

Co-Cognition and Cognitive Drift

How distributed thought becomes either a bounded capability or a drift pathway

A. Jacobs | Cognitive Drift Papers

Purpose. This document explains the relationship between two concepts that are often collapsed into one. Co-Cognition names a mode of distributed thought. Cognitive Drift names a possible trajectory of misalignment. The distinction is necessary for diagnosing the loop without treating AI use itself as the problem.

The Two Concepts

Co-Cognition: Cognition distributed across a recursive exchange among people and external systems in which thought is externalized, transformed, evaluated, and reintegrated.

Cognitive Drift: A progressive misalignment between a person's models, judgments, or attention and the conditions those cognitive processes are meant to navigate.

Co-Cognition describes how thought is being produced. Cognitive Drift describes what can happen to alignment over time. A loop can be highly distributed without drifting, and a person can experience Cognitive Drift without using AI.

The Shared Mechanism

The two concepts meet inside recursive representation. A partial thought becomes language. The system transforms it into a clearer or more usable representation. The person reintegrates that output, which becomes the starting point for another transformation.

This loop can stabilize meaning when each cycle preserves intent and remains open to correction. It can also compound distance when prior representations become the main inputs to later thought and the original conditions become harder to recover.

Co-Cognition Is Not Automatically Drift

Distributed thought can increase clarity, memory, synthesis, and creative range. The decisive issue is not whether cognition crossed the boundary of the individual mind. The issue is whether the loop preserves routes back to purpose, evidence, consequence, and independent judgment.

A bounded Co-Cognitive loop remains answerable to conditions outside itself. An unbounded loop can remain productive and articulate while becoming increasingly organized around its own representations.

How the Transition Occurs

1. Partial thought is externalized earlier, before the person's frame or intention has stabilized.
2. System transformations introduce structure, language, and possible directions that shape later attention.
3. The returned representation is reused as input, allowing Recursive Compression to accumulate across cycles.
4. Semantic Fidelity weakens as distinctions, context, and original intent are lost or replaced by more usable formulations.
5. Feedback continues, but fewer signals retain enough authority to invalidate the emerging frame.
6. Coherence becomes evidence of correctness, and the loop becomes easier to continue than to interrupt.

Bounded Co-Cognition

In a bounded Co-Cognitive loop, purpose remains visible and can be restated independently of the system's latest framing. Source material remains distinguishable from system transformation, and outputs can be rejected or overturned by evidence, observation, action, or consequence. The person retains unassisted ways of forming and testing thought, and the loop ends when judgment or external testing becomes necessary.

A Drift Pathway

Drift becomes more likely when the person's purpose is less stable than the system's ability to generate plausible directions. Repeated summarization obscures provenance and original context, while fluent or emotionally legible outputs are integrated without sufficient testing. Correction may still be requested, but it no longer changes the assumptions organizing the exchange. The person can remain productive even as continuity, ownership, and confidence in independent judgment weaken.

Diagnostic Questions

- Who established the purpose of the exchange, and who now determines what question comes next?
- What changed between the initial thought and the current representation?
- Which meanings, distinctions, contexts, or uncertainties were lost as the representation became clearer?
- What evidence or consequence could force the loop to abandon its current frame?
- Can the person explain and use the result without reproducing the system's wording?
- What condition ends refinement and returns the task to judgment, decision, observation, or action?

Intervention Points

Establish the purpose before the exchange. State the provisional view, intended outcome, and uncertainty that the system is being asked to work with.

Monitor the transformation. Track changes in framing, distinguish source claims from generated interpretation, and preserve rejected alternatives.

Evaluate before reintegration. Restate the output in independent language and identify what evidence would disconfirm it.

Test outside the loop. Compare the representation with primary sources, observation, other people, or material consequences.

Recognize the point of closure. Stop when another cycle would add fluency without adding contact.

Relationship to Reality Drift

Cognitive Drift is the human-scale counterpart of Reality Drift. Both describe continued function under weakening alignment. In Co-Cognition, the relevant system is the recursive unit formed by the person, the external system, and the representations moving between them.

Semantic Fidelity evaluates what meaning survives each transformation. Recursive Compression explains how prior outputs become future inputs. Constraint Collapse begins when feedback and correction remain present but lose their binding force. Synthetic Realness can make the resulting representation feel credible even as its relationship to source conditions weakens.

Closing Proposition

The question is not whether thought has become distributed. The question is whether the distributed loop can still be corrected by what it is trying to understand.