LEARNING DYNAMIC BELIEF GRAPHS FOR THEORY-OF-MIND REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Theory of Mind (ToM) reasoning in highly uncertain environments often involves complex, dynamic, and interdependent belief evolution that jointly shape the process of information seeking and sequential decision-making – especially in high-stakes settings such as disaster response, emergency medicine, and human-in-the-loop autonomy. However, prior approaches for ToM reasoning in LLMs often treat beliefs as static and independent, producing incoherent mental models over time and weak reasoning in dynamic contexts. We introduce a structured cognitive trajectory model for LLM-based ToM that represents mental state as a dynamic belief graph, jointly inferring latent beliefs, learning their time-varying dependencies, and linking belief evolution to information seeking and decisions. Our model contributes (i) a novel projection from textualized probabilistic statements to consistent probabilistic graphical model updates, (ii) an energy-based factor graph representation of belief interdependencies, and (iii) an ELBO-based objective that captures belief accumulation and delayed decisions. Across multiple real-world disaster evacuation datasets, our model significantly improves action prediction and recovers interpretable belief trajectories consistent with human reasoning, providing a principled module to augment LLMs using ToM reasoning with dynamic and interdependent beliefs in highly uncertain environment.

1 INTRODUCTION

Theory of Mind (ToM) refers to the ability to reason about people's latent mental states, such as beliefs, intentions, and expectations, and to use these inferences to explain and predict behavior Obiols & Berrios (2009); Wimmer & Perner (1983). While ToM has been central in psychology and cognitive science, it is increasingly important for AI systems built on Large Language Models (LLMs) that seek to model, predict, or interact with human behavior Baker et al. (2017); Zhang et al. (2025b); Jin et al. (2024). A key challenge in existing Theory-of-Mind reasoning is that the beliefs underlying human behavior are usually dynamic, interdependent, and unobserved. In real-world decision-making, people rarely act from a single isolated belief. Instead, they maintain multiple beliefs simultaneously, update them as new observations arrive, and allow different beliefs to reinforce or weaken one another Christian et al. (2024); Jara-Ettinger (2019); Lindell & Perry (2012).

These properties are especially important in high-stakes decision-making, such as emergency response, financial crises, or medical triage Lindell & Perry (2012); Croskerry (2009); Alfiana et al. (2025), where decisions often unfold through sequential belief updating rather than a one-shot judgment. For example, in disaster evacuation, cognitive theories such as the Protective Action Decision Model (PADM) Lindell & Perry (2012) emphasize that protective behavior depends on multiple beliefs, including perceived threats and risks to oneself, property, and others, which jointly shape people's sequential information seeking and evacuation decisions. Similarly, in emergency medicine, Recognition-Primed Decision theories describe how clinicians continuously revise beliefs about symptom severity, disease likelihood, treatment risk, and time pressure as new tests, observations, or patient responses become available before making treatment decisions Croskerry (2009). More broadly, Bayesian and predictive-processing accounts characterize cognition as a sequential process of latent belief updating under uncertainty Jara-Ettinger (2019); Baker et al. (2017). Across these settings, interdependent belief updates are often nonlinear, context-dependent, and rapidly evolving over time Champ & Brenkert-Smith (2016); Paton (2019); Cova & Drews (2024).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

In machine learning, Theory of Mind is typically formalized through Bayesian inverse planning (BIP), which infers mental states via inversion of a generative belief–action process Baker et al. (2017). BIP is principled but typically depends on synthetic state spaces (e.g., grid worlds) and specified dynamics, limiting their scalability to real-world settings. Recent LLM-based ToM methods leverage LLMs to infer beliefs or intentions from textual descriptions Zhang et al. (2025b); Jin et al. (2024); Shi et al. (2025) through inference-time prompting and hypothesis sampling to estimate belief distribution. However, most existing LLM approaches typically (1) rely on explicit belief signals: they either directly prompt LLMs to output belief states Wilf et al. (2024), or supervise or guide belief inference in LLMs using ground-truth belief labels, which are highly challenging to collect in real-world, high-stakes settings Chen et al. (2025); Zhang et al. (2025b); (2) reduce belief representations to a small set of independent or mutually exclusive hypotheses Gandhi et al. (2023); Le et al. (2019); Kim et al. (2023), which ignores that real mental states involve multiple interdependent beliefs whose interactions can reinforce or suppress one another, so that behavior depends on their joint configuration rather than any single belief; (3) are formulated in static or single-step settings and lack temporally aligned observation–action trajectories, not capturing how beliefs evolve over time through accumulation and continuous updates as new information arrives.

In this paper, we introduce a structured cognitive trajectory framework for LLM-based ToM reasoning. We represent mental states as dynamic belief graphs, capturing how multiple latent beliefs interact and evolve over time. The framework integrates LLM-derived semantic representations with a structured probabilistic model, enabling belief updates conditioned on observations and past states. Actions are predicted based on the evolving belief state, allowing the model to capture how belief dynamics drive decision-making. The entire system is trained end-to-end using a variational objective that ensures inferred beliefs are behaviorally grounded. At test time, the model performs online belief updating and action prediction from observations alone. An overview of the proposed framework is shown in Fig. 1. Our contributions include:

- We introduce a structured cognitive trajectory framework that models mental states as dynamic belief graphs rather than static and independent hypotheses. The framework captures key properties of real-world decisions: multiple co-existing beliefs and sequential belief updating over time.
- We develop a neuro-symbolic latent variable model that combines LLM-derived semantic evidences with structured probabilistic reasoning. The model jointly learns temporal belief evolution and inter-beleif dependencies through temporally conditioned unary and pairwise belief interactions.
- We derive a variational learning approach to jointly learn latent belief dynamics directly from observed action trajectories *without any belief-level supervision*. Unlike prior ToM approaches that infer beliefs through prompting or explicit labels, our framework behaviorally grounds latent beliefs through an ELBO objective linking observations, beliefs, and actions.
- We introduce a stronger evaluation setting for ToM reasoning beyond standard final-answer prediction [1]. In addition to action prediction, we evaluate whether inferred latent beliefs align with independently collected human self-reported cognitive states, recover human belief-interaction structures, and exhibit realistic temporal belief dynamics.

2 Related Works

Machine and LLM-Based Theory of Mind: Machine Theory of Mind (ToM) frames mental state inference as a probabilistic inverse problem, where latent beliefs are inferred from observed actions. A canonical formulation is Bayesian Inverse Planning, which frames ToM reasoning as inverting a generative decision-making model, instantiated as a POMDP Baker et al. (2017). Subsequent work scales this paradigm using neural and amortized inference, while largely retaining manually specified belief structures Jha et al. (2024). While classical Machine ToM provide a general framework for inferring latent mental states, they are typically evaluated in synthetic environments such as grid worlds or small-scale simulated tasks. Recent work has explored the use of Large Language Models (LLMs) to support mental state inference from textual descriptions and social narratives Zhang et al. (2025b); Jin et al. (2024); Shi et al. (2025); Yang et al. (2025); Kim et al. (2025a).

MuMToM Jin et al. (2024) integrates LLMs with a predefined POMDP-based agent model for Theory-of-Mind inference, using LLMs to interpret language observations and estimate action likelihoods for inverse planning. Building on this paradigm, AutoToM Zhang et al. (2025b) further removes

[1] https://anonymous.4open.science/r/ICLR_submission-2E2E/README.md

the need for a fixed agent model by leveraging LLMs to automatically generate belief hypotheses and estimate their posterior probabilities, effectively performing Bayesian-style ToM inference with model structure discovery directly in the language space. Nevertheless, these methods perform belief inference at inference time using frozen LLM priors, rather than learning belief representations from data. As a result, beliefs are treated as static and independent hypotheses without statistical grounding or temporal coherence, making them susceptible to hallucinations under unconstrained prompting.

Deep Temporal Generative Models. Deep generative models, particularly Variational Autoencoders (VAEs), provide a principled framework for latent variable inference via variational optimization of the ELBO. To model sequential data, researchers have extended the Variational Autoencoder (VAE) to time-series settings, broadly referred to as Deep Markov Models (DMMs) Krishnan et al. (2016) or Sequential VAEs Chung et al. (2015). These approaches model the temporal dynamics of a latent process $z_{1:T}$ via a transition distribution $p_\theta(z_t \mid z_{t-1})$, enforcing temporal coherence across latent states and allowing information to be integrated over time rather than treated independently.

Energy-Based Models and Factor Graphs. Energy-Based Models (EBMs) and factor graphs provide a natural way to model structured dependencies by assigning energies to joint variable configurations. Recent methods such as Structured Prediction Energy Networks (SPENs) Belanger & McCallum (2016) and Factor Graph Neural Networks Zhang et al. (2023) use neural networks to parameterize these energies and capture complex interactions. Our framework combines this structured modeling with temporal latent dynamics: a Deep Markov Model captures belief evolution over time, while a factor graph at each step models dependencies among beliefs.

3 Structured Cognitive Trajectory Modeling

3.1 Problem Formulation

Consistent with Theory-of-Mind formulations, we model human decision-making as a sequential decision-making process. Unlike traditional LLM-based ToM settings that represent an agent's mental state with a single belief or a small set of mutually exclusive beliefs, our setting allows reasoning over multiple co-existing beliefs. Specifically, an agent interacts with an environment over discrete time steps $t = 1, \ldots, T$: at each step, the environment has underlying physical states s_t. The agent receives partial observation o_t, and then updates their internal latent beliefs $\mathbf{b}_t$, and selects an action a_t, which further influences subsequent environmental states.

We represent the cognitive state as a structured belief graph $\mathcal{G}_t = (\mathbf{b}_t, \mathcal{E}_t)$, where $\mathbf{b}_t = [b_{t,1}, \ldots, b_{t,K}] \in \{0,1\}^K$ denotes the set of K latent belief variables and $\mathcal{E}_t$ represents edges, which capture how beliefs reinforce or suppress one another over time. We define an individual's belief about a proposition as their degree of confidence that the proposition is true, represented by a subjective probability between 0 and 1. For example, during a disaster evacuation, relevant propositions may include "my home will be damaged during this event" or "I will be injured during the evacuation". Each latent variable $b_{t,i}$ corresponds to one such proposition, and its marginal probability $P(b_{t,i} = 1)$ represents the strength of that belief. In practice, a candidate set of belief constructs can be built from well-established behavioral theories. Many behavioral theories are distilled from decades of empirical psycho-behavioral studies of real human behavior, retaining a set of verified cognitive factors that have been repeatedly observed to explain behavior across populations and events. For example, beliefs about threat, protective actions, and stakeholder perceptions in protective action decision theory Lindell & Perry (2012) for disaster evacuation; beliefs about symptom severity, treatment risk, and time pressure in recognition-primed decision theories for emergency medicine Croskerry (2009); behavioral, normative, and control beliefs in the Theory of Planned Behavior Ajzen (1991); and perceived usefulness and ease of use in the Technology Acceptance Model Davis (1989). Compared with ad-hoc, instance-specific belief hypotheses generated by LLMs in past LLM-augmented Bayesian ToM systems Zhang et al. (2025b); Kim et al. (2025b), such theory-informed belief constructs constrain the hypothesis space using empirical evidence and thus support more interpretable inference pathways that are better aligned with empirically documented patterns of human cognition, which we have validated in Appendix I, Table 10.

The resulting cognitive trajectory is represented as $(s_t, o_t, \mathcal{G}_t, a_t)_{t=1}^T$, where the environment states s_t and belief graphs $\mathcal{G}_t$ are latent, while observations o_t and actions a_t are observed. Examples of observations, actions, and belief variables used in our experiments are provided in Appendix K.

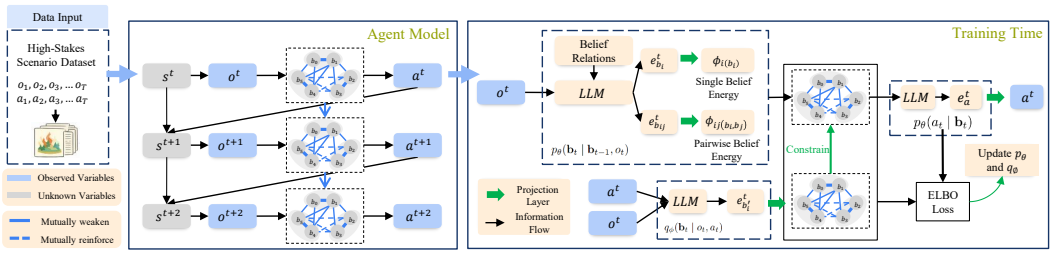


Figure 1: Overview of the Structured Cognitive Trajectory ToM framework. Here, s_t denotes the latent environment state, o_t the agent's observation, a_t the observed action, b_t the latent belief state, and e_t the semantic embedding extracted by a LLM.

Generative Process: We model belief evolution and action generation with a structured latent-variable model. Conditioned on the previous belief state and the current observation, the belief transition prior $p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$ captures belief persistence, accumulation, and inter-belief interactions. Given the latent belief state, the action model $p_\theta(a_t \mid \mathbf{b}_t)$ specifies the distribution over actions. The joint distribution factorizes as:

$$p(a_{1:T}, \mathbf{b}_{1:T} \mid o_{1:T}) = \prod_{t=1}^{T} \underbrace{p_\theta(a_t \mid \mathbf{b}_t)}_{\text{Action Model}} \cdot \underbrace{p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)}_{\text{Belief Transition Prior}}. \tag{1}$$

At each time step, $\mathbf{b}_t$ is modeled as a Markov Random Field with time-varying dependencies. We next describe the neuro-symbolic generative model that maps LLM-derived semantic evidence into structured belief dynamics and belief-conditioned actions.

3.2 Neuro-Symbolic Generative Model

Though LLMs provide rich semantic representations, independent outputs alone do not define a coherent joint belief distribution: they may be temporally inconsistent Steyvers et al. (2025), ignore inter-belief dependencies, and lack behavioral grounding Sclar et al. (2023). To address this, we treat LLM-derived semantic representations as contextual evidence rather than direct belief probabilities, and project them into a structured probabilistic latent-variable model with unary and pairwise belief interactions. This enables coherent temporal belief evolution and belief-conditioned decision-making under uncertainty, forming our neuro-symbolic generative model over latent belief trajectories.

Semantic-to-Potential Projection: Conditioned on the previous belief state $\mathbf{b}_{t-1}$ and the current observation o_t, we define the belief transition prior as a Gibbs distribution LeCun et al. (2006) over the joint configuration of K binary beliefs. And the corresponding energy function is:

$$E_\theta(\mathbf{b}_t) = -\sum_{i=1}^{K} \phi_i(b_{t,i}) - \sum_{1 \le i < j \le K} \phi_{ij}(b_{t,i}, b_{t,j}), \tag{2}$$

where $\phi_i(b_{t,i})$ and $\phi_{ij}(b_{t,i}, b_{t,j})$ denote the unary and pairwise potentials, respectively. These potentials are parameterized by LLM-based semantic evidence and are implicitly conditioned on $(\mathbf{b}_{t-1}, o_t)$. The belief transition prior is defined as $p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t) = \frac{\exp\left(-E_\theta(\mathbf{b}_t)\right)}{Z_q(\mathbf{b}_{t-1}, o_t)}$, where $Z_q(\mathbf{b}_{t-1}, o_t)$ is the partition function that normalizes the distribution over all 2^K belief configurations.

We next describe how LLM embeddings are used to parameterize the unary and pairwise potentials. For each belief i, we query a frozen LLM to obtain semantic embedding representations that assess whether the resident is likely to hold the target belief under current observation o_t, while accounting for belief persistence over time. Specifically, we construct two prompt variants that differ only in the assumed previous belief state: one with $b_{t-1,i} = 1$ ("Yes") and one with $b_{t-1,i} = 0$ ("No"), yielding $\mathbf{h}_{t,i}^{\text{Yes}} = \text{LLM}(b_{t,i} \mid o_t,\ b_{t-1,i}{=}1)$ and $\mathbf{h}_{t,i}^{\text{No}} = \text{LLM}(b_{t,i} \mid o_t,\ b_{t-1,i}{=}0)$. To incorporate belief history into current evidence extraction in a differentiable and semantically consistent manner, we take the expectation of the LLM embedding over the previous belief state under the belief transition model:

$$\mathbf{h}_{t,i} = p_\theta(b_{t-1,i} = 1)\, \mathbf{h}_{t,i}^{\text{Yes}} + \big(1 - p_\theta(b_{t-1,i} = 1)\big)\, \mathbf{h}_{t,i}^{\text{No}}. \tag{3}$$

where $p_\theta(b_{t-1,i} = 1)$ denotes the model-implied marginal belief probability from prior timestep $t{-}1$.

Unary Potential and Pairwise Potential Function: Because belief states are unobserved, the sign of the unary energy is not identifiable in an unsupervised setting: flipping the semantic interpretation of a belief can lead to an equivalent likelihood. To preserve the intended semantics of each belief, we anchor the unary potential to the LLM's semantic direction by contrasting its alignment with the "Yes" and "No" reference embeddings. We define the base unary score as $\phi_{t,i}^{\text{base}} = \tau\big(\cos(\mathbf{h}_{t,i}, \mathbf{h}_{t,i}^{\text{Yes}}) - \cos(\mathbf{h}_{t,i}, \mathbf{h}_{t,i}^{\text{No}})\big)$, where τ is a temperature parameter. This provides a semantic anchor for the latent belief while still allowing model to learn from data. We therefore add a learnable residual term to capture evidence not explained by the semantic anchor from observation:

$$\phi_{t,i} = \phi_{t,i}^{\text{base}} + \mathbf{w}_u^\top \mathrm{ReLU}(\mathbf{h}_{t,i}) + \beta_u. \tag{4}$$

We further design a pairwise embedding for each belief pair as:

$$\phi_{ij} = \mathbf{w}_p^\top \mathrm{ReLU}(\mathbf{h}_{ij}) + \beta_p. \tag{5}$$

where $\mathbf{h}_{ij} = \mathrm{LLM}(b_i, b_j)$ is derived from LLM embedding. Intuitively, $\phi_{ij} > 1$ encourages co-activation of beliefs i and j, while $0 < \phi_{ij} < 1$ discourages co-activation.

Marginal Belief Probabilities: Given the belief transition prior $p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$ defined in Eq. (2), we compute marginal belief probabilities by summing over all joint belief configurations. Let $\mathbf{b}_t = (b_{t,1}, \ldots, b_{t,K}) \in \{0,1\}^K$ denote a joint belief assignment. The marginal probability that belief i is active at time t is given by $p_\theta(b_{t,i} = 1 \mid \mathbf{b}_{t-1}, o_t) = \sum_{\mathbf{b}_t : b_{t,i}=1} p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$.

Action Generation: Because actions may depend nonlinearly on combinations of beliefs, with one belief changing the relevance of another, we use action-specific self-attention to model belief–action interactions. Given belief marginals, we construct belief-conditioned embeddings for each action $j \in \mathcal{A}$. For each belief i, we precompute two action embeddings corresponding to $b_i = 1$ and $b_i = 0$: $\mathbf{e}_{j,i}^{(1)} = \mathrm{LLM}(a_j \mid b_i{=}1)$ and $\mathbf{e}_{j,i}^{(0)} = \mathrm{LLM}(a_j \mid b_i{=}0)$, and mix them using the marginal belief probability:

$$\mathbf{x}_{t,j,i} = p_\theta(b_{t,i} = 1)\, \mathbf{e}_{j,i}^{(1)} + \big(1 - p_\theta(b_{t,i} = 1)\big)\, \mathbf{e}_{j,i}^{(0)}, \qquad \mathbf{A}_{t,j} = \mathrm{softmax}\left(\frac{\mathbf{Q}_{t,j}\mathbf{K}_{t,j}^\top}{\sqrt{d_k}}\right). \tag{6}$$

Stacking $\{\mathbf{x}_{t,j,i}\}_{i=1}^K$ yields an action-specific belief token matrix $\mathbf{X}_{t,j} \in \mathbb{R}^{K \times d}$. Self-attention maps $\mathbf{X}_{t,j}$ to an interaction-aware representation $\mathbf{Z}_{t,j} \in \mathbb{R}^{K \times d}$, where $\mathbf{Q}_{t,j}, \mathbf{K}_{t,j}, \mathbf{V}_{t,j}$ are linear projections of $\mathbf{X}_{t,j}$ Vaswani et al. (2023). The entry $\mathbf{A}_{t,j}[i,k]$ captures how belief k modulates the contribution of belief i for action j. For each candidate action j, the resulting interaction-aware features $\mathbf{Z}_{t,j}$ are aggregated into an action-level representation and mapped to an action likelihood. The likelihoods across candidate actions define $p_\theta(a_t \mid \mathbf{b}_t)$, from which the action is selected.

3.3 Inference Model

In realistic settings, latent beliefs are typically not directly observable, and ground truth of evolving beliefs is often unavailable. Here we formulate belief inference as an unsupervised latent-variable problem: given observations o_t and observed actions a_t, the model must infer the latent belief configuration that plausibly explains the behavior. Formally, this corresponds to inference over the posterior $p_\theta(\mathbf{b}_t \mid o_t, a_t)$, which is intractable due to the structured temporal dependencies and inter-belief interactions in the generative model. To enable efficient inference, we introduce an amortized variational posterior $q_\phi(\mathbf{b}_t \mid o_t, a_t)$ to approximate the intractable posterior over beliefs during learning. Because the observed action a_t provides additional evidence about the latent beliefs that likely generated the behavior, the inference model conditions on both observations and realized actions. We parameterize the variational posterior in a factorized form:

$$q_\phi(\mathbf{b}_t \mid o_t, a_t) = \prod_{i=1}^{K} q_\phi(b_{t,i} \mid \mathbf{u}_{t,i}^{\text{inf}}), \quad \mathbf{u}_{t,i}^{\text{inf}} = \mathrm{LLM}(b_{t,i} \mid o_t, a_t). \tag{7}$$

where $\mathbf{u}_{t,i}^{\text{inf}}$ is a belief-specific semantic embedding extracted by a frozen LLM. Conditioning on both o_t and a_t allows the inference model to attribute observed actions to plausible underlying beliefs. Each belief posterior is modeled as a Bernoulli distribution with logit: $\ell_{t,i} = \mathbf{W}_1\, \mathrm{ReLU}\big(\mathbf{u}_{t,i}^{\text{inf}}\big) + \mathbf{b}_1$. Here, $i \in \{1, \ldots, K\}$ indexes the belief dimension, and $\ell_{t,i} \in \mathbb{R}$ denotes the Bernoulli logit for $b_{t,i}$.

Table 1: Action prediction performance on wildfire datasets with sequential multi-belief updates. All metrics are higher-is-better (↑); the best result in each row is shown in **bold**. ‡ and † denote multi-step and single-step reasoning, respectively; for benchmarks, they indicate multi-step or single-step tasks. ⋆ denotes unsupervised latent belief learning. ○ denotes methods that rely on supervised belief learning or external feedback during belief inference. LLM Inference and SimToM are direct inference methods; SimToM uses a two-stage prompting framework.

Benchmark	Backbone	Metric	**Ours***‡		AutoToM$^{\circ\ddagger}$		LLM Inference†		FLARE$^{\circ\dagger}$		Reconciliation$^{\circ\dagger}$	
			Inter	Final	Inter	Final	Inter	Final	Inter	Final	Inter	Final
Wildfire‡ (Six Beliefs)	Qwen3-8B	Action Acc.	**0.8443**	**0.9157**	0.3860	0.6959	0.2796	0.6888	0.6790	0.7760	0.2427	0.6194
		Macro-F1	**0.7960**	**0.9023**	0.1825	0.5215	0.1629	0.4079	0.4104	0.7741	0.1522	0.6190
	Qwen3-32B	Action Acc.	**0.9201**	**0.9490**	0.3833	0.7742	0.3096	0.6888	0.6607	0.7763	0.4568	0.7373
		Macro-F1	**0.9310**	**0.9368**	0.1390	0.5330	0.1182	0.4079	0.4492	0.6556	0.1571	0.6508
	Qwen3-72B	Action Acc.	**0.8208**	**0.9005**	0.1499	0.7788	0.3277	0.6888	0.4821	0.5520	0.4741	0.8341
		Macro-F1	**0.7855**	**0.9000**	0.0790	0.5600	0.1766	0.4079	0.3409	0.5448	0.1875	0.7187
	Llama-3-8B	Action Acc.	**0.7477**	**0.8677**	0.3660	0.8387	0.3156	0.6888	0.5714	0.7230	0.6523	0.8111
		Macro-F1	**0.5842**	**0.8392**	0.2750	0.4560	0.1433	0.4079	0.4620	0.6249	0.3458	0.6193
	Llama-3-70B	Action Acc.	**0.9367**	**0.9286**	0.3746	0.8387	0.3096	0.6888	0.6607	0.7760	0.4569	0.8571
		Macro-F1	**0.9456**	**0.9151**	0.1530	0.4560	0.1182	0.4079	0.4488	0.6964	0.1571	0.5999
	Deepseek-R1 70B	Action Acc.	**0.9084**	**0.9592**	0.3862	0.8433	0.1353	0.6167	0.6250	0.6447	0.5747	0.8710
		Macro-F1	**0.9133**	**0.9500**	0.1390	0.4580	0.2287	0.4204	0.4221	0.6016	0.3170	0.6721

Belief absorption: The formulation above assumes that the theory-informed candidate belief space contains major beliefs relevant to the observed decision trajectory. If there exist unknown relevant belief constructs, likelihood evidence generated by those unknown beliefs may be incorrectly attributed to one or more candidate beliefs and thus bias belief attribution. To address this challenge, we introduce a *belief absorption* mechanism. Specifically, we introduced unconstrained meta-auxiliary belief nodes that allow alternation explanations for those action trajectory-level evidences not adequately captured by the theory-informed candidates. Because genuinely unknown beliefs cannot be measured directly, we evaluate this mechanism through controlled omission experiments. Results show that auxiliary nodes become increasingly correlated with survey-measured beliefs as these beliefs are removed from the candidate space, and the absorption signals are minimal when all survey-measured beliefs are included in the candidate space. These results are consistent with the mechanism capturing information about unknown beliefs (Appendix D).

3.4 Variational Learning Objective

Since beliefs $\mathbf{b}_t$ are latent and not directly observed, we train the model by maximizing an evidence lower bound (ELBO) Kingma & Welling (2022a); Rezende et al. (2014) on the conditional log-likelihood of actions. At each time step t, the generative model defines the joint distribution $p_\theta(a_t, \mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t) = p_\theta(a_t \mid \mathbf{b}_t)\, p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$. where $p_\theta(\mathbf{b}t \mid \mathbf{b}t-1, o_t)$ is the temporal belief transition prior defined by the energy-based belief model, and $p_\theta(a_t \mid \mathbf{b}_t)$ is the belief-conditioned action likelihood parameterized by the attention-based action model described above. Because the posterior $p_\theta(\mathbf{b}_t \mid o_t, a_t)$ does not admit a tractable form for stable end-to-end optimization and amortized inference, we approximate it with the inference model $q_\phi(\mathbf{b}_t \mid o_t, a_t)$. Applying standard variational inference, the training objective is defined as a sum of per-time-step ELBOs:

$$\mathcal{L} = -\sum_{t=1}^{T} \big(\mathcal{L}_t^{\text{act}} - \mathcal{L}_t^{\text{KL}}\big). \tag{8}$$

The action likelihood term encourages belief configurations that explain the observed action, and the KL term enforces consistency between the inference posterior and the belief transition prior:

$$\mathcal{L}_t^{\text{act}} = \mathbb{E}_{q_\phi(\mathbf{b}_t \mid o_t, a_t)}\big[\log p_\theta(a_t \mid \mathbf{b}_t)\big], \quad \mathcal{L}_t^{\text{KL}} = \mathrm{KL}(q_\phi(\mathbf{b}_t \mid o_t, a_t) \,\|\, p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)). \tag{9}$$

As mentioned earlier, the inference model conditions on the realized action a_t, whereas the generative prior does not. This standard variational asymmetry allows action information to guide posterior belief inference during training and be transferred to the generative model through the KL term. A detailed derivation of the ELBO is provided in Appendix B. At test time, only the generative components $p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$ and $p_\theta(a_t \mid \mathbf{b}_t)$ are used.

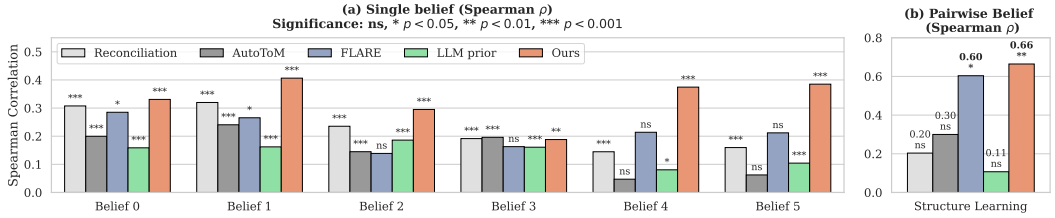


Figure 2: (a) Spearman correlation between model-predicted beliefs and human ratings for individual beliefs. (b) Spearman correlation for pairwise belief structure learning.

4 RESULTS

4.1 EXPERIMENTAL SETUP

Dataset: Existing ToM benchmarks provide controlled evaluation settings but are typically single-step and only contain one to two types of beliefs in each sample (e.g, BigToM Gandhi et al. (2023), SimpleToM Le et al. (2019), and OpenToM Xu et al. (2024)). To address this gap, we leverage a real-world wildfire evacuation dataset Kuligowski et al. (2022); Forrister et al. (2024), which contains multi-step observation–action trajectories collected from participants affected by major wildfire events (e.g., Kincade and Marshall fires). The dataset records updates of individuals' six types of belief over time through structured survey responses (e.g., Likert-scale ratings of perceived risks) and free-text descriptions of their evolving perceptions, information processing, and decision-making. It includes multiple co-existing belief variables, enabling analysis of inter-belief interactions and temporal belief evolution. The exact prompts used to extract belief- and action-conditioned hidden embeddings from the LLM are detailed in Appendix L.

We also evaluate on standard ToM benchmarks for comparison (e.g, BigToM, SimpleToM, and OpenToM mentioned above), while noting that these datasets involve limited belief variables and lack temporal structure.

Baselines: We compare against these representative baselines. *AutoToM* Zhang et al. (2025b) performs Bayesian-style ToM inference using LLM-generated hypotheses over latent mental variables. *Model Reconciliation (LLM-based)* Tang et al. (2025) uses LLMs to generate post-hoc causal explanations when model predictions disagree with observed human decisions. *FLARE* Chen et al. (2025) integrates PADM with LLM reasoning to model latent risk perceptions and predict evacuation behavior. *SimToM* Wilf et al. (2024) performs two-stage prompting with explicit perspective-taking before answering ToM questions. For the wildfire dataset, we compare against AutoToM, Model Reconciliation, and FLARE. For general ToM benchmarks, we compare against AutoToM and SimToM, as these methods are specifically proposed for ToM reasoning tasks. We also include an LLM prior baseline (direct LLM inference). LLM Inference and SimToM are direct inference baselines without belief learning. FLARE relies on supervised belief signals, while AutoToM and Reconciliation use iterative feedback during inference. In contrast, our model learns latent beliefs without supervision. Both our model and AutoToM model multi-step belief dynamics, whereas the remaining baselines are single-step.

4.2 RESULTS ON DYNAMIC MULTI-BELIEF WILDFIRE DATASET

Action Prediction Performance: For wildfire dataset, we evaluate two tasks: intermediate action prediction (4-way classification) and final evacuation prediction (binary). Accuracy alone is insufficient since biased models may over-predict dominant classes. We therefore report macro-F1 to assess prediction balance. Plain LLM shows strong bias, with very low macro-F1 despite reasonable accuracy, indicating highly imbalanced predictions. As shown in Table 1, our model improves both accuracy and macro-F1, especially on the intermediate task, indicating better capture of diverse decision patterns. These gains suggest that the learned belief graph supports both coarse-grained final decisions and fine-grained intermediate actions. Additional results on ELBO convergence and training dynamics are provided in Appendix G.

Unary Belief Correlation Results: A central goal of our framework is to infer latent cognitive beliefs that mediate decision-making. Although beliefs are unobserved during training, the post-event wildfire

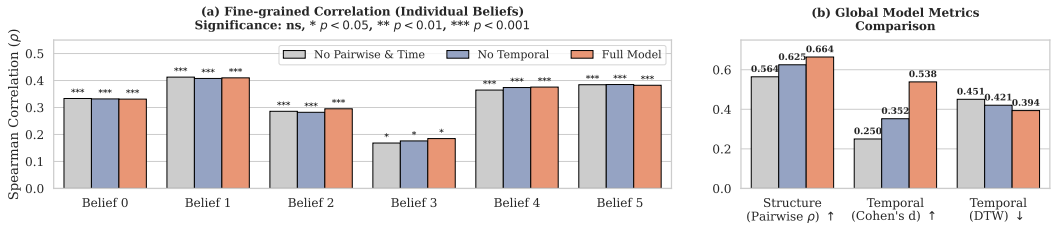


Figure 3: Ablation results: (a) Per-belief Spearman correlation under different ablations. (b) Global metrics for belief structure learning and temporal consistency.

surveys include self-reported belief assessments that can be used as proxy ground truth for evaluation. This enables validation of inferred beliefs without violating the model's latent-variable assumption. The survey measures six belief dimensions relevant to wildfire evacuation, with respondents rating each belief on a 1–5 ordinal scale. Because these ratings are subjective and not directly comparable across individuals, we evaluate belief quality using Spearman correlation between predicted belief scores and survey responses. Spearman correlation is a rank-based measure that captures monotonic agreement and is invariant to scale transformations, making it a standard choice for comparing latent belief predictions against ordinal human ratings Nan et al. (2024); Keles et al. (2021). Figure 2 (a) reports per-belief Spearman correlations for our method and all baselines. Across most belief dimensions, our model achieves substantially higher correlations than all baseline models. Results for beliefs on additional backbones are in Appendix C,

*Pairwise Belief Correlation Results:*We evaluate whether the model captures belief–belief interactions by testing if co-variation patterns among beliefs are preserved: i.e., when belief i tends to increase together with belief j in the survey data, the same tendency should also appear in the model predictions. For each belief pair (i, j) with $i < j$, we compute Spearman correlations r_{ij}^{gt} and r_{ij}^{pred} over individuals using ground-truth and predicted beliefs, respectively. Agreement between belief structures is then measured by the Spearman correlation, which evaluates whether belief pairs that co-vary more strongly in human data are ranked similarly by the model. Figure 2 (b) shows that our model best recovers the belief–belief co-variation structure, while the other baselines exhibit weaker agreement with the survey-derived interaction ranking.

Belief Scalability Analysis Our model performs belief inference via exact enumeration over the joint belief space, which grows exponentially with the number of beliefs. We further consider mean-field and MCMC as approximation strategies and compare them against exact enumeration to evaluate their efficiency and performance trade-offs. Results show that exact inference remains practical up to $K{=}15$, while mean-field provides a more efficient alternative as the belief space grows. Detailed results are provided in Appendix F.

4.3 Results on Other ToM Benchmarks

We evaluate both belief and action prediction on standard text-based ToM benchmarks (Table 2). Unlike the baselines, which use belief labels for supervision or feedback during inference, our model learns beliefs as latent variables without direct belief supervision. Its belief predictions are among the strongest in most settings, though not always the best. The model also matches or outperforms the baselines on action prediction in many cases, showing that useful decision predictions can be learned from behaviorally grounded latent beliefs without explicit belief labels. These benchmarks, however, are mostly single-step and contain only a few mutually exclusive beliefs, so there is limited need to model temporal dynamics or interactions between beliefs. Our model therefore remains competitive, but its advantage is naturally smaller in these simpler settings.

For additional analysis, we examine the role of latent beliefs through belief mediation (Appendix H), where intervening on inferred beliefs produces systematic changes in action predictions. We also analyze the action-specific self-attention patterns (Appendix G) to show how one belief can modulate the contribution of another under different target actions. The model is also stable across different prompt settings, with only minor performance changes (Appendix E). Appendix F further studies the effect of belief-space size and compares exact and approximate inference as the number of beliefs increases. Sensitivity to the temperature parameter τ is reported in Appendix G.

Table 2: Comparison across Our Model, AutoToM, SimToM, and LLM Inference on SimpleToM, BigToM, and OpenToM. ‡ and † denote multi-step and single-step reasoning, respectively (or tasks for benchmarks), ⋆ denotes unsupervised latent belief learning. ∘ denotes methods that rely on supervised belief inference or external feedback during belief inference.

Benchmark	Backbone	Our Model ⋆‡		AutoToM ∘‡		SimToM †		LLM Inference †	
		Belief Acc.	Action Acc.	Belief Acc.	Action Acc.	Belief Acc.	Action Acc.	Belief Acc.	Action Acc.
SimpleToM‡ (Single Belief)	DeepSeek-R1	0.917	**0.996**	**0.942**	0.775	0.610	0.580	0.860	0.732
	Llama-3-8B	**0.987**	**0.996**	0.905	0.661	0.620	0.900	0.689	0.505
	Llama-3-70B	0.857	**0.933**	0.979	0.847	**0.980**	0.930	0.968	0.756
	Qwen3-32B	0.948	**0.991**	0.906	0.746	**0.980**	0.950	0.947	0.767
	Qwen3-8B	**0.961**	**0.991**	0.867	0.446	0.870	0.780	0.823	0.518
BigToM‡ (Two Beliefs)	Qwen3-8B	**0.949**	**0.880**	0.695	0.700	0.615	0.780	0.830	0.675
	Llama-3-70B	0.966	0.870	**0.978**	0.870	0.530	0.540	0.812	**0.880**
	DeepSeek-R1	**0.983**	**0.870**	0.975	0.863	0.740	0.830	0.890	0.790
	Qwen2.5-72B	0.949	**0.950**	**0.985**	0.885	0.765	0.820	0.950	0.890
	Qwen3-32B	0.875	**0.890**	**0.900**	0.870	0.355	0.740	**0.900**	0.870
OpenToM‡ (Two Beliefs)	DeepSeek-R1	**0.985**	**0.940**	0.730	0.680	0.850	0.760	0.760	0.860
	Llama-3-8B	0.940	**0.970**	0.720	0.670	**0.970**	0.930	0.930	**0.970**
	Llama-3-70B	**0.945**	0.895	0.800	0.820	0.860	0.620	0.860	**0.910**
	Qwen3-32B	**0.910**	0.885	0.740	0.850	0.820	0.830	0.810	**0.910**
	Qwen3-8B	**0.845**	**0.830**	0.810	0.820	0.830	0.780	0.810	0.700

4.4 Ablation Study

We conduct ablation studies by selectively removing key elements of the proposed model. In the first ablation (No-Pairwise), we remove the pairwise belief interaction terms ϕ_{ij} from the belief transition prior, reducing the temporal MRF to independent unary potentials. In the second ablation (No-Temporal), we remove temporal belief transitions by discarding dependence on $\mathbf{b}_{t-1}$ and training the model independently at each time step. All other components are kept unchanged. Figure 3(b) compares model variants in terms of how well they recover belief–belief co-variation patterns. The full model achieves the highest Spearman correlation, while removing pairwise potentials or temporal transitions substantially degrades performance.

We evaluate belief dynamics using two complementary metrics. The Standardized Mean Difference (Cohen's d) Lakens (2013); Funder & Ozer (2019) assesses whether belief updates are preferentially aligned with action changes, capturing whether temporal belief evolution is decision-relevant rather than uniformly drifting. Dynamic time warping (DTW) distance Zhou et al. (2013); Pasteuning et al. (2025) evaluates whether the overall temporal shape of predicted belief trajectories matches human-reported belief evolution, reflecting global temporal consistency. Together, these metrics distinguish action-aligned belief updates from mere temporal smoothness; implementation details are provided in Appendix A. As shown in Figure 3(b), the full model achieves the highest Cohen's d and the lowest DTW distance among all variants. Figure 3 (a) shows per-belief Spearman correlations across ablations. We observe that even without pairwise or temporal components, the model achieves comparable performance at the single-belief level, indicating that the ELBO-based latent-variable framework alone suffices to recover marginal belief signals. Further ablations on the ELBO objective and factor graph structure, disentangling their contributions to belief learning and action prediction, are provided in Appendix H.

5 Conclusion

We introduced a structured cognitive trajectory framework for Theory-of-Mind reasoning that models mental states as dynamic belief graphs with multiple interacting beliefs jointly evolving over time. Our framework integrates LLM-derived semantic representations with structured probabilistic latent variable modeling to learn temporally coherent multi-belief dynamics and belief-conditioned decision-making from observation-action trajectories without belief supervision. Across real-world wildfire evacuation datasets containing multi-step multi-belief and standard ToM benchmarks, our approach improves action prediction while recovering interpretable latent belief trajectories and human-like belief interaction structures without using any belief ground truth. These results show that structured latent belief modeling provides a promising direction for accurate and behaviorally grounded ToM reasoning beyond static or independent belief representations.

AI USE STATEMENT

In this work, we did not use generative AI tools for tasks with required disclosure. We have not used generative AI tools to generate the scientific content of this paper, including research ideas, methodology, experiments, results, analysis, or conclusions, and the rest of the required disclosure tasks are not applicable to this work.

Additionally, we used generative AI tools for language polishing, including grammar checking and improving the clarity and readability of the manuscript. We have reviewed all AI-assisted edits to ensure that they did not change the scientific content or meaning of the paper. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses previously collected human behavioral survey data from disaster-affected populations. All datasets were collected under institutional review board (IRB) approval with informed consent from participants and were de-identified before analysis. We use these data solely for research on belief inference and behavioral prediction and do not attempt to identify individual participants.

REPRODUCIBILITY STATEMENT

We provide the full model formulation and training objective in the main text, with a detailed ELBO derivation in Appendix. Additional implementation details, ablations, and computational settings are reported in the appendix, and the exact LLM prompts used for embedding extraction are also provided in Appendix. We also provide an anonymous code repository with the submission to support reproduction of the experiments.

REFERENCES

Icek Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991. ISSN 0749-5978. doi: https://doi.org/10.1016/0749-5978(91)90020-T. URL https://www.sciencedirect.com/science/article/pii/074959789190020T. Theories of Cognitive Self-Regulation.

Alfiana Alfiana, Muhammad Azizi, Andi Primafira, I Prayana, and Srifatmawati Ahmad. Behavioral finance and its impact on corporate financial decision making. *Journal of Economic, Bussines and Accounting (COSTING)*, 8:941–946, 03 2025. doi: 10.31539/costing.v8i2.14394.

Chris L. Baker, Julian Jara-Ettinger, Rebecca Saxe, and Joshua B. Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1: 0064, 2017. doi: 10.1038/s41562-017-0064.

David Belanger and Andrew McCallum. Structured prediction energy networks, 2016. URL https://arxiv.org/abs/1511.06350.

Patricia A Champ and Hannah Brenkert-Smith. Is seeing believing? perceptions of wildfire risk over time. *Risk Analysis*, 36(4):816–830, 2016. doi: https://doi.org/10.1111/risa.12465.

Ruxiao Chen, Chenguang Wang, Yuran Sun, Xilei Zhao, and Susu Xu. From perceptions to decisions: Wildfire evacuation decision prediction with behavioral theory-informed LLMs. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 29754–29778, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1438.

Philipp Christian, Jan Kaiser, Peter C. Taylor, M. George, Simone Schütz-Bosbach, and Alexander Soutschek. Belief updating during social interactions: Neural dynamics and causal role of dorsomedial prefrontal cortex. *The Journal of Neuroscience*, 44(22):e1669232024, 2024. doi: 10.1523/JNEUROSCI.1669-23.2024.

Junyoung Chung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C Courville, and Yoshua Bengio. A recurrent latent variable model for sequential data. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/b618c3210e934362ac261db280128c22-Paper.pdf.

Thomas Cova and Frank Drews. *Wildfire Protective Actions and Collective Spatial Cognition*, pp. 237–250. Routledge, 05 2024. doi: 10.4324/9781003202738-11.

Pat Croskerry. A universal model of diagnostic reasoning. *Academic Medicine*, 84(8):1022–1028, 2009.

Fred D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989. doi: 10.2307/249008.

Ana Forrister, Erica D. Kuligowski, Yuran Sun, Xiang Yan, Ruggiero Lovreglio, Thomas J. Cova, and Xilei Zhao. Analyzing risk perception, evacuation decision and delay time: A case study of the 2021 marshall fire in colorado. *Travel Behaviour and Society*, 35:100729, 2024. ISSN 2214-367X. doi: https://doi.org/10.1016/j.tbs.2023.100729. URL https://www.sciencedirect.com/science/article/pii/S2214367X23001801.

David C. Funder and Daniel J. Ozer. Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2):156–168, 2019. doi: 10.1177/2515245919847202.

Kanishk Gandhi, Jan-Philipp Fraenken, Tobias Gerstenberg, and Noah Goodman. Understanding social reasoning in language models with language models. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 13518–13529. Curran Associates, Inc., 2023.

URL https://proceedings.neurips.cc/paper_files/paper/2023/file/2b9efb085d3829a2aadffab63ba206de-Paper-Datasets_and_Benchmarks.pdf.

Guiyang Hou, Wenqi Zhang, Yongliang Shen, Zeqi Tan, Sihao Shen, and Weiming Lu. Egosocialarena: Benchmarking the social intelligence of large language models from a first-person perspective, 2025. URL https://arxiv.org/abs/2410.06195.

Yuyi Huang, Runzhe Zhan, Lidia S. Chao, Ailin Tao, and Derek F. Wong. Path drift in large reasoning models: How first-person commitments override safety. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 19602–19616, Suzhou, China, November 2025. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.990. URL https://aclanthology.org/2025.emnlp-main.990/.

Julian Jara-Ettinger. Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29:105–110, 2019. ISSN 2352-1546. doi: https://doi.org/10.1016/j.cobeha.2019.04.010. URL https://www.sciencedirect.com/science/article/pii/S2352154618302055. Artificial Intelligence.

Kunal Jha, Tuan Anh Le, Chuanyang Jin, Yen-Ling Kuo, Joshua B. Tenenbaum, and Tianmin Shu. Neural amortized inference for nested multi-agent reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):530–537, Mar. 2024. doi: 10.1609/aaai.v38i1.27808.

Chuanyang Jin, Yutong Wu, Jing Cao, Jiannan Xiang, Yen-Ling Kuo, Zhiting Hu, Tomer Ullman, Antonio Torralba, Joshua B. Tenenbaum, and Tianmin Shu. Mmtom-qa: Multimodal theory of mind question answering, 2024.

Umit Keles, Chujun Lin, and Ralph Adolphs. Efficient prediction of trait judgments from faces using deep neural networks. *PsyArXiv*, 2021. doi: 10.31234/osf.io/t7hw4.

Hyunwoo Kim, Melanie Sclar, Xuhui Zhou, Ronan Bras, Gunhee Kim, Yejin Choi, and Maarten Sap. FANToM: A benchmark for stress-testing machine theory of mind in interactions. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 14397–14413, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.890. URL https://aclanthology.org/2023.emnlp-main.890/.

Hyunwoo Kim, Melanie Sclar, Tan Zhi-Xuan, Lance Ying, Sydney Levine, Yang Liu, Joshua B. Tenenbaum, and Yejin Choi. Hypothesis-driven theory-of-mind reasoning for large language models. In *Second Conference on Language Modeling*, 2025a. URL https://openreview.net/forum?id=yGQqTuSJPK.

Hyunwoo Kim, Melanie Sclar, Tan Zhi-Xuan, Lance Ying, Sydney Levine, Yang Liu, Joshua B. Tenenbaum, and Yejin Choi. Hypothesis-driven theory-of-mind reasoning for large language models, 2025b. URL https://arxiv.org/abs/2502.11881.

Markus T. Kinateder, Erica D. Kuligowski, Paul A. Reneke, et al. Risk perception in fire evacuation behavior revisited: definitions, related concepts, and empirical evidence. *Fire Safety Science Review*, 4(1), 2015. doi: 10.1186/s40038-014-0005-z.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022a. URL https://arxiv.org/abs/1312.6114.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022b. URL https://arxiv.org/abs/1312.6114.

Michael J. Kinsey, Steven M. V. Gwynne, Erica D. Kuligowski, et al. Cognitive biases within decision making during fire evacuations. *Fire Technology*, 55:465–485, 2019a. doi: 10.1007/s10694-018-0708-0.

Michael J. Kinsey, Steven M. V. Gwynne, Erica D. Kuligowski, et al. Cognitive biases within decision making during fire evacuations. *Fire Technology*, 55(2):465–485, 2019b. doi: 10.1007/s10694-018-0708-0.

Rahul G. Krishnan, Uri Shalit, and David Sontag. Structured inference networks for nonlinear state space models, 2016. URL https://arxiv.org/abs/1609.09869.

Erica D. Kuligowski, Xilei Zhao, Ruggiero Lovreglio, Ningzhe Xu, Kaitai Yang, Aaron Westbury, Daniel Nilsson, and Nancy Brown. Modeling evacuation decisions in the 2019 kincade fire in california. *Safety Science*, 146:105541, 2022. ISSN 0925-7535. doi: https://doi.org/10.1016/j.ssci.2021.105541. URL https://www.sciencedirect.com/science/article/pii/S0925753521003842.

Daniël Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, 4:863, November 2013. doi: 10.3389/fpsyg.2013.00863.

Matthew Le, Y-Lan Boureau, and Maximilian Nickel. Revisiting the evaluation of theory of mind through question answering. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5872–5877, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1598. URL https://aclanthology.org/D19-1598/.

Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu Jie Huang. A tutorial on energy-based learning. In Gökhan Bakir, Thomas Hofmann, Bernhard Schölkopf, Alexander Smola, and Ben Taskar (eds.), *Predicting Structured Data*. MIT Press, 2006.

Michael K. Lindell and Ronald W. Perry. The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis*, 32(4):616–632, 2012. doi: https://doi.org/10.1111/j.1539-6924.2011.01647.x.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024. doi: 10.1162/tacl_a_00638.

Yiping Lyu and Thomas Adams. Preparing for real-time weather risk management: the decision models of household evacuation under uncertainty for taiwanese and us residents. *Natural Hazards*, 114:405–425, 2022. doi: 10.1007/s11069-022-05395-8.

Jason Nan, Matthew S. Herbert, Suzanna Purpura, Andrea N. Henneken, Dhakshin Ramanathan, and Jyoti Mishra. Personalized machine learning-based prediction of wellbeing and empathy in healthcare professionals. *Sensors*, 24(8):2640, 2024.

Jordi Obiols and German Berrios. The historical roots of theory of mind: The work of james mark baldwin. *History of psychiatry*, 20:377–92, 09 2009. doi: 10.1177/0957154X08337334.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022. URL https://arxiv.org/abs/2203.02155.

Jasmin M. Pasteuning, Caroline Broeder, Milou S.C. Sep, Bernet M. Elzinga, Brenda W.J.H. Penninx, Christiaan H. Vinkers, and Erik J. Giltay. Dynamic time warping to model daily life stress reactivity in a clinical and non-clinical sample – an ecological momentary assessment study. *Journal of Affective Disorders*, 390:119833, 2025. ISSN 0165-0327. doi: https://doi.org/10.1016/j.jad.2025.119833.

Douglas Paton. Disaster risk reduction: Psychological perspectives on preparedness. *Australian Journal of Psychology*, 71(4):327–341, 2019. doi: 10.1111/ajpy.12237.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, pp. II–1278–II–1286. JMLR.org, 2014.

Sander Schulhoff, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, Aayush Gupta, HyoJung Han, Sevien Schulhoff, Pranav Sandeep Dulepet, Saurav Vidyadhara, Dayeon Ki, Sweta Agrawal, Chau Pham, Gerson C. Kroiz, Feileen Li, Hudson Tao, Ashay Srivastava, Hevander Da Costa, Saloni Gupta, Megan L. Rogers, Inna Goncearenco, Giuseppe Sarli, Igor Galynker, Denis Peskoff, Marine Carpuat, Jules White, Shyamal Anadkat, Alexander Hoyle, and Philip Resnik. The prompt report: A systematic survey of prompting techniques. *arXiv preprint arXiv:2406.06608*, 2024.

Melanie Sclar, Sachin Kumar, Peter West, Alane Suhr, Yejin Choi, and Yulia Tsvetkov. Minding language models' (lack of) theory of mind: A plug-and-play multi-character belief tracker. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13960–13980, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.780. URL https://aclanthology.org/2023.acl-long.780/.

Freda Shi, Xinyun Chen, Kanishka Misra, Nathan Scales, David Dohan, Ed Chi, Nathanael Schärli, and Denny Zhou. Large language models can be easily distracted by irrelevant context, 2023. URL https://arxiv.org/abs/2302.00093.

Haojun Shi, Suyu Ye, Xinyu Fang, Chuanyang Jin, Leyla Isik, Yen-Ling Kuo, and Tianmin Shu. Muma-tom: Multi-modal multi-agent theory of mind, 2025.

Mark Steyvers, Hector Tejeda, Aviral Kumar, et al. What large language models know and what people think they know. *Nature Machine Intelligence*, 7:221–231, 2025. doi: 10.1038/s42256-024-00976-7. URL https://doi.org/10.1038/s42256-024-00976-7.

Yinxu Tang, Stylianos Loukas Vasileiou, Vincent Derkinderen, and William Yeoh. Model reconciliation via cost-optimal explanations in probabilistic logic programming. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Karthik Valmeekam, Matthew Marquez, Sarath Sreedharan, and Subbarao Kambhampati. On the planning abilities of large language models - a critical investigation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=X6dEqXIsEW.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023. URL https://arxiv.org/abs/1706.03762.

Alex Wilf, Sihyun Lee, Paul Pu Liang, and Louis-Philippe Morency. Think twice: Perspective-taking improves large language models' theory-of-mind capabilities. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8292–8308, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.451. URL https://aclanthology.org/2024.acl-long.451/.

Heinz Wimmer and Josef Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103–128, 1983.

Yuxi Xia, Pedro Henrique Luz De Araujo, Klim Zaporojets, and Benjamin Roth. Influences on LLM calibration: A study of response agreement, loss functions, and prompt styles. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3740–3761, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.188. URL https://aclanthology.org/2025.acl-long.188/.

Hainiu Xu, Runcong Zhao, Lixing Zhu, Jinhua Du, and Yulan He. Opentom: A comprehensive benchmark for evaluating theory-of-mind reasoning capabilities of large language models, 2024. URL https://arxiv.org/abs/2402.06044.

Bo Yang, Jiaxian Guo, Yusuke Iwasawa, and Yutaka Matsuo. Large language models as theory of mind aware generative agents with counterfactual reflection, 2025. URL https://arxiv.org/abs/2501.15355.

Lance Ying, Tan Zhi-Xuan, Lionel Wong, Vikash Mansinghka, and Joshua B. Tenenbaum. Understanding epistemic language with a language-augmented Bayesian theory of mind. *Transactions of the Association for Computational Linguistics*, 13:613–637, 2025. doi: 10.1162/tacl_a_00752. URL https://aclanthology.org/2025.tacl-1.30/.

Zhehao Zhang, Ryan A. Rossi, Branislav Kveton, Yijia Shao, Diyi Yang, Hamed Zamani, Franck Dernoncourt, Joe Barrow, Tong Yu, Sungchul Kim, Ruiyi Zhang, Jiuxiang Gu, Tyler Derr, Hongjie Chen, Junda Wu, Xiang Chen, Zichao Wang, Subrata Mitra, Nedim Lipka, Nesreen Ahmed, and Yu Wang. Personalization of large language models: A survey, 2025a. URL https://arxiv.org/abs/2411.00027.

Zhen Zhang, Mohammed Haroon Dupty, Fan Wu, Javen Qinfeng Shi, and Wee Sun Lee. Factor graph neural networks, 2023. URL https://arxiv.org/abs/2308.00887.

Zhining Zhang, Chuanyang Jin, Mung Yao Jia, Shunchi Zhang, and Tianmin Shu. Autotom: Scaling model-based mental inference via automated agent modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b.

Tony Z. Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. Calibrate before use: Improving few-shot performance of language models, 2021. URL https://arxiv.org/abs/2102.09690.

Feng Zhou, Fernando De la Torre, and Jessica K. Hodgins. Hierarchical aligned cluster analysis for temporal clustering of human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):582–596, 2013. doi: 10.1109/TPAMI.2012.137.

A METRICS

This appendix details the quantitative metrics used to evaluate belief prediction quality, belief interaction structure, and temporal belief dynamics. All metrics are computed *post hoc* for evaluation only and are not used for training.

A.1 STANDARDIZED MEAN DIFFERENCE (ACTION-CONDITIONED BELIEF CHANGE)

Let $\mathbf{b}_t \in [0,1]^K$ denote the vector of belief marginals at time t, and let a_t denote the discrete action at time t. We define the belief change magnitude between consecutive steps as

$$\Delta \mathbf{b}_t = \mathbf{b}_t - \mathbf{b}_{t-1}, \qquad x_t = \|\Delta \mathbf{b}_t\|_1 = \sum_{i=1}^{K} |b_{t,i} - b_{t-1,i}|. \tag{10}$$

We introduce an action-change indicator $y_t = \mathbb{I}[a_t \neq a_{t-1}]$, where $y_t = 1$ indicates an action change and $y_t = 0$ otherwise. To quantify whether belief updates are selectively amplified at action-changing timesteps, we compute the standardized mean difference (Cohen's d):

$$d = \frac{\mu_1 - \mu_0}{s_{\text{pooled}}}, \tag{11}$$

where $\mu_1 = \mathbb{E}[x_t \mid y_t = 1]$ and $\mu_0 = \mathbb{E}[x_t \mid y_t = 0]$ are the conditional means of belief change magnitude, and

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2}{n_1 + n_0 - 2}} \tag{12}$$

is the pooled standard deviation computed from the within-group variances s_1^2, s_0^2 and sample sizes n_1, n_0.

A larger positive d indicates that belief updates are stronger at timesteps where actions change, reflecting action-aligned belief dynamics rather than uniform temporal drift.

A.2 DYNAMIC TIME WARPING (DTW)

To measure trajectory-level alignment between predicted and ground-truth belief dynamics, we use Dynamic Time Warping (DTW). Let $b_{t,i}^{\text{pred}}$ and $b_{t,i}^{\text{gt}}$ denote the predicted and ground-truth belief values for belief i at time t, respectively. For each belief dimension i and individual sequence, we compute the DTW distance

$$\text{DTW}_i = \text{DTW}\Big(\{b_{t,i}^{\text{pred}}\}_{t=1}^{T}, \{b_{t,i}^{\text{gt}}\}_{t=1}^{T}\Big), \tag{13}$$

where $\text{DTW}(\cdot,\cdot)$ denotes the standard dynamic time warping distance.

To account for sequence length, we normalize by the valid trajectory length and report the average DTW across beliefs and individuals:

$$\text{DTW}_{\text{avg}} = \mathbb{E}_{i,n}\left[\frac{\text{DTW}_{i,n}}{T_n}\right], \tag{14}$$

where n indexes individuals and T_n is the valid length of sequence n. Lower DTW indicates better temporal alignment between predicted and ground-truth belief trajectories.

A.3 SPEARMAN RANK CORRELATION

Given two variables $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$, let $\{R_i\}$ and $\{S_i\}$ denote their rank-transformed values. The Spearman correlation coefficient is defined as

$$\rho = \frac{\sum_{i=1}^{N}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{N}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{N}(S_i - \bar{S})^2}}, \tag{15}$$

where $\bar{R}$ and $\bar{S}$ are the mean ranks.

B Derivation of the Evidence Lower Bound (ELBO)

In this section, we provide the detailed derivation of the training objective ELBO Kingma & Welling (2022b). We formulate the objective by decomposing the sequence likelihood into time-step specific contributions. Specifically, we derive the ELBO for a single time step t, conditioned on the history (encapsulated in the previous belief state $\mathbf{b}_{t-1}$) and the current observation o_t.

B.1 Problem Definition

Let $\mathbf{b}_t$ denote the discrete latent belief vector, a_t denote the observed action, and o_t denote the observation at time t. The generative model factorizes the joint distribution at step t as:

$$p_\theta(a_t, \mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t) = p_\theta(a_t \mid \mathbf{b}_t)\, p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t), \tag{16}$$

where $p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$ is the belief transition prior and $p_\theta(a_t \mid \mathbf{b}_t)$ is the action likelihood.

We aim to maximize the conditional log-likelihood of the observed action a_t given the previous state:

$$\log p_\theta(a_t \mid \mathbf{b}_{t-1}, o_t) = \log \sum_{\mathbf{b}_t} p_\theta(a_t, \mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t). \tag{17}$$

B.2 Derivation via Jensen's Inequality

We introduce a variational approximation $q_\phi(\mathbf{b}_t \mid o_t, a_t)$ to approximate the posterior distribution of the beliefs. Rewriting the log-likelihood by multiplying and dividing by q_ϕ:

$$\log p_\theta(a_t \mid \mathbf{b}_{t-1}, o_t) = \log \sum_{\mathbf{b}_t} q_\phi(\mathbf{b}_t \mid o_t, a_t) \frac{p_\theta(a_t, \mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)}{q_\phi(\mathbf{b}_t \mid o_t, a_t)}. \tag{18}$$

This summation is equivalent to the expectation under the variational distribution:

$$\log p_\theta(a_t \mid \mathbf{b}_{t-1}, o_t) = \log \mathbb{E}_{\mathbf{b}_t \sim q_\phi} \left[\frac{p_\theta(a_t, \mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)}{q_\phi(\mathbf{b}_t \mid o_t, a_t)} \right]. \tag{19}$$

Applying Jensen's Inequality ($\log \mathbb{E}[X] \geq \mathbb{E}[\log X]$) yields the Evidence Lower Bound (ELBO) for time step t:

$$\log p_\theta(a_t \mid \mathbf{b}_{t-1}, o_t) \geq \mathbb{E}_{\mathbf{b}_t \sim q_\phi} \left[\log \frac{p_\theta(a_t, \mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)}{q_\phi(\mathbf{b}_t \mid o_t, a_t)} \right]. \tag{20}$$

B.3 Objective Function Decomposition

We expand the joint probability in the numerator using the chain rule:

$$\text{ELBO}_t = \mathbb{E}_{q_\phi} \left[\log \frac{p_\theta(a_t \mid \mathbf{b}_t)\, p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)}{q_\phi(\mathbf{b}_t \mid o_t, a_t)} \right]. \tag{21}$$

By separating the terms, we obtain the final objective function consisting of a reconstruction term and a regularization term:

$$\begin{aligned} \text{ELBO}_t &= \mathbb{E}_{q_\phi}\big[\log p_\theta(a_t \mid \mathbf{b}_t)\big] + \mathbb{E}_{q_\phi}\left[\log \frac{p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)}{q_\phi(\mathbf{b}_t \mid o_t, a_t)}\right] \\ &= \underbrace{\mathbb{E}_{q_\phi(\mathbf{b}_t \mid o_t, a_t)}\big[\log p_\theta(a_t \mid \mathbf{b}_t)\big]}_{\mathcal{L}_t^{\text{act}}\text{: Action Reconstruction}} - \underbrace{D_{\text{KL}}\big(q_\phi(\mathbf{b}_t \mid o_t, a_t) \,\|\, p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)\big)}_{\mathcal{L}_t^{\text{KL}}\text{: Belief Consistency}}. \end{aligned} \tag{22}$$

The total training objective is the sum of these negative step-wise ELBOs over the sequence length T:

$$\mathcal{L} = -\sum_{t=1}^{T} \text{ELBO}_t. \tag{23}$$

C RESULTS ACROSS DIFFERENT LLM MODELS

We have extended our experiments to multiple LLM backbone including Qwen3-8B, Qwen3-32B, Qwen2.5-72B, LLaMA3-8B, LLaMA-3-70B, and DeepSeek-R1 70B.

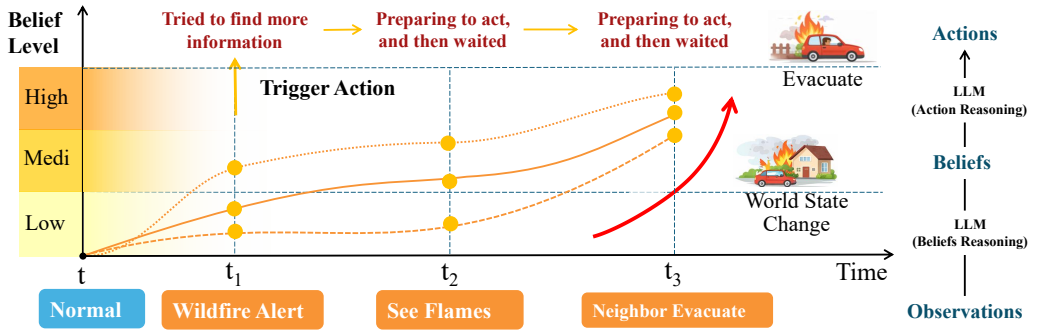


Figure 4: An example of how belief trajectories evolve with observations in high-stakes decision-making like wildfire evacuation, actions are triggerred upon threshold crossing.



Figure 5: Results on wildfire dataset across all LLM backbone. Including belief spearman and action prediction accuracy

D BELIEF ABSORPTION BEYOND THE PREDEFINED BELIEF SPACE

Our predefined belief space is carefully constructed based on established behavioral theory and domain knowledge. However, in more general settings, some behaviorally relevant beliefs may fall outside this predefined space. To improve the flexibility and generalizability of the framework, we introduce a *belief absorption* mechanism that adds an additional pseudo belief node to capture residual belief information not represented by the explicitly modeled beliefs.

The pseudo node is not associated with any predefined beliefs. Its input is constructed by averaging the observation embeddings of the currently included beliefs, and its belief probability is inferred using the same inference head as the original belief nodes, $q_{p,t} = P(z_{p,t} = 1 \mid o_t)$. The pseudo belief is then treated as an additional latent belief variable and passed to the action decoder together with the predefined beliefs, $p_\theta(a_t \mid z_{1,t}, \ldots, z_{K,t}, z_{p,t})$. Because the pseudo node has no predefined semantics, it can use the original action objective and ELBO to capture behaviorally relevant residual information that is not represented by the predefined belief space.

To test whether the pseudo node actually absorbs residual information, we use two complementary measures. We first quantify its behavioral relevance by comparing the final action distributions when the pseudo belief is active and inactive using the Jensen–Shannon divergence, $D_{\mathrm{JS}}^{(q_p)}(p(a \mid z_p = 1, o) \,\|\, p(a \mid z_p = 0, o))$. This metric measures how much action-relevant variation is carried by the pseudo node without relying on action correctness. As shown in Fig. 6a, it decreases from 1.87×10^{-4} for $K = 2$ to 1.23×10^{-4} for $K = 4$ and 0.06×10^{-4} for $K = 6$.

We then examine whether the pseudo node captures the semantics of omitted beliefs using the partial Spearman correlation, $\rho_{\mathrm{S}}(q_p, y_k \mid \mathbf{Y}_{\mathrm{included}})$. By controlling for the explicitly modeled beliefs, this metric isolates the semantic information uniquely captured by the pseudo node. As shown in Fig. 6b, the mean absolute correlations are 0.106 for $K = 2$ and 0.102 for $K = 4$. Together, these results show that the pseudo node captures residual semantic and action-relevant information when the predefined belief space is incomplete, while becoming nearly inactive when all beliefs are explicitly modeled.

The absorption mechanism can therefore diagnose whether the current belief space is sufficiently expressive. As more relevant beliefs are explicitly included, the action-absorption score should decrease toward zero, as observed from $K = 2$ to $K = 6$ in Fig. 6a. A non-negligible score indicates that the pseudo node still captures behaviorally relevant information missing from the current belief space. In this case, methods such as AutoToM, ThoughtTracing, and LaBToM Zhang et al. (2025b); Kim et al. (2025b); Ying et al. (2025) can be used to propose new candidate beliefs for inclusion in the explicit belief space.

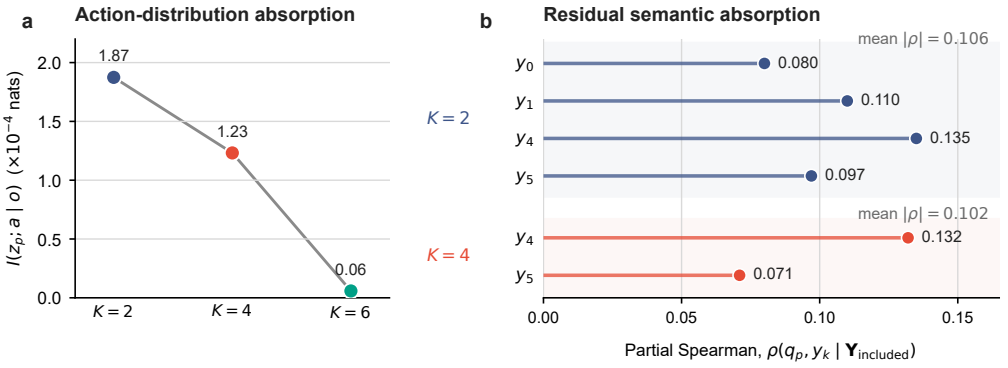


Figure 6: Belief absorption results. The pseudo node captures omitted belief semantics, while its influence on the action distribution decreases as more predefined beliefs are included.

E PROMPTS SENSITIVITY ANALYSIS

Prior work has shown that prompt perspective alone can affect reasoning outcomes (first-person perspective and third-person perspective) Huang et al. (2025), and it is argued that first-person prompting better aligns with natural human cognition in social reasoning tasks Hou et al. (2025). In disaster contexts, this is especially relevant due to optimism bias, where individuals underestimate their own risk while judging others as more vulnerable Kinsey et al. (2019b). Motivated by this, our original setup adopts first-person prompting. Here, we further compare it with third-person prompting. In addition, prior work has shown that performance can vary across prompting strategies such as chain-of-thought (CoT) and few-shot prompting Xia et al. (2025); Schulhoff et al. (2024). Therefore, we also evaluate CoT and few-shot prompting under the first-person setting to assess robustness to prompt design.

As shown in Table 3, compared to third-person prompting, first-person prompting consistently performs better in recovering belief structure, suggesting that adopting the agent's perspective enables the LLM to more naturally reproduce human cognitive biases such as optimism bias. In contrast, third-person prompting weakens this alignment and leads to less stable belief estimates. Meanwhile, other prompting strategies (e.g., CoT, few-shot, probabilistic prompting) exhibit relatively minor and inconsistent differences, indicating that prompt format has limited impact compared to the choice of perspective.

	First person	Third person	Probabilistic prompting	Few-shot prompting	CoT prompting
Belief 0	0.3311	0.3207	0.3588	0.3638	0.3495
Belief 1	0.4099	0.4241	0.4101	0.4021	0.4055
Belief 2	0.2952	0.1046	0.2504	0.2478	0.2396
Belief 3	0.1848	0.1029	0.2628	0.1747	0.2083
Belief 4	0.3757	0.3329	0.3604	0.3614	0.3223
Belief 5	0.3824	0.3241	0.3680	0.3551	0.3327
p-value 0	3.27×10^{-6}	6.83×10^{-6}	3.98×10^{-7}	2.67×10^{-7}	8.25×10^{-7}
p-value 1	4.73×10^{-9}	1.20×10^{-9}	4.62×10^{-9}	9.76×10^{-9}	7.10×10^{-9}
p-value 2	3.72×10^{-5}	1.52×10^{-1}	5.10×10^{-4}	5.85×10^{-4}	8.97×10^{-4}
p-value 3	1.09×10^{-2}	1.59×10^{-1}	2.59×10^{-4}	1.62×10^{-2}	4.02×10^{-3}
p-value 4	9.98×10^{-8}	2.86×10^{-6}	3.50×10^{-7}	3.24×10^{-7}	6.12×10^{-6}
p-value 5	5.64×10^{-8}	5.37×10^{-6}	1.89×10^{-7}	5.34×10^{-7}	2.92×10^{-6}
Pairwise correlation	0.6643	0.4071	0.7000	0.6179	0.4036
Pairwise p-value	6.91×10^{-3}	1.32×10^{-1}	3.67×10^{-3}	1.41×10^{-2}	1.36×10^{-1}
Intermediate action Acc	0.8443	0.8439	0.8712	0.8932	0.8057
Final action acc	0.9157	0.8383	0.8567	0.9128	0.8735

Table 3: Comparison across different prompting strategies. Compared on eliefs spearman correlations and action predictions accuracy.

F BELIEFS SCALABILITY TEST

In our current setting, the number of belief dimensions K=6 is not arbitrarily chosen by us, but follows the behavioral theory, PADM, and PADM-guided survey design for the wildfire evacuation survey [9, 10]. To further analyze the effect of belief dimensionality, we conduct experiments with K=2,4,6,10,15.

For K<6, we use subsets of the original belief variables. As shown in Table 4. When reducing the belief space to K=4, the model still preserves strong performance, while K=2 leads to a sharp drop, indicating that too few beliefs fail to capture sufficient decision-relevant information. Also, under K=2, belief estimation metrics appear higher. This is because the model is forced to rely on the limited belief variables to explain actions, and therefore must align them more closely with the true human beliefs.

When extending the belief space, a key challenge is that the number of joint belief configurations grows exponentially as 2^K, which may lead to computational intractability for large K. To address this, we consider standard approximate inference methods, including mean-field and MCMC. As shown in Table 5, exact inference remains efficient in our setting due to fully vectorized GPU computation, with only modest increases in runtime as K grows. Meanwhile, mean-field provides the fastest approximation with small performance degradation, whereas MCMC is significantly slower due to its sequential nature and limited parallelizability on GPUs. These results indicate that while exact inference is sufficient for the current theory-grounded setting, the framework can be naturally extended to scalable approximate inference methods when K becomes large.

	2 beliefs	4 beliefs	Original Setup (6 beliefs)
Belief 0	-	0.3187 (7.83×10^{-6})	0.3311 (3.27×10^{-6})
Belief 1	-	0.4067 (6.33×10^{-9})	0.4099 (4.73×10^{-9})
Belief 2	0.2637 (2.47×10^{-4})	0.1795 (1.35×10^{-2})	0.2952 (3.72×10^{-5})
Belief 3	0.3315 (3.16×10^{-6})	0.1100 (1.32×10^{-1})	0.1848 (1.09×10^{-2})
Belief 4	-	-	0.3757 (9.98×10^{-8})
Belief 5	-	-	0.3824 (5.64×10^{-8})
Pairwise corr.	-	-	0.6643 (6.91×10^{-3})
Action acc	0.6180	0.8349	0.8443
Final action acc	0.9041	0.9140	0.9157

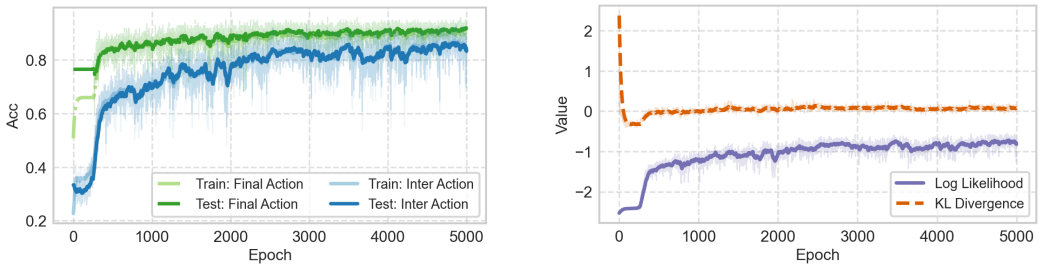
Table 4: Ablation study on the number of beliefs $K \in \{2, 4, 6\}$. Each entry reports Spearman correlation with p-value in parentheses.

Methods	Intermediate Acc	Final Acc	Inference Time (s)
K=6 (Exact)	0.8443	0.9157	0.003031
K=6 (MF)	0.8240	0.9167	0.001569
K=6 (MCMC)	0.8223	0.9204	0.018991
K=10 (Exact)	-	-	0.003165
K=10 (MF)	-	-	0.001895
K=10 (MCMC)	-	-	0.031219
K=15 (Exact)	-	-	0.003718
K=15 (MF)	-	-	0.002751
K=15 (MCMC)	-	-	0.043742

Table 5: Scaling the number of beliefs ($K = 6, 10, 15$) and comparing approximation methods.

G MORE RESULTS

Action Prediction and ELBO change Analysis. We first examine training dynamics and action prediction performance. As shown in Figure 7b, the KL term starts at a high positive value and drops sharply in early training, reflecting rapid alignment between the inference distribution and the belief-transition prior. It then stabilizes around 0, indicating a steady balance between prior regularization and data-driven inference. Meanwhile, the action log-likelihood $\log p_\theta(a_t \mid \mathbf{b}_t)$ increases steadily, indicating that the inferred beliefs become increasingly predictive of actions. Together, these trends suggest that action-explanatory belief configurations are first identified through inference and subsequently consolidated into stable generative belief dynamics. Figure 7a shows action prediction accuracy for both intermediate actions and final evacuation decisions. The smooth convergence and stable test performance are consistent with the ELBO dynamics, confirming that improved likelihood and stabilized belief learning translate into reliable action prediction.



(a) Action prediction accuracy.

(b) ELBO component dynamics.

Figure 7: Left: Action prediction accuracy over training epochs for intermediate actions and final evacuation decisions. Right: Evolution of the action likelihood term and the KL divergence.

Sensitivity Analysis over τ. We study the effect of the temperature parameter τ in the base unary score defined in Section 3.2: $\phi_{t,i}^{\text{base}} = \tau\big(\cos(\mathbf{h}_{t,i}, \mathbf{h}_{t,i}^{\text{Yes}}) - \cos(\mathbf{h}_{t,i}, \mathbf{h}_{t,i}^{\text{No}})\big)$.

The role of τ is to anchor the semantic orientation of each belief dimension. As shown in Table 6, when τ is reduced or removed, several belief correlations flip sign, indicating instability in semantic alignment. However, the overall pairwise structure and action prediction performance remain relatively stable across different τ values. This suggests that τ mainly serves to fix the direction of the latent belief space, while having limited impact on downstream tasks. In practice, τ only needs to be sufficiently large to stabilize semantic orientation during early training.

Methods	$\tau = 15$	$\tau = 3$	$\tau = 0$
Belief 0	0.3311 (3.27×10^{-6})	0.3395 (1.76×10^{-6})	-0.3246 (5.19×10^{-6})
Belief 1	0.4099 (4.73×10^{-9})	0.3983 (1.38×10^{-8})	-0.3294 (3.69×10^{-6})
Belief 2	0.2952 (3.72×10^{-5})	-0.2590 (3.16×10^{-4})	-0.2545 (4.09×10^{-4})
Belief 3	0.1848 (1.09×10^{-2})	-0.1320 (7.03×10^{-2})	-0.2340 (1.19×10^{-3})
Belief 4	0.3757 (9.98×10^{-8})	0.2663 (2.12×10^{-4})	-0.3395 (1.77×10^{-6})
Belief 5	0.3824 (5.64×10^{-8})	0.3006 (2.64×10^{-5})	-0.3221 (6.20×10^{-6})
Pairwise corr.	0.6643 (6.91×10^{-3})	0.6150 (1.48×10^{-2})	0.6929 (4.19×10^{-3})
Action acc	0.8443	0.8680	0.8695
Final action acc	0.9157	0.8470	0.8560

Table 6: Sensitivity analysis over τ. Each belief entry reports Spearman correlation (p-value in parentheses).

Interpreting the Attention Matrix. The self-attention matrix $A_{t,j}[i,k]$ captures how belief k modulates the contribution of belief i to action j, i.e., belief–belief interactions conditioned on the target action.

We consider four intermediate actions: preparing and waiting; acting immediately (e.g., evacuating, defending property, or taking shelter); seeking additional information; and no reaction. The final action is binary, indicating whether the resident evacuates or not.

The belief set consists of six dimensions: potential damage to one's home, damage to the neighborhood, personal injury, death, injury to others/pets/livestock, and death of others/pets/livestock.

For actions associated with evacuation or immediate response (intermediate a_0, a_1 and final evacuation A_1), attention is consistently concentrated on beliefs related to personal risk and property damage (e.g., injury, death, and home destruction). This suggests that evacuation decisions are driven by the joint—and often competing—influence of personal safety and asset protection, a trade-off well documented in the Protective Action Decision Model (PADM) Lindell & Perry (2012); Champ & Brenkert-Smith (2016).

In contrast, for inaction (intermediate a_3) and non-evacuation (A_0), attention shifts toward beliefs about harm to others (e.g., pets or livestock), while personal-risk-related beliefs are relatively down-weighted. This indicates that when perceived self-risk is low, decision-making is less dominated by survival concerns and instead influenced by social and responsibility-related signals. This observation is also consistent with PADM, which highlights the role of social cues and responsibility in shaping protective behavior.

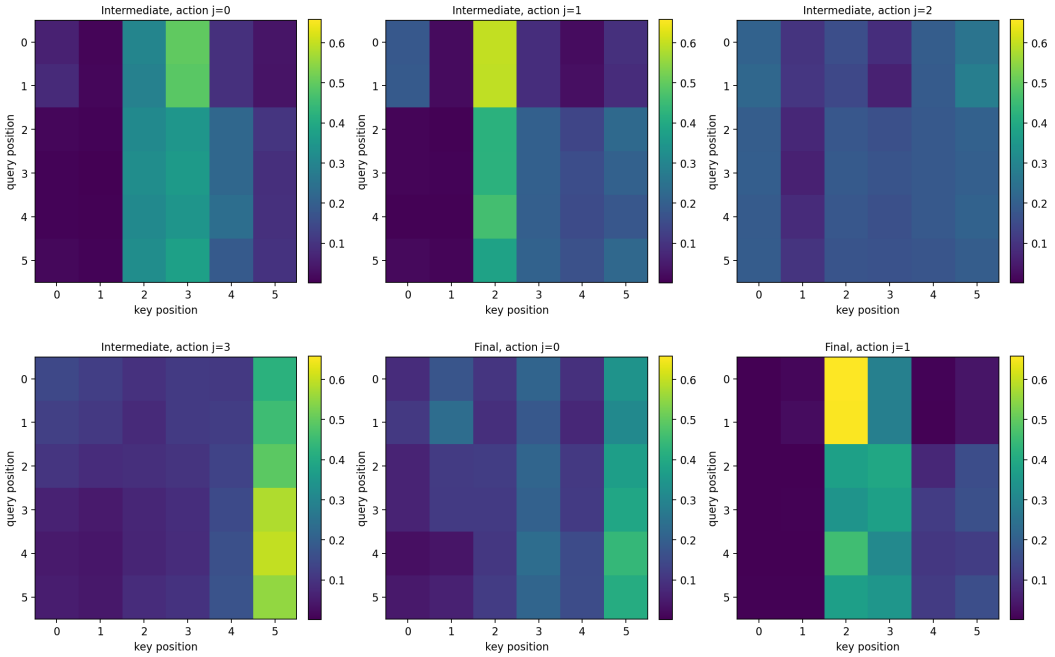


Figure 8: Attention matrix for four intermediate actions and two final actions.

H BELIEFS MEDIATING ROLE AND MORE ABLATION STUDY

H.1 BELIEFS MEDIATING ROLE

To test whether beliefs actually play a mediating role, we performed belief interventions experiments at inference time. Specifically, we keep the observation fixed and manually modify belief logits before the action decoder. We use two types of interventions: flipping extreme beliefs (e.g., $b < 0.2$ flipped to 1, $b > 0.8$ flipped to 0) flipping neutral beliefs (e.g.,$0.4 < b < 0.5$ flipped to 1, $0.5 < b < 0.6$ flipped to 0). We then measure: flip rate: how often the predicted action changes. probability change: how much the action probability shifts. Results in Table 7 show that changing beliefs leads to clear changes in action predictions. When beliefs are strong (close to 0 or 1), flipping them leads to large changes in action predictions, while flipping weak or uncertain beliefs (around 0.5) has very little effect. This suggests that beliefs function as mediators inside the model, and beliefs the model are confident about have a strong influence on actions. This also corresponds to the core theory in PADM, which strongly emphasize how critical that threat and risk perceptions are for high-stake dynamic decision-making like disaster responses.

H.2 ADDITIONAL ABLATION ON ELBO AND FACTOR GRAPHS

It is necessary to disentangle the effect of learning from data and the contribution of the ELBO objective and factor graph structure. To address this, we include additional ablations: (1) No ELBO, no factor graph: removing pairwise belief interactions so that beliefs are modeled independently, while still treating them as latent variables in the generative process. (2) No ELBO: removing the variational objective and training the model using only the action likelihood, while still maintaining latent beliefs in the generative process.

As shown in Table 8, removing ELBO leads to a substantial drop in both belief estimation and action prediction. Removing the factor graph results in a drop in pairwise correlation.

Belief interventions	Inter action flip rate	Inter action prob. change	Final action flip rate	Final action prob. change
Flipping extreme beliefs ($b < 0.2 \to 1,\ b > 0.8 \to 0$)	0.6508	0.2869	0.1633	0.2408
Flipping neutral beliefs ($0.4 < b < 0.5 \to 1,\ 0.5 < b < 0.6 \to 0$)	0.1640	0.0550	0.0306	0.0538

Table 7: Belief intervention test. With flip rate and probability change reported for intermediate actions and final actions.

	Our Model	No ELBO	No ELBO, no factor graph
Belief 0	0.3311 (3.27×10^{-6})	0.1605 (2.74×10^{-2})	0.1767 (1.50×10^{-2})
Belief 1	0.4099 (4.73×10^{-9})	0.3049 (2.00×10^{-5})	0.2707 (1.65×10^{-4})
Belief 2	0.2952 (3.72×10^{-5})	0.1826 (1.19×10^{-2})	0.2032 (5.05×10^{-3})
Belief 3	0.1848 (1.09×10^{-2})	0.1526 (3.61×10^{-2})	0.1755 (1.57×10^{-2})
Belief 4	0.3757 (9.98×10^{-8})	0.2921 (4.53×10^{-5})	0.3454 (1.13×10^{-6})
Belief 5	0.3824 (5.64×10^{-8})	0.2938 (4.07×10^{-5})	0.2983 (3.07×10^{-5})
Pairwise correlation	0.6643	0.5464	0.4857
Pairwise p value	6.91×10^{-3}	3.51×10^{-2}	6.64×10^{-2}
Action acc	0.8443	0.4794	0.4802
Final action acc	0.9157	0.7830	0.7806

Table 8: Ablation on ELBO and factor graph. Belief rows show Spearman correlation (p-value in parentheses); pairwise correlation captures belief relations; action metrics are prediction accuracy.

I BELIEFS ANALYSIS

We consider six belief dimensions derived from the survey, capturing residents' perceived wildfire risk at the property, self-safety, and others-safety levels. Specifically, the first two beliefs concern potential damage to one's home and neighborhood (Table 11).

Belief Update Trajectories. Throughout the paper, beliefs are explicitly modeled as time-varying latent states and jointly inferred across multiple time steps. Here, we complement this formulation by clustering and visualizing belief trajectories directly from the survey data, providing empirical evidence that such temporal belief dynamics are present in the data and exhibit structured heterogeneity across individuals.

To focus on the trend and shape of belief evolution rather than the absolute score level, we apply within-resident z-score normalization to each three-point trajectory. After normalization, each resident's trajectory has zero mean and unit variance across time, which removes individual-specific baseline and scale differences in subjective belief ratings and ensures that trajectories are compared based on temporal variation rather than absolute magnitude.

Specifically, for resident i (and a fixed belief dimension), let the observed survey trajectory be $\mathbf{x}_i = [x_{i,0}, x_{i,1}, x_{i,2}]$. We compute the resident-specific mean and standard deviation

$$\mu_i = \frac{x_{i,0} + x_{i,1} + x_{i,2}}{3}, \qquad \sigma_i = \sqrt{\frac{(x_{i,0} - \mu_i)^2 + (x_{i,1} - \mu_i)^2 + (x_{i,2} - \mu_i)^2}{3}}, \tag{24}$$

and normalize each time point as, where ϵ is a small constant for numerical stability.

$$z_{i,t} = \frac{x_{i,t} - \mu_i}{\sigma_i + \epsilon}, \quad t \in \{0, 1, 2\}, \tag{25}$$

As shown in Figure 9, applying k-means clustering with $k = 3$ to the normalized trajectories reveals three distinct belief trajectory types: a panic-type pattern characterized by monotonically increasing beliefs over time, a recovery-type pattern exhibiting an initial increase followed by a decline, and a calm-type pattern with relatively stable belief levels across time.

Belief Distributions. We examine the empirical score distributions of the six belief dimensions derived from the survey. As shown in Figure 10, while most beliefs exhibit relatively balanced distributions, we observe that Belief 2 (risk to one's own injury) and Belief 3 (risk to one's own death) display pronounced long-tailed distributions, with a substantial mass concentrated at low scores.

This pattern is consistent with the well-documented phenomenon of *optimistic bias* in disaster psychology Kinsey et al. (2019a); Kinateder et al. (2015); Lyu & Adams (2022), whereby individuals tend to underestimate the likelihood of negative events occurring to themselves, even when acknowledging risks to property or to others. This distributional property provides a plausible explanation for why Belief 2 and Belief 3 exhibit weaker predictive performance compared to other belief dimensions in the main results. The strong skewness and limited variance in these beliefs reduce the effective signal available for modeling and correlation analysis, leading to systematically lower performance relative to beliefs with more balanced score distributions.

Theory-guided vs. non-theory-guided belief spaces. As discussed above, our framework uses predefined belief spaces grounded in well-established psychological or behavioral theories, such as PADM in the wildfire setting studied in this paper. More broadly, theory-guided belief ontologies provide a structured latent space centered on meaningful cognitive factors, reducing the risk of learning arbitrary variables that fit observed actions but do not reflect valid belief dynamics. To examine this effect, we trained the same framework on the wildfire action-prediction task after replacing the six PADM-based beliefs with six LLM-generated beliefs that cover only a limited part of the PADM belief space, as listed below. We then compared action-prediction performance under the two belief sets.

The theory-guided belief set achieves better action-prediction performance. This supports the use of psycho-behavioral theory as a structural prior that guides the model toward established determinants of human behavior rather than incomplete or arbitrary belief factors.

Table 9: Non-theory-guided beliefs used in the comparison.

ID	Belief
B0	I believed the cost of essential items would be a concern when evacuating.
B1	I believed the cost of safety supplies would be a concern when evacuating.
B2	I believed the cost of shelter would be a concern when evacuating.
B3	I was worried about traffic during the evacuation before making an evacuation decision.
B4	I was worried about finding housing for the evacuation before making an evacuation decision.
B5	I believed that my personal belongings would be stolen before making an evacuation decision.

Table 10: Action-prediction performance using theory-guided and non-theory-guided belief spaces.

Metric	Theory-guided	Non-theory-guided
Intermediate Action Accuracy	0.844	0.795
Intermediate Action F1	0.796	0.751
Final Action Accuracy	0.915	0.7767
Final Action F1	0.902	0.6627

Table 11: Survey belief dimensions and semantic grouping.

Category	Belief ID	Belief Statement
Property Risk	B0	My home would be damaged or destroyed by fire.
	B1	My neighborhood would be damaged or destroyed by fire.
Self Safety Risk	B2	I might become injured.
	B3	I might die.
Others Safety Risk	B4	Other people, pets, or livestock might become injured.
	B5	Other people, pets, or livestock might die.

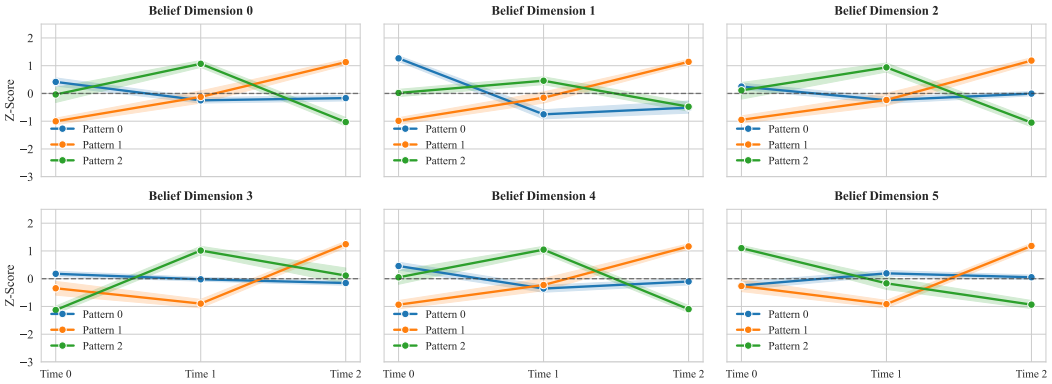


Figure 9: Clustered belief update trajectories derived from survey responses.

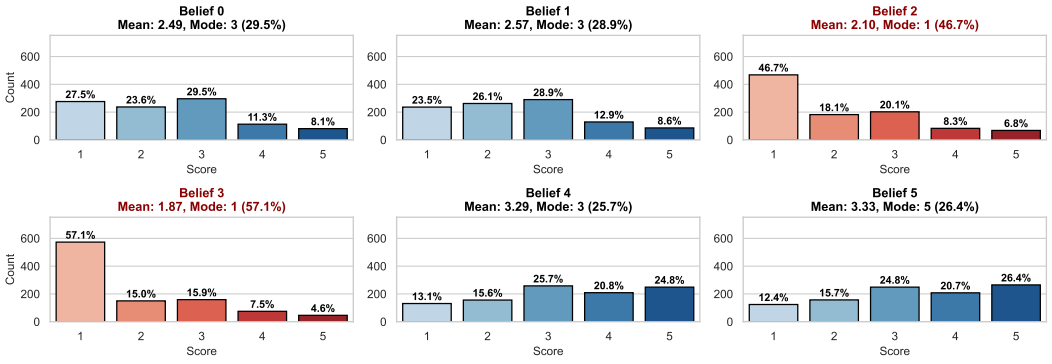


Figure 10: Empirical score distributions of the six survey belief dimensions.

J EXPERIMENTS DETAILS

We provide additional implementation details, ablations, and analyses in the appendix. All experiments are conducted on NVIDIA A100 GPUs. Training, inference, and evaluation use a single GPU, while larger LLM backbones for embedding extraction are run on 2–3 GPUs.

Why ELBO Is Necessary. Across both unary belief prediction and pairwise belief structure learning, our model consistently outperforms all baselines. These results indicate that the ELBO objective is critical for learning meaningful latent beliefs, as it forces belief variables to jointly explain observations and actions, positioning beliefs as a mediating latent layer rather than post-hoc representations. Crucially, this alignment is achieved without belief-level supervision, which is essential in real-world settings where reliable belief annotations are unavailable. In contrast, inference-time prompting or heuristic belief extraction lacks optimization-based constraints tying beliefs to actions, leading to weakly identifiable belief representations that are highly sensitive to model priors.

Ablation Study: Roles of Each Component. When removing the pairwise interaction component ϕ_{ij}, the pairwise Spearman correlation drops. This gap indicates that inter-belief structure is hard to recover from marginal belief prediction or action supervision alone. The reason is that ϕ_{ij} provides a direct modeling pathway for belief–belief dependencies, allowing the ELBO objective to allocate gradients specifically to pairwise interactions rather than relying solely on indirect signals mediated through actions or marginal belief terms. Removing temporal modeling degrades trajectory-level metrics. This occurs because the temporal component enables gradients to propagate across the entire time horizon, allowing beliefs to accumulate evidence and maintain consistency over long-term decision trajectories. In contrast, the per-belief Spearman correlation is relatively insensitive to the removal of structural or temporal components. This suggests that individual belief dimensions are primarily learned through the ELBO objective itself, which provides the dominant learning signal for aligning each belief with observed actions. Taken together, the ablation study confirms a clear division of labor: ELBO-based training determines what beliefs are present, the pairwise component learns how beliefs interact, and the temporal component captures how beliefs evolve over time.

Generality of the Framework. The framework is general at both the modeling and problem levels. At the modeling level, it imposes no restrictions on dimensionality: individual beliefs are represented by unary energies ϕ_i, belief–belief interactions by pairwise energies ϕ_{ij}, and actions by unary energy terms with a softmax, allowing belief and action spaces of arbitrary sizes to be learned jointly. For larger belief spaces ($K > 6$), approximate inference methods such as mean-field or MCMC can be used in place of exact enumeration; we empirically study these approximations in Appendix F.

At the problem level, the framework applies to human decision-making settings that follow a common psychological reasoning process, in which observations are interpreted into internal beliefs that in turn guide actions. Structure learning and an ELBO objective constrain this reasoning to ensure coherent belief propagation and behavioral grounding. Beyond human cognition, the same formulation extends to general reasoning settings in which observations give rise to latent variables that interact to support downstream conclusions. For example, in medical reasoning, clinical notes and patient reports induce latent variables representing underlying conditions, whose interactions determine diagnostic outcomes; in legal reasoning, evidence descriptions give rise to latent variables such as intent that jointly support legal judgments.

Implications for LLM-Based Reasoning Our framework positions LLMs within a structured Theory-of-Mind reasoning process guided by dynamic belief graphs, offering a principled approach for building autonomous agents. In terms of alignment, existing methods either rely on implicit optimization signals such as RLHF Ouyang et al. (2022), which lack interpretability, or on rule-based prompting, which lacks flexibility Zhao et al. (2021); Valmeekam et al. (2023). Our approach provides a middle ground by explicitly modeling latent belief graphs, enabling interpretable reasoning, auditability, and targeted intervention through belief-level modifications. For long-horizon reasoning, LLMs often suffer from context drift as irrelevant information accumulates Liu et al. (2024); Shi et al. (2023). Our temporal belief prior $p_\theta(\mathbf{b}_t \mid \mathbf{b}_{t-1}, o_t)$ serves as a compact and structured memory, maintaining temporal consistency by selectively updating and preserving beliefs over time. Finally, unlike implicit personalization methods based on retrieval or fine-tuning Zhang et al. (2025a), our framework represents individuals as explicit, evolving belief graphs. This enables transparent and controllable personalization through direct manipulation of belief states and their interactions.

K SELECTED QUESTIONS IN SURVEY DATA

Table 12: Example of the Selected Survey Questions

Category	Survey Question
Observations	**(1)** Before you decided to evacuate (or stay), did one or more emergency officials let you know that your area was under an evacuation warning, pre-evacuation notice, or an evacuation order? **(2)** Before you decided to evacuate (or stay), did someone you know tell you to evacuate or that an evacuation warning or order was issued for your area? **(3)** Before you decided to evacuate (or stay), did you see, hear, or feel flames or embers in your immediate vicinity (that is, your neighborhood)?
Beliefs	**(1)** After first finding out from the emergency official(s), how much did you believe each of the following statements on a scale of 1 to 5? **(2)** After first finding out from someone you know, how much did you agree with each of the following statements on a scale of 1 to 5? **(3)** After you first observed fire cues, how much did you believe each of the following statements on a scale of 1 to 5? **(A)** My home would be damaged or destroyed by fire. **(B)** My neighborhood would be damaged or destroyed by fire. **(C)** I might become injured. **(D)** I might die. **(E)** Other people, pets or livestock might become injured. **(F)** Other people, pets or livestock might die.
Actions	**(1)** What was your immediate reaction when the emergency official(s) first let you know? **(2)** What was your immediate reaction when someone you know first told you? **(3)** What was your immediate reaction to observing fire cues? **(A)** No reaction; I continued my activities. **(B)** I tried to find more information. **(C)** I started preparing to act, and then waited. **(D)** No reaction; I continued my activities.
Past Experiences	**(1)** Before the Kincade fire, how many times in the past 10 years did you evacuate because of a wildfire? **(2)** How long had you lived at that residence? **(3)** Before the Kincade fire, did you know that wildfires could be a problem in your community? **(4)** Before the Kincade fire, had you or others taken any measures to protect your residence from wildfires?
Demographic Info	**(1)** How old are you? **(2)** What is the highest level of education you have completed? **(3)** Which of the following categories best describes your employment status at the time the Kincade fire started? **(4)** How many children under 13 years old lived in your household at the time the Kincade fire started? **(5)** How many adults (18 years old to 64 years old) lived in your household at the time the Kincade fire started?
T-1 Beliefs	**(1)** Before the Kincade fire, how would you have described the possibility that a wildfire would threaten your property, on a scale from 1 to 5? **(2)** Please rate the extent to which you agreed or disagreed with each statement (before or at the time of the Kincade fire). **(A)** I believed that my home was well-built and safe for a wildfire. **(B)** I believed that staying at home was safer than moving to a nearby building or shelter.

L PROMPTS

Table 13: List of LLM Prompts used in the experiments.

Task Type	Prompt Template Content
Probability of beliefs given observations	**Context:** You are a resident in a wildfire evacuation scenario. Based on your history provided above, analyze the current situation. **Hypothetical Scenario:** **1.** Your Past Observation: `{past_observation}` **2.** Your Current Observation: `{observation}` **3.** Your Previous Belief State: `{last_belief}` **Hypothesis to Evaluate:** Target Belief: `{belief}` **Task:** Evaluate if you are likely to hold the "Target Belief" given the observation and their previous belief state. Answer with only the letter (A or B). **Options:** (A) Likely (B) Unlikely
Probability of actions given beliefs	**Context:** You are a resident in a wildfire evacuation scenario. Based on your history provided above, analyze the current situation. **Hypothetical Scenario:** **1.** You hold this belief: `{belief}` **2.** You are evaluating the Action: `{action}` **Task:** Analyze if this specific belief **encourages** or **discourages** you to take the target action. Considering your background, is the target action likely? Answer with only the letter (A or B). **Options:** (A) Likely (B) Unlikely
Describe states after actions	**Context:** You are a resident in a wildfire evacuation scenario. Based on your history provided above, analyze the current situation. **Current Situation:** `{observation}` **Your Decision:** `{action}` **Task:** Describe the IMMEDIATE consequences of this action from your first-person perspective. Focus on sensory details and safety status change. **Constraints:** Keep it brief (1-3 sentences); Be objective; Avoid speculation.
Belief Relations	**Context:** You are a resident in a wildfire evacuation scenario. Analyze the relationship between two beliefs held by yourself. **Input:** Belief 1: `{belief 1}` Belief 2: `{belief 2}` **Task:** Output only one number: a value between 0 and 1 indicates they weaken each other; a value greater than 1 indicates they enhance each other. Do not explain.
Infer belief given action and observation	**Context:** You are a resident in a wildfire evacuation scenario. Based on your history provided above, analyze the current situation. **Hypothetical Scenario:** **1.** Resident's Current Observation: `{observation}` **2.** The resident has made this action: `{action}` **Hypothesis to Evaluate:** Target Belief: `{belief}` **Task:** Evaluate if the resident likely holds the "Target Belief" given the observation and the action they made. Answer with only the letter (A or B). **Options:** (A) Likely (B) Unlikely