative result would be evidence that persistent Parts merely entrench early errors; the architecture should be judged

partly by whether local prediction-driven updating can soften such distortions.

6. Position

Agentic AI may benefit from abandoning the assumption that good cognition should converge internally before action selection begins. Persistent disagreement can itself be computationally useful.

Rather than forcing honesty, caution, curiosity, task completion, uncertainty, and other concerns into one temporarily balanced voice, an agent can preserve specialized advocates for those concerns while maintaining a single external identity. The executive does not have to agree with every Part; it gains the benefit of having had the thought.

The resulting architecture is neither a swarm nor a hive mind. It is one persistent actor whose cognition is allowed to remain internally differentiated: **one identity, one body, many Parts**.

References

- [1] C. G. Jung. *The Structure and Dynamics of the Psyche*. Collected Works, Vol. 8. Princeton University Press, 1960.
- [2] R. C. Schwartz. *Internal Family Systems Therapy*. Guilford Press, 1995.
- [3] M. Minsky. *The Society of Mind*. Simon & Schuster, 1986.
- [4] B. J. Baars. *A Cognitive Theory of Consciousness*. Cambridge University Press, 1988.
- [5] A. Clark. “Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.” *Behavioral and Brain Sciences*, 36(3), 2013.