

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Dynamic Theory of Mind as a Temporal Memory Problem: Evidence from Large Language Models

Permalink

<https://escholarship.org/uc/item/2bt8d7kv>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Nguyen, Thuy Ngoc

Phan, Duy Nhat

Gonzalez, Cleotilde

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Dynamic Theory of Mind as a Temporal Memory Problem: Evidence from Large Language Models

Thuy Ngoc Nguyen

Department of Computer Science, University of Dayton

Duy Nhat Phan

University of Dayton Research Institute, University of Dayton

Cleotilde Gonzalez

Department of Social and Decision Sciences, Carnegie Mellon University

Abstract

Theory of Mind (ToM) is central to social cognition and human–AI interaction, and Large Language Models (LLMs) have been used to help understand and represent ToM. However, most evaluations treat ToM as a static judgment at a single moment, primarily relying on tests of false beliefs. This overlooks a key dynamic dimension of ToM: the ability to represent, update, and retrieve others’ beliefs over time. We investigate dynamic ToM as a temporally extended representational memory problem. We introduce DToM-Track, an evaluation framework testing temporal belief reasoning in controlled multi-turn conversations through recall of prior beliefs, inference of current beliefs, and detection of belief change. Using LLMs as computational probes, we find a consistent asymmetry where models reliably infer current beliefs but struggle to retrieve prior belief states once updates occur. This pattern persists across model families and scales, consistent with recency bias and interference effects documented in cognitive science.