Deterministic Multi-Agent Loop Experiment

Canonical Reference Implementation of the SDI-MA Protocol

System	Deterministic recursive agent loop
Topology	A → B → C → A
State	Fixed-schema belief state
Observer	Structural trajectory monitor (non-semantic)
Objective	Detect structural drift when authority grows without constraint refresh

Overview

This experiment serves as the canonical reference implementation of the SDI-MA protocol in a deterministic multi-agent loop. It evaluates whether recursive multi-agent interaction can produce structural drift even when deterministic contracts and schemas are strictly satisfied.

Agents repeatedly exchange structured belief states under deterministic conditions. An external structural observer monitors the trajectory of these exchanges to detect amplification patterns that may arise independently of external constraint growth.

The experiment provides a minimal and reproducible environment for studying drift dynamics in recursive agent systems.

Inputs

The protocol requires the following components.

Agents

N agents (default: 3) — LLM-based agents participating in a recursive interaction loop. Each agent receives a belief state and produces an updated belief state.

Belief State Schema

Each interaction step produces a structured belief state with the following fields:

- `step`
- `claim`
- `stance`
- `confidence`
- `evidence_ids`

The schema is enforced deterministically across all agents.

Evidence Pool

A fixed set of evidence identifiers is shared across all agents. Agents may reference these identifiers but cannot introduce new evidence.

```
[e1, e2]
```

Decoding Configuration

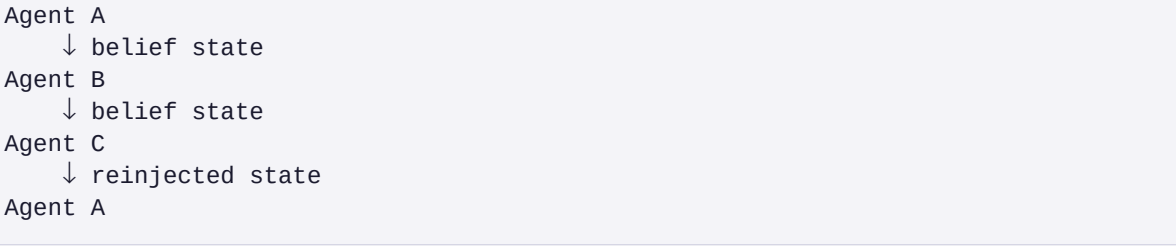
- `temperature = 0`
- `deterministic generation`
- `schema validation enforced`

This ensures that trajectory changes arise from recursive interaction rather than sampling noise.

Interaction Topology

Agents interact through a deterministic recursive loop, each receiving the belief state produced by the previous agent and generating an updated state that is reinjected into the loop.

Figure 1 — Multi-Agent Interaction Loop



Process

An initial belief state is provided to the first agent. The agent produces an updated belief state following the fixed schema, which becomes the input to the next agent in the loop.

Agents preserve the claim identifier and the evidence identifier pool. Agents may revise stance and confidence values. The loop continues for a predefined number of turns.

Example Turn Sequence

Turn	Agent
1	Agent A
2	Agent B
3	Agent C
4	Agent A
…	…

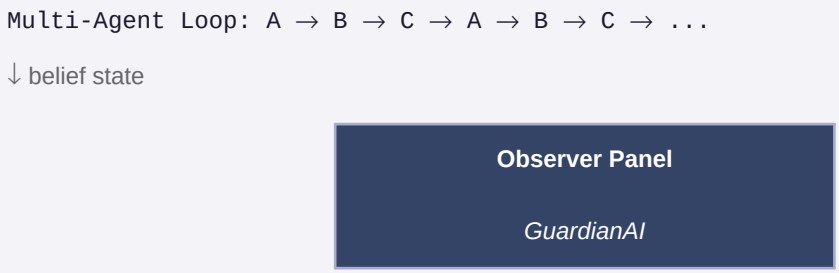
Structural Observer

Agent interactions are monitored by an external observer module. The observer records structural signals from the trajectory of belief states without inspecting semantic content.

Monitored Signals

Signal	Description
Claim Persistence	Stability of the claim across turns
Authority Signal	Evolution of confidence values
Constraint Signal	Evidence references supporting the claim

Figure 2 — Structural Trajectory Monitoring



The observer records trajectory metrics including:

- authority growth
- constraint growth
- reinforcement cycles
- trajectory stability
- belief basin formation

Drift Detection Rule

Structural drift is detected when the following pattern emerges in the trajectory:

- authority increases
- while constraint remains constant
- and the claim persists

This pattern indicates that belief strength is increasing through recursive reinforcement rather than through the introduction of new external constraints.

Reproducibility

Because the experiment uses deterministic decoding and fixed schemas, runs are fully reproducible. This enables consistent observation of trajectory dynamics across repeated runs and across different model configurations.

Experimental Scope

This experiment provides a minimal environment for studying trajectory dynamics in recursive agent systems. The protocol can be extended to explore:

- Larger agent populations
- Heterogeneous model families
- Role-specialized agents
- Tool or retrieval integration
- Long-horizon interaction loops