



OPEN ACCESS

EDITED BY

Mikkel Lepperød,
Simula Research Laboratory, Norway

REVIEWED BY

Jungsun Yoo,
University of California, Irvine, United States
Miguel A. F. Sanjuan,
Rey Juan Carlos University, Spain
Md Al Masrur Khan,
Indiana University, United States

*CORRESPONDENCE

Daria de Tinguy
✉ Daria.detinguyatugent.be

RECEIVED 18 September 2024

ACCEPTED 19 November 2024

PUBLISHED 11 December 2024

CITATION

de Tinguy D, Verbelen T and Dhoedt B (2024)
Learning dynamic cognitive map with
autonomous navigation.
Front. Comput. Neurosci. 18:1498160.
doi: 10.3389/fncom.2024.1498160

COPYRIGHT

© 2024 de Tinguy, Verbelen and Dhoedt. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Learning dynamic cognitive map with autonomous navigation

Daria de Tinguy^{1*}, Tim Verbelen² and Bart Dhoedt¹

¹Department of Engineering and Architecture, Ghent University/IMEC, Ghent, Belgium, ²Verses, Los Angeles, CA, United States

Inspired by animal navigation strategies, we introduce a novel computational model to navigate and map a space rooted in biologically inspired principles. Animals exhibit extraordinary navigation prowess, harnessing memory, imagination, and strategic decision-making to traverse complex and aliased environments adeptly. Our model aims to replicate these capabilities by incorporating a dynamically expanding cognitive map over predicted poses within an active inference framework, enhancing our agent's generative model plasticity to novelty and environmental changes. Through structure learning and active inference navigation, our model demonstrates efficient exploration and exploitation, dynamically expanding its model capacity in response to anticipated novel un-visited locations and updating the map given new evidence contradicting previous beliefs. Comparative analyses in mini-grid environments with the clone-structured cognitive graph model (CSCG), which shares similar objectives, highlight our model's ability to rapidly learn environmental structures within a single episode, with minimal navigation overlap. Our model achieves this without prior knowledge of observation and world dimensions, underscoring its robustness and efficacy in navigating intricate environments.

KEYWORDS

autonomous navigation, active inference, cognitive map, structure learning, dynamic mapping, knowledge learning

1 Introduction

Humans effortlessly discern their position in space, plan their next move, and rapidly grasp the layout of their surroundings (Tyukin et al., 2021; Rosenberg et al., 2021) when faced with ambiguous sensory input (Zhao, 2018). Replicating these abilities in autonomous artificial agents is a significant challenge, requiring robust sensory systems, efficient memory management, and sophisticated decision-making algorithms. Unlike humans, artificial agents lack inherent cognitive abilities and adaptive learning mechanisms, particularly when confronted with aliased observations, where sensory inputs are ambiguous or misleading (Lajoie et al., 2018).

To replicate human navigational abilities, an agent must capture the dynamic spatial layout of the environment, localize itself, and predict the consequences of its actions. Most attempts to achieve this combine those fundamental elements in SLAM algorithms (Simultaneous Localization and Mapping), often based on Euclidian maps (Campos et al., 2020; Placed et al., 2023). However, these methods require substantial memory as the world expands. Other strategies involve deep learning models, which depend on large datasets and struggle to adapt to unexpected events not encountered during training (Levine and Shah, 2022). A more efficient alternative lies in cognitive graphs or maps and learning a mental representation of the world from partial observations (Ha and Schmidhuber, 2018; Peer et al., 2021), creating a symbolic structure of the environment (Epstein et al., 2017; Friston et al., 2021). Cognitive graphs, by definition,

represent a “mental understanding of an environment” derived from contextual cues such as spatial relationships (American-Psychological-Association, 2024). Alongside this structure, the ability to imagine the outcomes of actions enables more reliable navigation decisions based on preferences (Kaplan and Friston, 2017; Gershman, 2017).

Our approach integrates those biological capabilities into a unified model. Using visual observations and proprioception (Bush et al., 2015), we construct a cognitive map through a generative model, enabling navigation with an active inference (AIF) framework. This model links states by incorporating observations and positions through transitions, as illustrated in Figure 1B, showing the processed observation of the agent and Figure 1C presenting the resulting cognitive graph.

Using Bayesian inference, the model predicts future states and positions, growing its cognitive map by forming prior beliefs about un-visited locations. As new observations are made, the agent updates its internal model, dynamically refining its representation of the environment (Friston et al., 2021). This continual adjustment allows the agent to effectively navigate complex environments by anticipating and learning from uncharted areas (Whittington et al., 2022). Our internal positioning system draws inspiration from the neural positioning system found in rodents and primates, aiding in self-localization and providing an intrinsic metric for measuring distance and relative direction between locations (Bush et al., 2015; Zhao, 2018; Edvardsen et al., 2019).

To achieve goal-directed navigation and exploration, we employ AIF to model the agent’s intrinsic behavior in a biologically plausible way. Unlike methods relying on pre-training for specific environments, our approach introduces a navigation and dynamic cognitive map growing based on the free energy (FE) principle. This map is inspired by mechanisms observed in animals, such as border cells for obstacle detection (Solstad et al., 2008) and the visual cortex for visual perception. It continuously expands by predicting new observations and adapting dynamically to changing environmental structures.

This study aims to develop an autonomous agent that can determine where it is, decide where to navigate, and learn the structure of complex, unknown environments without prior training, mimicking the adaptability and spatial awareness observed in biological organisms. Traditional exploration approaches and deep learning models struggle in dynamic settings, requiring extensive memory and pre-collected datasets to predict future settings, or they face difficulties in adapting to untrained situations. The challenge is to design a model that allows agents to autonomously build, update, and expand an internal map based on current sensory data and past beliefs, efficiently managing ambiguous observations (such as aliased states) and responding flexibly to unexpected environmental changes.

Our contribution to this problem encompasses several key aspects:

- Proposing a novel dynamic cognitive mapping approach that allows agents to predict and extend their internal map over imagined trajectories, enabling anticipatory navigation and rapid adaptation to new environments.

- Developing a navigation model that operates without pre-training or prior exposure, allowing the agent to successfully explore and make decisions in unfamiliar environments.
- Proposing a flexible navigation behavior fully explicit by relying upon the AIF framework.
- Outperforming in environmental learning and decision-making efficiency the Clone-Structured Cognitive Graph (CSCG) model (George et al., 2021), a prominent model for cognitive map representation.
- Showcasing robust adaptability, where the model responds seamlessly to dynamic environmental changes, replicating rat maze-like scenarios, thus emphasizing its practical application in flexible, real-world navigation tasks.
- Incorporating biologically inspired processes such as border cells and visual cortex perception, our agent’s navigation strategy is theoretically grounded and scalable to more realistic settings.

2 Related work

As agents navigate their surroundings, they connect observations to construct an internal map or graph of the environment. This intuitive decision-making process is propelled by incentives such as food, safety, or exploration, rapidly guiding agents toward their objective (Rosenberg et al., 2021).

Motion planning

Navigation tasks are often categorized into two primary scenarios based on the agent’s familiarity with the environment and its objectives. In scenarios where the environment is partially known and the agent is already aware of its destination, the primary focus is achieving efficient retrieval and executing actions reliably. This entails leveraging existing knowledge about the environment’s structure and landmarks to navigate swiftly and accurately toward the predetermined goal (Levine and Shah, 2022). Typically, this task centers around solving the motion planning problem: “How to move from point A to B?” Methods such as those in Mohamed et al. (2022, 2023) propose enhanced versions of Model Predictive Path Integral (MPPI) navigation, which calculates the optimal sequence of control actions by simulating multiple future trajectories and selecting the one minimizing a cost function. They allow real-time adaptation to obstacles in dynamic environments by continuously updating their navigation based on the most recent state and cost estimates of the control trajectory. However, these methods often rely on precise localization within a local map, require substantial computational capacity, and may be vulnerable to sensor failures.

Goal identification

Our current model is not focussing specifically on solving motion planning but on determining where the agent is and where it should go based on available information. Unlike traditional models, it does not require precise or absolute knowledge of its position; instead, it relies on its internally inferred belief about its location and considers obstacles relative to this position rather than to a global map. As a result, sensor failures are less critical, provided the agent can adapt to situations with either no sensory input or multiple inputs. In this study, the agent is expected to receive visual

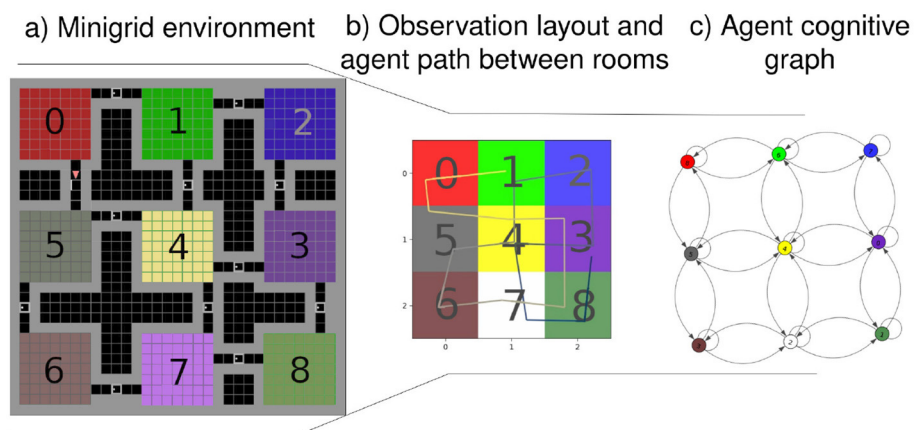


FIGURE 1

(A) From a full 3 by 3 rooms mini-grid environment (Chevalier-Boisvert et al., 2018; de Tinguy et al., 2024) to (B) rooms observation layout as perceived by the agent and the path it has taken between rooms - composed of a line from black to white-, (C) shows the agent final internal topological graph (cognitive graph) linking all the locations between them.

observations and detect obstacles through a system analogous to LiDAR. Failure of these sensors, however, would halt navigation.

When the agent must determine both its location and intended direction in addition to planning its movement, the task becomes significantly more complex, especially in unknown environments. The agent must engage in map building and employ a form of reasoning to effectively operate and navigate through unfamiliar surroundings (Chaplot et al., 2020a; Ritter et al., 2021). This involves dynamically constructing a cognitive map of the environment, integrating sensory information, and adapting behavior based on learned spatial relationships and environmental cues (Mirowski et al., 2016; Gupta et al., 2017). Works relying on Gaussian processes such as (Ali et al., 2023) unify navigation, mapping, and exploration, reducing redundancy and improving exploration efficiency; however, it is weak to featureless environments, over-reliant on a correct position estimation and are weak to sudden displacements. Our research addresses the challenges posed by determining where it should go and how it should do it by proposing an agent capable of seeking information gain and reaching desired observations in both familiar and unfamiliar environments. Our approach ensures robust performance even in scenarios involving potential disruptions, such as kidnapping, repeated observations, or environmental modifications, without requiring any environment-specific pre-training. Through this framework, we seek to advance the capabilities of fully autonomous agents in navigating and solving tasks in diverse and dynamic environments.

Spatial representation

Spatial representation plays a significant role in robot navigation and boasts a rich history, offering diverse approaches with their own set of advantages and challenges (Placed et al., 2023). While many SLAM systems rely on metric maps to navigate, which provide precise spatial information (Lajoie et al., 2018; Campos et al., 2020), there is a growing interest in topological mapping (Chaplot et al., 2020b; Ali et al., 2023) due to its biological plausibility and lower computational memory requirements. Cognitive maps are mental representations of spatial

knowledge that explain how agents navigate and apprehend their environment (George et al., 2021). A cognitive map encompasses the layout of physical spaces (Peer et al., 2021), landmarks, distances between locations, and the relationships among different elements within the environment (Foo et al., 2005; Epstein et al., 2017). Models such as the clone-structured cognitive graph (CSCG) (Guntupalli et al., 2023) and transformer representations (Dedieu et al., 2024) create cognitive maps (usually represented as topological graphs) using partial observations, offering reusable, and flexible representations of the environment. However, these models often entail significant training time, typically involving fixed policies or random motions and require a statically defined cognitive map dimension.

In biological systems, such as animals, the hippocampus plays a crucial role in managing episodic memory, spatial reasoning, and rapid learning (Stachenfeld et al., 2016), while structured knowledge about the environment is gradually acquired by the neocortex (Zhao, 2018; Raju et al., 2022). This enables remarkable adaptability and efficiency in navigation, with animals often requiring minimal instances to learn and navigate complex environments (Tyukin et al., 2021). They leverage cognitive mapping strategies to adapt to changes, swiftly grasp the layout of their surroundings, and efficiently return to previously visited places.

Drawing inspiration from these natural mechanisms, the compact cognitive map (Zeng and Si, 2019) proposes an extendable internal map based on movement information (how far the agent is from any past location) and current visual recognition. Our system shares this goal of remembering significant observations, forming a spatial representation, and resolving ambiguity through contextual cues (Tomov et al., 2018). Furthermore, we extend its adaptability by proactively forecasting the extension of the internal cognitive map before the experience of new spatial information. This anticipatory capability empowers the system to predict potential action outcomes in unobserved areas, enhancing its ability to navigate in unknown territory. This proactive approach aligns more closely with the efficient decision-making processes observed

in biological agents, allowing for rapid adaptation and effective navigation in diverse environments.

Active inference

Animals navigate their environments adeptly by combining sensory inputs with proprioception (Zhao, 2018). This enables them to avoid being misled by repeated evidence, a phenomenon known as aliasing. Such navigation strategies can be elucidated through active inference (Kaplan and Friston, 2017). Active inference applied to navigation entails the continual refinement of internal models based on sensory feedback, facilitating adaptive and effective decision-making within the environment. Serving as a normative framework, AIF elucidates cognitive processing and brain dynamics in biological organisms (Friston, 2013; Parr et al., 2022). It posits that both action and perception aim to minimize an agent's free energy, acting as an upper limit to surprise. Central to active inference are generative models, encapsulating causal relationships among observable outcomes, agent actions, and hidden environmental states. These environmental states remain "hidden" as they are shielded from the agent's internal states by a Markov blanket (Ha and Schmidhuber, 2018). Leveraging partial observations, the agent constructs its own beliefs regarding hidden states, enabling action selection and subsequent observation to refine its beliefs relying on the partially observable Markov decision model (POMDP) (Friston et al., 2016).

At the heart of decision-making and adaptive behavior lies a delicate balance between exploitation and exploration (Schwartenbeck et al., 2019). Exploitation involves selecting the most valuable option based on existing beliefs about the world, while exploration entails choosing options to learn and understand the environment (Friston et al., 2021). Recent behavioral studies indicate that humans engage in a combination of both random and goal-directed exploration strategies (Gershman, 2017). In our model, policies are chosen stochastically using the concept of FE. Regardless of the strategy employed, the agent attempts to minimize surprise by formulating policies that increase the likelihood of encountering preferred states, whether it is a specific observation (e.g., food) or the curiosity to comprehend the environment's structure (Schwartenbeck et al., 2019). This approach facilitates active learning by swiftly diminishing uncertainty regarding model parameters and enhancing knowledge acquisition about unknown contingencies. It also regulates model parameter updates in response to new evidence in a changing environment. Typically, this decision-making relies on priors over the current world structure and past observed outcomes to predict the next policies, usually leaning on static structures. Past research such as Bayesian model reduction techniques (Friston et al., 2019), or selection mechanisms (Friston et al., 2023), aims to expand models to accommodate emerging patterns of observation, relying on past and current observations. We propose to go one step further by extending our cognitive map over predicted outcomes, allowing the agent to reason over future un-visited states.

All those mechanisms lead an agent to efficiently explore or forage autonomously in any structured environment, even if it changes. It is learning as it goes in a few-shot or one-shot learning as would mice in a labyrinth (Rosenberg et al., 2021) relying on predictions and observations.

3 Method

The method section describes our approach to dynamic cognitive mapping and navigation in unfamiliar environments. Figure 2 provides a step-by-step overview of our model's process, visualized as a flowchart. This includes observing the environment, updating internal matrices with new information, inferring the current state, predicting possible motions, and selecting the next action based on updated beliefs.

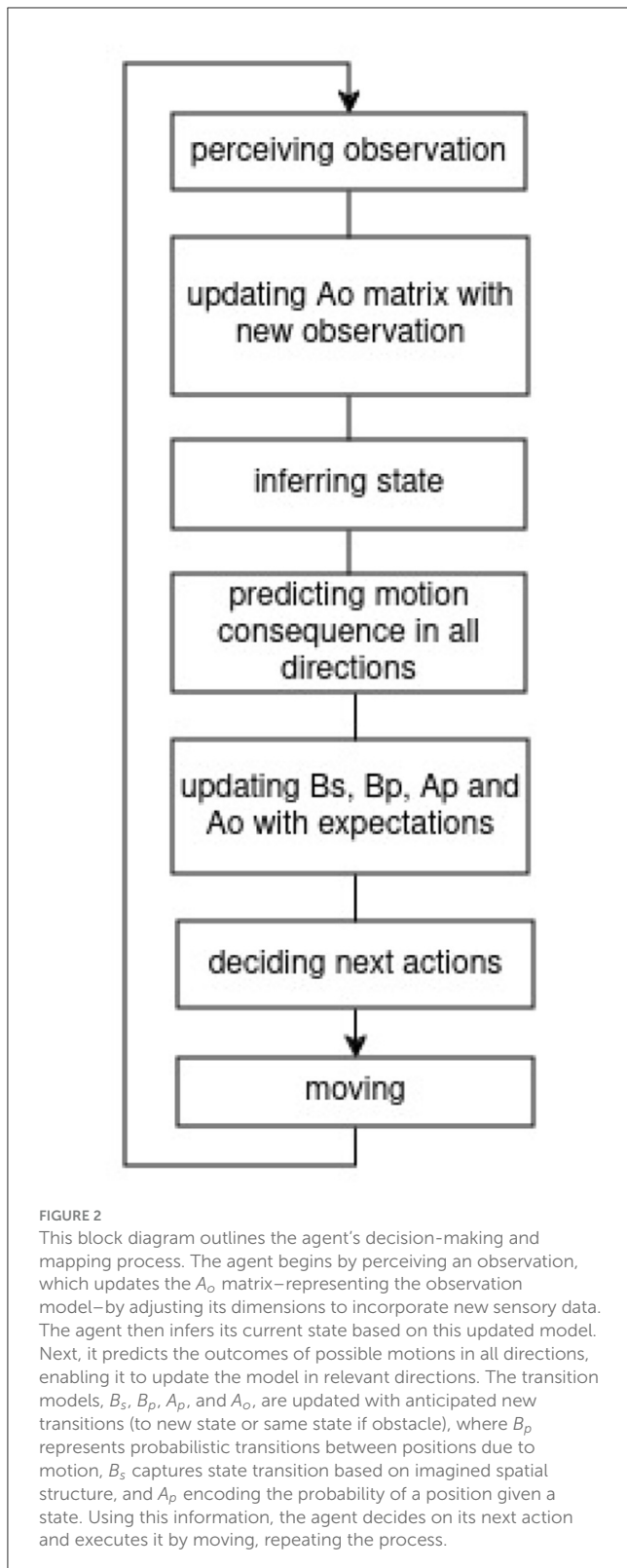
To demonstrate the agent's navigation abilities, we tested it in a series of mini-grid environments, as illustrated in Figure 1, where each map comprises rooms with a consistent floor color connected by randomly positioned corridors separated by closed doors, all the map layouts used in this study are presented in Appendix 1 and will be further detailed in Section 4.

Our agent starts exploring without prior knowledge of the environment's dimensions or potential observations. It starts by inferring the initial state and pose from the initial observation and expands its model based on anticipating the outcomes of moving in the four cardinal directions. It accounts for potential unexplored adjacent areas before formulating decision-making policies, based on the model's existing beliefs. In Figure 1A, the agent's traversal of an environment is depicted, with each room encoded as a state by the model and the visual observation being the color of the room's floor. Motion between rooms through a door suggests a transition between states, while walls are recognized as obstacles through RGB observations. The presented inference mechanism is expected to operate at the highest level of abstraction within a hierarchical framework such as de Tinguy et al. (2024), where lower layers handle motion, the observation process and detect the presence of obstacles, akin to the output expected from, respectively, our motor cortex, visual cortex, and border cell mechanism (Solstad et al., 2008).

Figure 1B illustrates our agent navigation through processed observations; in the case of our mini-grid environments, the lower levels of the hierarchical model summarize the room information to a single color (i.e., the floor color) and obstacle detection (i.e., walls or doors) fed to the highest level of abstraction (i.e., presented model). The presented model generates an internal topological map given motions between rooms such as the final map shown in Figure 1C. Each state contains information about the room's floor color and its inferred position.

3.1 State inference

Before each step, the agent engages in state inference by integrating the latest observation and internal positioning, following the POMDP graph depicted in Figure 3, where the current state s_t and position p_t are inferred based on the previous state s_{t-1} , position p_{t-1} , and action a_{t-1} leading to the current observation o_t . The generative model capturing this process is described by Equation 1, where the joint probability distribution over time sequences of states, observations, and actions is formulated. Tildes ($\tilde{\cdot}$) denote sequences over time.



The inference process operates within an AIF framework, where sensory inputs and prior beliefs are integrated to infer the agent's current state.

Due to the posterior distribution over a state becoming intractable in large state spaces, we use variational inference instead. This approach introduces an approximate posterior denoted as $Q(\tilde{s}, \tilde{p} | \tilde{o}, \tilde{a})$ and presented in Equation 2 (Smith et al., 2022).

$$Q(\tilde{s}, \tilde{p} | \tilde{o}, \tilde{a}) = Q(s_0, p_0 | o_0) \prod_{t=1}^{\tau} Q(s_t, p_t | s_{t-1}, p_{t-1}, a_{t-1}, o_t) \quad (2)$$

The inference scheme defined in Equation 2 heavily relies on priors and observations to localize the agent within its environment.

Therefore, we associate observations o_t , stemming from visual information, with our inferred position p_t , generated by the agent's proprioception. This improves state inference in an aliased environment, where the agent is simultaneously inferring its state and building its cognitive map. In the absence of prior information, the internal positioning p is initialized at the start of exploration as an origin [i.e., a tuple (0,0)]. It is then updated when the agent transitions between rooms (e.g., by passing through a door).

When the agent is certain about its current state, it updates past beliefs about observations and transitions in response to environmental changes. This ensures rapid re-alignment between the agent's model and the evolving environment, within one or a few iterations, depending on the agent's confidence in its outdated beliefs. The specific learning mechanism of our model will be explained in Section 3.4.

In scenarios where the agent is unexpectedly relocated to an unfamiliar position, as tested in some of our experiments, its certainty about its location decreases since its current state and previous actions no longer match its present observations. This drop in confidence occurs as the probability of being in a particular state—considering both position and observation—falls below a predefined certainty threshold (see Appendix 2.1 for exact values). When this threshold is crossed, the agent suspends reliance on its position information to infer state, focussing instead exclusively on observations to regain sufficient confidence over its state. The position and model updates are paused during this re-estimation of its location until the agent achieves adequate confidence in its state. This approach, guided by free energy minimization, prioritizes gathering information about its whereabouts, fostering informed decision-making in uncertain conditions and encouraging a resilient comprehension of its new state and position.

3.2 Free energy

Typically, agents are assumed to minimize their variational free energy denoted F , which can serve as a metric to quantify the discrepancy between the joint distribution P and the approximate posterior Q as presented in Equation 3 (Parr et al., 2022). This equation shows how adequate is the model to explain the past and current observations.

$$P(\tilde{o}, \tilde{s}, \tilde{p}, \tilde{a}) = P(o_0 | s_0) P(s_0) P(p_0) \prod_{t=1}^{\tau} P(o_t | s_t) P(s_t, p_t | s_{t-1}, p_{t-1}, a_{t-1}) \quad (1)$$

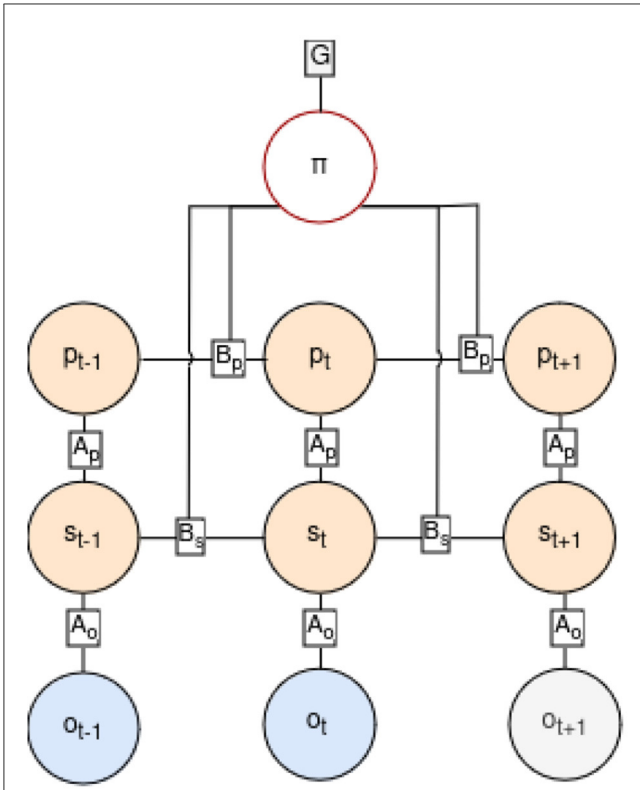


FIGURE 3
Factor graph of the POMDP in our generative model, showing transitions from the past to the present (up to time-step t) and extending into the future (time-step $t + 1$). Past observations are marked in blue, indicating they are known. In the future steps, actions follow a policy π influencing the new states and position in orange and new predictions in gray. The position at time t , p_t , is determined by the policy and the prior position p_{t-1} , while the current state s_t is inferred from the observation o_t , the position p_t , and the previous state s_{t-1} . Transitions between states are ruled by the B matrices, which define how prior conditions contribute to the current one considering taken actions. A matrices represent conditional probabilities of the quantities they connect.

$$\begin{aligned}
 F &= \mathbb{E}_{Q(\tilde{s}, \tilde{p} | \tilde{a}, \tilde{o})} [\log[Q(\tilde{s}, \tilde{p} | \tilde{a}, \tilde{o})] - \log[P(\tilde{s}, \tilde{p}, \tilde{a}, \tilde{o})]] \\
 &= \underbrace{D_{KL}[Q(\tilde{s}, \tilde{p} | \tilde{a}, \tilde{o}) || P(\tilde{s}, \tilde{p} | \tilde{a}, \tilde{o})]}_{\text{posterior approximation}} - \underbrace{\log[P(\tilde{o})]}_{\text{log evidence}} \\
 &= \underbrace{D_{KL}[Q(\tilde{s}, \tilde{p} | \tilde{a}, \tilde{o}) || P(\tilde{s}, \tilde{p} | \tilde{a})]}_{\text{complexity}} - \underbrace{\mathbb{E}_{Q(\tilde{s}, \tilde{p} | \tilde{a}, \tilde{o})} [\log[P(\tilde{o} | \tilde{s})]]}_{\text{accuracy}}
 \end{aligned} \tag{3}$$

Active inference agents aim to minimize their free energy by engaging in three main processes: learning, perceiving, and planning. Learning involves optimizing the model parameters, perceiving entails estimating the most likely state, and planning involves selecting the policy or action sequence that leads to the lowest expected free energy. Essentially, this means that the process implicates forming beliefs about hidden states offering a precise and concise explanation of observed outcomes while minimizing complexity (Friston, 2013).

While planning, we minimize the expected free energy (EFE) instead, denoted G , indicating the agent’s anticipated free energy

after implementing a policy π . Unlike the variational free energy F , which focuses on current and past observations, the expected free energy incorporates future expected observations generated by the selected policy.

To calculate this expected free energy $G(\pi)$ over each step τ of a policy, we sum the expected free energy of each time-step.

$$G(\pi) = \sum_{\tau} G(\pi, \tau) \tag{4}$$

$$\begin{aligned}
 G(\pi, \tau) &= \underbrace{\mathbb{E}_{Q(o_{\tau}, s_{\tau}, p_{\tau} | \pi)} [\log(Q(s_{\tau}, p_{\tau} | \pi)) - \log(Q(s_{\tau}, p_{\tau} | o_{\tau}, \pi))]}_{\text{information gain term}} \\
 &\quad - \underbrace{\mathbb{E}_{Q(o_{\tau}, s_{\tau}, p_{\tau} | \pi)} [\log(P(o_{\tau}))]}_{\text{utility term}}
 \end{aligned} \tag{5}$$

The expected information gain quantifies the anticipated shift in the agent’s belief over the state from the prior $Q(s_{\tau} | \pi)$ to the posterior $Q(s_{\tau} | o_{\tau}, \pi)$ when pursuing a particular policy. On the other hand, the utility term assesses the expected log probability of observing the preferred outcome under the chosen policy. This value intuitively measures the likelihood that the policy will guide the agent toward its prior preferences. Prior preferences are embedded within the agent’s model as an objective or target state the agent should work toward, usually given as a preferred state or observation. Free energy indirectly encourages outcomes that align with its preferences or target states. This approach makes the utility term less about “reward” in the traditional sense of reinforcement learning and more about achieving coherence with the agent’s built-in preferences, balancing this with exploration (information gain). The agent therefore chooses the optimal policy to follow among possible policies by applying

$$P(\pi) = \sigma(-\gamma G(\pi)) \tag{6}$$

where σ , the softmax function, is tempered with a temperature parameter γ converting the expected free energy of policies into a categorical distribution over policies. Playing with the temperature parameter alters the stochasticity of the navigation, the agent being more or less likely to choose the optimal policy rather than any other one. Details about the definition of policies can be found in Appendix 2.2.

3.3 Active inference and cognitive map navigation

When navigating within an environment, the agent uses AIF to continuously refine its knowledge of the world by updating its model parameters, more specifically, transition probabilities (how likely it is to move from one state to another) and likelihoods (how likely it is to observe certain features given a state). These quantities are updated, considering the actions taken by the agent and the observations gathered resulting from them. These updates help the agent to reduce the gap between its predicted and actual observations, effectively fine-tuning its internal model to match the real environment better. This approach allows the agent to

anticipate future states better and select actions that will minimize future EFE, thus optimizing its navigation strategy. Typically, this process assumes that the agent has some prior knowledge about the environment, such as its dimensions or the types of observations it might encounter (Neacsu et al., 2022; Kaplan and Friston, 2017). This prior knowledge allows the agent to form expectations over observations or states, even though it might not know which observations correspond to which locations. For instance, the agent might expect to encounter walls, doors, or specific floor colors, but it does not know where exactly these will be.

Identifying a model architecture with the right level of complexity to capture accurately the environment is problematic. Therefore, we take the approach of letting the model expand dynamically and explore as needed. Model expansion, however, requires a trigger indicating the need for an increase in model complexity.

Some have expanded their model upon receiving new patterns of observations (de Tinguy et al., 2024; Friston et al., 2023). When considering the case of exploring room structured mazes, it implies the creation of new states only when the agent physically transitions to a new location/room.

In an environment where multiple actions could have potentially created new states, only generating states upon actual observation means the agent will forget that previously visited rooms could have led to other unexplored spaces. Consequently, the generative model cannot leverage this knowledge to predict accurately the presence or absence of adjacent rooms from past visited rooms since these potential states have not been included in the model.

Therefore, when the agent evaluates policies using EFE, it fails to accurately predict action consequences if it has not directly experienced those transitions. This short-sighted exploration strategy neglects the potential connections between visited and un-visited locations, leading to sub-optimal exploration.

To illustrate this, consider our agent expanding its internal map only after directly observing the new rooms. Initially, the agent observes its current room and knows that doors may lead to other rooms, but it has no prior knowledge of what lies behind these doors. This uncertainty results in a high predicted information gain in the EFE. When the agent finally crosses a door and encounters a new observation, it generates a new state corresponding to this observation. However, the doors it did not explore are forgotten as only really observed outcomes are considered to update the model. Thus, the agent will not consider these un-visited areas as offering significantly greater potential for information gain in its future planning. This leads to a slower exploration process, where the agent takes longer to visit all rooms because it fails to anticipate the existence of un-visited distant locations and ignores how to reach them. The model overlooks exploration opportunities that could help it select the most effective exploration strategy overall.

To address this, our model learns to expand its internal map based on all available opportunities and grows based on predicted observations or states rather than solely on actual observations. This way, the agent retains awareness of all potential actions, including those it has not yet executed, allowing it to systematically explore the environment by considering both visited and un-visited locations. This strategy enhances the connectivity between different areas of the environment (enabling the agent to predict

that different doors are likely to lead to the same room) and enables the agent to explore more efficiently, reducing the time required to map its surroundings fully.

Figure 4 presents the three approaches:

- when we know the environment dimension (static cognitive map dimension),
- when the internal map grows upon receiving new observations,
- when the internal map grows upon predicting new states yet to be explored.

the first plan o) presents an agent's motion in a room maze, the arrow in orange illustrates the path the agent follows from the green to the purple room. a) Requires a prior over the environment, thus not being adapted to fully unknown mazes. b) Is not considering the blue room in its model even though it knew it existed when in the red room (S_1), but since it chose to move toward the purple room, this knowledge is lost. c) Considers all the motions it could have taken to change rooms. Thus, it remembers that the blue room exists and that it has never been there, improving drastically exploration as it can plan to move there from anywhere in the maze.

3.4 Cognitive map extension

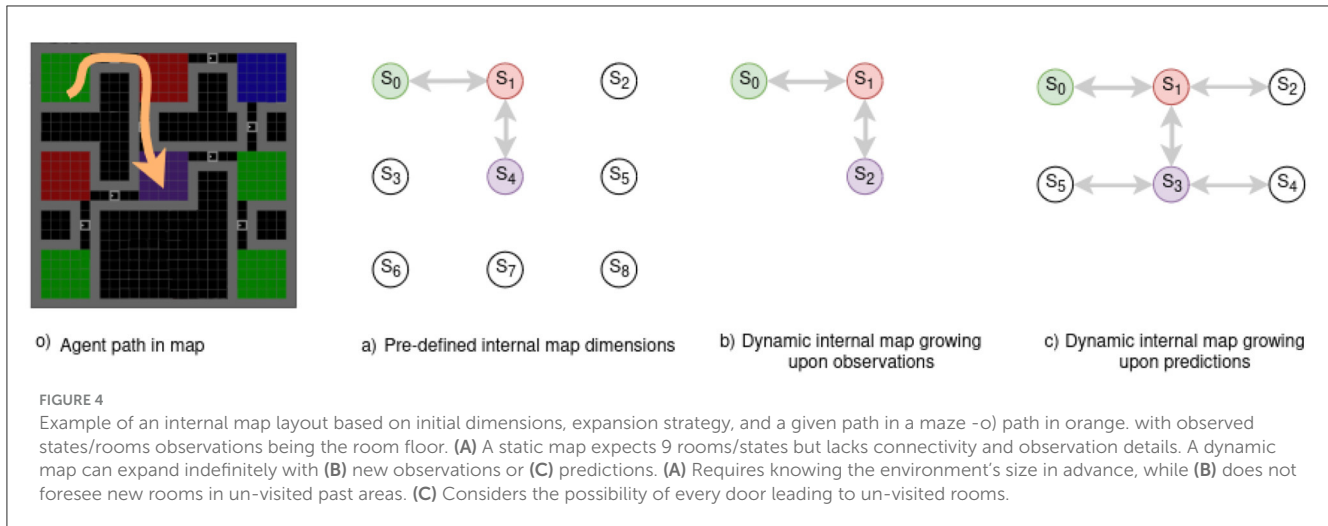
In this section, we present how the model learns the structure of the environment by expanding the cognitive map based on predictions, before obtaining actual observations, and updating beliefs upon given new evidence.

Each time the agent moves, the model infers its current state and pose by integrating its motion and observations, subsequently updating the model parameters. In line with the Bayesian reduction model (Friston et al., 2019), if the observation likelihood A_o fails to encompass the current observation, this indicates the novelty of the observation. As a result, the dimensionality of the observation likelihood is expanded to incorporate the new observation, and the model parameters θ are updated accordingly to reflect this extended observation likelihood.

The optimisation of beliefs regarding the generative model parameters θ involves minimizing the free energy F_θ , while accounting for prior beliefs and uncertainties related to both parameters and policies, as outlined in Parr et al. (2022):

$$\begin{aligned} \theta &= (A_o, A_p, B_s) \\ F_\theta &= \mathbb{E}_{Q(\pi, \theta)} [F(\pi, \theta)] + D_{KL}[Q(\theta) || P(\theta)] + D_{KL}[Q(\pi) || P(\pi)] \end{aligned} \quad (7)$$

This optimisation in Equation 7 balances the expected accuracy of the model's predictions and the necessity of maintaining coherence with prior beliefs. This ensures that the agent's learned representations and actions are both accurate and consistent with existing knowledge. The parameter set θ includes the following Markov matrices: the state transition $B_s = P(s_t | s_{t-1}, a_{t-1})$, the observation likelihood $A_o = P(o_t | s_t)$, and the position likelihood $A_p = P(p_t | s_t)$. The position transition matrix $B_p = P(p_t | p_{t-1}, a_{t-1})$ is not updated through active inference; instead, it is a deterministic metric in which the size grows as the agent explores, guided by the policy π . This indicates that the transition between two positions is determined by the given action and the previous position.



In contrast, Equation 7 indicates that the state transition model B_s is updated based on the transitions the agent experiences during exploration. Further details regarding the learning rates and matrix initialization can be found in Appendix 2.1.

Once the current model is fully updated, policy outcomes are predicted using expected free energy (EFE) in the four cardinal directions from the agent's current state. Walls and doors are identified from RGB observations and incorporated to determine the feasible directions for room transitions, influencing the transition probabilities. Equation 8 presents the EFE of a policy π with collision observation c . The probability $P(c)$ is binary, meaning there either is or is not a collision, with probabilities of 1 or 0, respectively. The learning term of the equations shows how much we learn about the position likelihood considering an obstacle or not while the inference term evaluates the state s_{t+1} and position p_{t+1} given the collision observation c_{t+1} . It reflects how well the model infers the state and position based on that observation.

$$\begin{aligned}
 G(\pi) &= \mathbb{E}_{Q_\pi} [\log Q(s_{t+1}, p_{t+1}, A_p | \pi) - \log Q(s_{t+1}, p_{t+1}, A_p | c_{t+1}, \pi) - \log P(c_{t+1})] \\
 &= - \underbrace{\mathbb{E}_{Q_\pi} [\log Q(A_p | s_{t+1}, p_{t+1}, c_{t+1}, \pi) - \log Q(A_p | s_{t+1}, p_{t+1}, \pi)]}_{\text{expected information gain (learning)}} \\
 &\quad - \underbrace{\mathbb{E}_{Q_\pi} [\log Q(s_{t+1}, p_{t+1} | c_{t+1}, \pi) - \log Q(s_{t+1}, p_{t+1} | \pi)]}_{\text{expected information gain (inference)}} \\
 &\quad - \underbrace{\mathbb{E}_{Q_\pi} [\log P(c_{t+1})]}_{\text{expected risk of collision}}
 \end{aligned} \tag{8}$$

The transition model B_p tracks possible transitions between positions. It determines the next position by incrementing the previous one based on the agent's motion. When calculating transition probabilities between positions, detected obstacles are taken into account. If an obstacle is detected in a given direction, the transition probability to that position is set to zero, and the predicted new position is disregarded. Conversely, if no obstacles are detected and the agent predicts a new position p_{t+1} that does

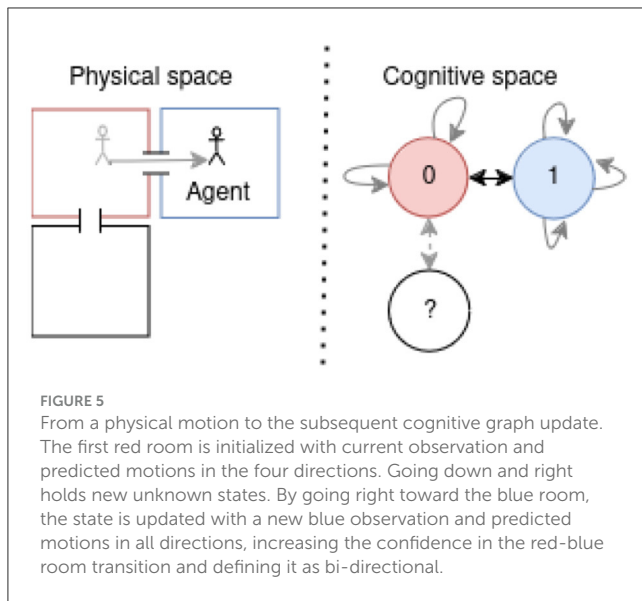
not match any previously known positions (i.e., the position has not been imagined before), the size of B_p is expanded to include this newly discovered position. This approach enables the model to adapt and remember new positions as they are encountered. Considering any pose p , Equation 9 represents the expected free energy (EFE) of the prior over the position likelihood parameters, summed across all policies. The sigmoid function, σ , transforms the negative EFE into a probability value (Friston et al., 2023). This equation evaluates how effectively the current position likelihood A_p explains the relationship between the state s , the pose p , and the collision observation c .

$$\begin{aligned}
 P(A_p) &= \sigma(-G) \\
 G(A_p) &= \mathbb{E}_{Q_{A_p}} [\log P(p, s | A_p) - \log P(p, s | c, A_p) - \log P(c)] \\
 &= - \underbrace{\mathbb{E}_{Q_{A_p}} [\log P(p, s | c, A_p) - \log P(s | p, A_p) - \log P(p | A_p)]}_{\text{expected information gain}} \\
 &\quad - \underbrace{\mathbb{E}_{Q_{A_p}} [\log P(c)]}_{\text{expected value}}
 \end{aligned} \tag{9}$$

If the predicted next position p_{t+1} does not correspond to any existing state in the model (considering the state probability conditioned on the position), the model recognizes this position as corresponding to a new, unexplored state. Thus, the model introduces a new state in its representation to account for this new place, expanding all probability matrices to accommodate the additional state dimension. This ensures that the state dimensions remain consistent across all matrices. The newly introduced state is assigned a high probability in the position likelihood matrix A_p , such that $P(s_{t+1} | p_{t+1}) = 1$, ensuring the model accurately reflects this new position in its state representation.

The observation likelihood matrix A_o assigns uniform probabilities across its distribution for un-visited states, reflecting the agent's limited knowledge about the unexplored states. Essentially, the agent cannot predict which observation to expect at a newly predicted location due to the lack of prior experience there.

The model first evaluates the likelihood of reaching the predicted position given a specific motion to calculate the transition



probabilities between a current and predicted state based on position. If an obstacle is detected in the direction of this expected position, the transition probability is null, meaning the action results in the agent staying in the same state. Conversely, if no obstacle is detected, the transition likelihood between the current and predicted states is maximized. This state transition is determined by the position transition and the corresponding probability $P(s_t|p_t)$, ensuring that the position-based transition is correctly mapped to the state transition within the model. In both scenarios, the Dirichlet distribution of the transition matrix B_s is updated similarly to how it is with experienced transitions but with a lower learning rate. This lower learning rate accounts for the greater uncertainty associated with predicted transitions compared to those already experienced. Details about the learning rates can be found in [Appendix 2.1](#).

These unexplored states are highly attractive in the EFE framework because visiting them offers significant information gain.

The schematic depicted in [Figure 5](#) illustrates a straightforward process of updating the state from the physical world to the cognitive graph. Unknown states are weakly defined as existing but lacking any observations, resulting in weaker transition probabilities compared to discovered states through experimental transitions holding observations.

4 Results

Our experiments are designed to assess the efficacy of our model in constructing cognitive graphs by linking visited locations with anticipated ones, thereby enabling efficient exploration and goal achievement.

We conducted comparative analyses with CSCG to assess several functionalities. These include learning spatial maps under aliased observations and decision-making strategies aimed at internal objectives, such as exploration or reaching specific observations. Those tests were done with and without prior

environmental information (i.e., is the model familiar with the environment layout). Finally, we validated our agent's self-localization capability following a kidnapping and re-localization at a random place and how it re-plans after observing obstacles on its path.

The clone-structured cognitive graph (CSCG) is a graph-based cognitive architecture where nodes represent beliefs and edges are transitions between these states. Clones of these nodes allow the agent to maintain multiple hypotheses about its environment, enabling parallel exploration and evaluation of different action sequences and their outcomes. This structure supports dynamic and adaptive decision-making in aliased environments ([George et al., 2021](#)). More about the CSCG can be found in [Appendix 2.3](#).

The environments comprise interconnected rooms, ranging from configurations with 7 to 23 rooms. These arrangements vary from T-shaped layouts to 3 by 3, 4 by 4 room grids, donuts-shaped layouts and a reconstruction resembling a Tolman's maze, all with or without aliased room colors. Environments containing several rooms with the same floor color are considered aliased and named such in our experiments.

We consider both models (CSCG and ours) as part of a hierarchical framework ([de Tinguy et al., 2024](#)). Information is processed at the lower levels, generating for each room:

- a single color per room as observation
- spatial boundary information (i.e., expected risk of collision) indicating obstacles or possible ways out (doors) ([Solstad et al., 2008](#)).

The agents can move in the four cardinal directions or remain stationary in the current room. Moving toward a door leads to entering a new room, while moving toward a wall does not alter the observation. Detailed observation layouts of the 8 environments used can be found in [Appendix 1](#), with an example of the 3x3 rooms layout and observations illustrated in [Figure 1](#).

4.1 Dynamic extension

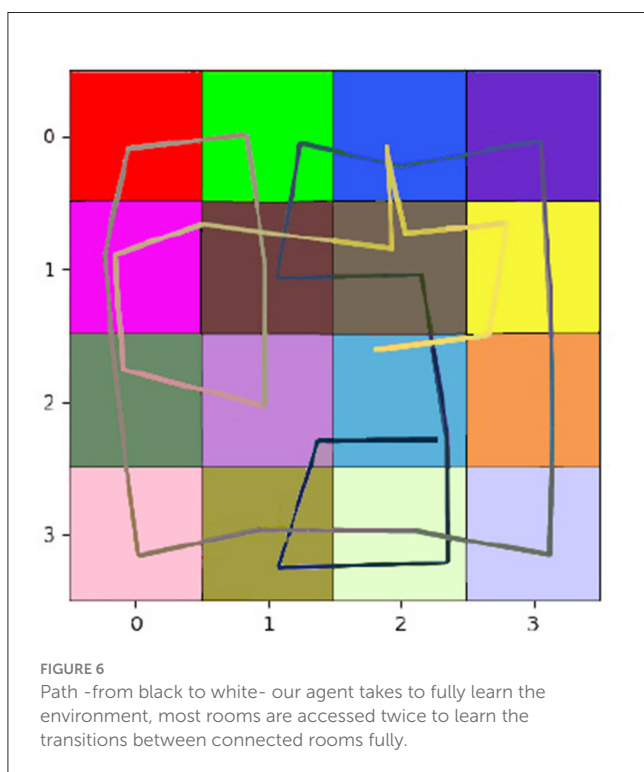
By moving through space, gathering observations, and updating their internal beliefs, both our model and CSCG exhibit the ability to learn the underlying spatial map of the environment, akin to the navigational capabilities observed in animals ([Peer et al., 2021](#)). However, while a CSCG has a static cognitive map and learns through random exploration, our model starts with a smaller state dimension and makes informed decisions at each step based on its internal beliefs and preferences. This mechanism enhances the relevance of each movement and accelerates the learning process, mirroring the efficient navigation strategies seen in animals ([Gershman, 2017](#)).

Both models were made to learn the connectivity between rooms. The agents spawned at random start locations and were tasked with seeking information gain from the environment - in other words, they were expected to learn the layout of the environment. Our model is led by information gain only as no preferred observations are given to the model, and the utility term is silent during pure exploration. The results averaged over

TABLE 1 Average number of steps required to fully learn the environments for each model.

Models\ environments	Oracle	Ours	CSCG	CSCG pose ob	CSCG random policy	CSCG pose ob random policy
3x3	11	14.4 ± 2	231.9 ± 59	221.3 ± 35	214.6 ± 44	205.7 ± 22
3x3_alias	11	18.1 ± 2	376.3 ± 35	201.5 ± 20	338.8 ± 30	210.4 ± 22
4x4	15	34.0 ± 4	375.7 ± 36	405.5 ± 78	419.6 ± 98	380.5 ± 53
4x4_alias	15	48.9 ± 9	509.5 ± 107	387.5 ± 67	554.1 ± 70	391.3 ± 66
T_maze	9	12.5 ± 1	284.5 ± 69	297.7 ± 108	261.7 ± 96	232.5 ± 73
T_maze_alias	9.5	14.3 ± 4	592.5 ± 28	282.7 ± 124	602.0 ± 69	341.0 ± 75
Donuts	13	15.6 ± 2	519.6 ± 127	450.0 ± 87	416.3 ± 105	493.1 ± 132

The best performances for each environment is presented in bold.



20 tests per model and environment are summarized in Table 1 and Figure 7. Our model demonstrates efficient exploration in all those environments. Exploration is considered complete when the internal belief regarding transitions between places aligns with the ground-truth transition matrix of the environment, with a minimum confidence of 60% over all correct transitions. This level of confidence was deemed enough to guarantee a good understanding of the maze structure for all models. To ensure a fair comparison, we provided CSCG with either color observations alone (referred to as “CSCG” in our results) as in our proposed model or color-position pairs observations (termed “CSCG pose ob” in our results). We considered providing ground-truth position and color as an observation during exploration without prior as equivalent to our model’s proprioceptive ability. In addition, the CSCG can navigate randomly as in the original

study (George et al., 2021) or by considering a localization belief similar to our model, coupled with the CSCG Viterbi navigation module as was realized in de Maele et al. (2023). Additional information about the CSCG architecture and training process can be found in Appendix 2.3. Two major differences highlight the capabilities of our model compared to the CSCG model: Our model can expand its internal map dynamically, adapting to the environment without requiring a predefined dimension size, and it can reason over unexplored areas, whereas the CSCG model needs to encounter a new place to integrate that state into its internal map.

Table 1 shows the impact of this difference in design. We can compare the number of steps required by each agent to explore diverse environments, with the oracle, an A-star algorithm, defining the minimum number of steps needed to visit all rooms once. In the T-shaped maze, the oracle’s number of steps is averaged considering their starting poses. Our model demonstrates significantly faster learning of all map structures than the CSCG algorithm. All CSCG models require a substantial number of steps to explore the environment, regardless of whether the observations are ambiguous or not. Neither random exploration nor Viterbi-based navigation significantly improves its performance. The decision-making process does not prioritize policies guiding the agent toward unexplored areas, limiting its overall effectiveness. Figure 6 illustrates a typical path undertaken by our agent within a 4×4 observation environment. Despite a few instances of overlapping paths, our agent demonstrates a discerning approach, making informed decisions rather than resorting to random exploration. This strategic decision-making results in an average maximum number of steps of our model being about 20 times smaller than those of the CSCG. To delve deeper into the exploration process, we examine the first discovery of all rooms, as depicted in Figure 7 on a logarithmic scale. Remarkably, our model’s performance closely resembles the oracles, suggesting a tendency to prioritize novelty exploration over confirming connections between rooms. Emphasizing imagined beliefs leads to faster convergence of locations connectivity; however, it also carries the risk of forming false connections not substantiated by observation. To mitigate this risk, we opted to maintain state connections made out of imagined beliefs weaker than those formed through observation.

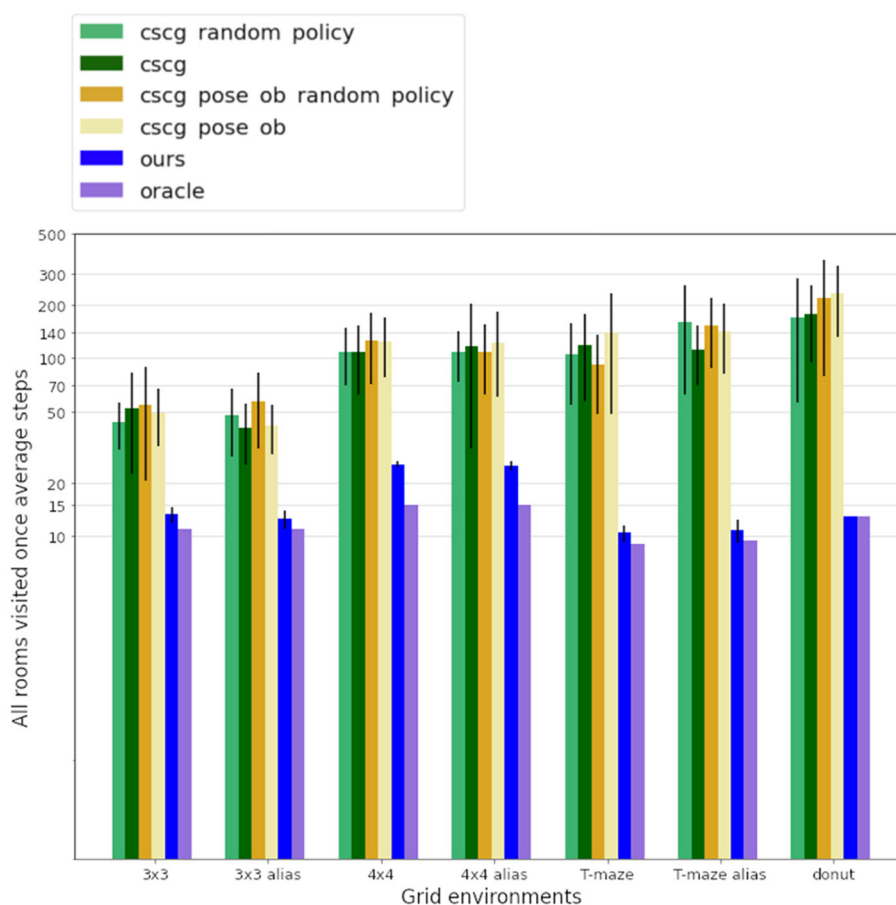


FIGURE 7

Figure displays the average number of steps required to discover all the rooms on a logarithmic scale, with the oracle serving as the benchmark for the minimum steps needed to visit all rooms once. Aliased rooms, marked by identical observations across different locations, present a challenge by potentially misleading the agent about its current position. Nonetheless, our agent has a meaningful pathway, discovering all rooms at least 30% faster than CSCG.

4.2 Autonomous localization

While exploring, our agent prioritized information gain, in the following experiments we give a preferred observation (a floor color) to stimulate exploitative behavior instead. The agent can navigate guided by a preference regardless of whether the environment is familiar or not with the desire to reach that objective. A goal is considered reached if the agent decides to stay at the desired location, thus discarding random stumbling upon it as a successful attempt. Figure 8 shows the average number of steps each model requires to reach the goal, considering both aliased and non-aliased environments. The results are averaged over environment type (aliased maze or not); each model had 20 runs per goal distance and environment. Without prior knowledge about the environment, our model systematically explores the map until it encounters the goal, resulting in a higher number of steps than the oracle, which has precise knowledge of the goal location. However, when we relocate our agent after exploration, keeping the map in memory (i.e., ours wt prior), the results align closely with the oracle in non-aliased environments (Figure 8B) and approach oracle performance in aliased environments (Figure 8A), where the

agent may need a few steps to localize itself while searching for the goal. A demonstration of our localization process is presented in Figure 9. CSCG, on the other hand, exclusively operates with prior knowledge of the environment and employs a Viterbi algorithm for navigation (Kanungo, 1999). Despite having prior information similar to our agent, CSCG's performance is not consistently superior when provided only with color observations. As our agent, CSCG also requires self-localization based on multiple observations in such scenarios. Overall, our agent demonstrates excellent efficiency in reaching and recognizing the goal. Even without prior over the map, the agent does not return to already explored areas unless necessary, demonstrating a biologically plausible path planning.

The impact of aliased observations over our model is demonstrated in Figure 9. The left panel of each figure presents the agent (represented by an X) provided with a prior distribution over the map (agent's state value reported at the upper left of each tile) and the policies' expected free energy with a darker gray color signifying higher preference to proceed toward that path. The right panel depicts the agent's confidence in its location (i.e., state) given collected observations. The agent starts at the bottom

of the T-maze (Figure 9A), observing only the present color, which is observable in three different locations (here in states 0, 1, and 7). The agent cannot tell from just this single observation where it is yet. This results in divided confidence between those three states when inferring its potential localization. In the right panel, we can see that the agent considers going forward as the best option. Notably, the imagined T turns are incorrect due to the agent's confidence in its localization being mainly split between states 0 and 1, while the agent is, in reality, at state 7. Moving to Figure 9B, representing step 1, the agent adjusts its internal beliefs regarding localization based on the color of the previous room and the current observation. This correction in beliefs leads to a refinement in the agent's perceived localization, as can be seen in the associated bar plot. In Figure 9C, corresponding to step 2, the agent exhibits a significantly higher level of certainty regarding its whereabouts in location 1. It confidently determines that the goal is located either to the left or right but rules out the possibility of it being behind. Finally, Figure 9D portrays the fourth step, where the agent demonstrates full confidence in its localization and successfully identifies the objective with dark gray shading on the goals. This high level of confidence indicates the agent's strong belief in its internal representations. Therefore, the next steps will correctly lead the agent either right or left toward the preferred observation.

4.3 Dynamic remapping

Can the model correctly update its internal cognitive map given new evidence contradicting its prior? To verify that, we conduct two experiments: one in the Donut environment, where there are only two paths to reach a goal (a long and a shorter path), by going left or right, and a second one reproducing Tolman's second maze experiment with obstacles located at several positions.

In our Donut experiment, the agent has previously learned the environment from a randomly selected starting point without any obstacle. Then, the agent was kidnapped to a corner of the map with the newly implemented objective of observing a lightly pink color (highlighted by a red square in Figure 10). Our model is given a floor color as objective, and exploitation and exploration have equivalent weights in the navigation. Path planning steps should be followed following the sequence from (1) to (10) depicted in Figure 10. At first, the agent successfully plans its path to accommodate the new objective by moving through the upper path, as depicted in the first frame (1) of Figure 10. The dark gray shading indicates the high level of confidence in this path, the darker the shade, the higher the certitude; this imagined policy leads to a desired outcome. We disrupted the shortest path by adding an obstacle, as highlighted by a white square in Figure 10. This obstruction invalidated the agent's original route, leading to a notable decrease in the probability of reaching the goal along the intended path (as evident in frame 2) of Figure 10. Consequently, the agent initiates a remapping process promptly, updating its beliefs regarding graph connectivity. The subsequent frames, from frames (3) to (10) in Figure 10, depict the agent's increasing confidence as it navigates closer to the objective using the longest path.

In this scenario where a room along the imagined path was unexpectedly closed off, the agent demonstrates its capability to dynamically adapt its navigation strategy in response to changes in the environment, effectively leveraging new observations to revise its path and achieve its objective. Failure to imagine an alternative path would have resulted in the agent stubbornly trying to pass a closed door. This flexible connectivity exists despite using a generative model, which is known to have a hard time revising strong beliefs. This is due to two things: First, when experimenting with a blockage, our agent updates its internal model with negative parameter learning (see Appendix 2.1). Second, the inherent growth of our agent, as it grows in its state or observation dimension, the probabilities get more distributed, so transitions to past states that have not been experimented for a long time naturally weaken over time, reproducing the mechanism of animal-like synaptic plasticity (Eichenbaum, 2015).

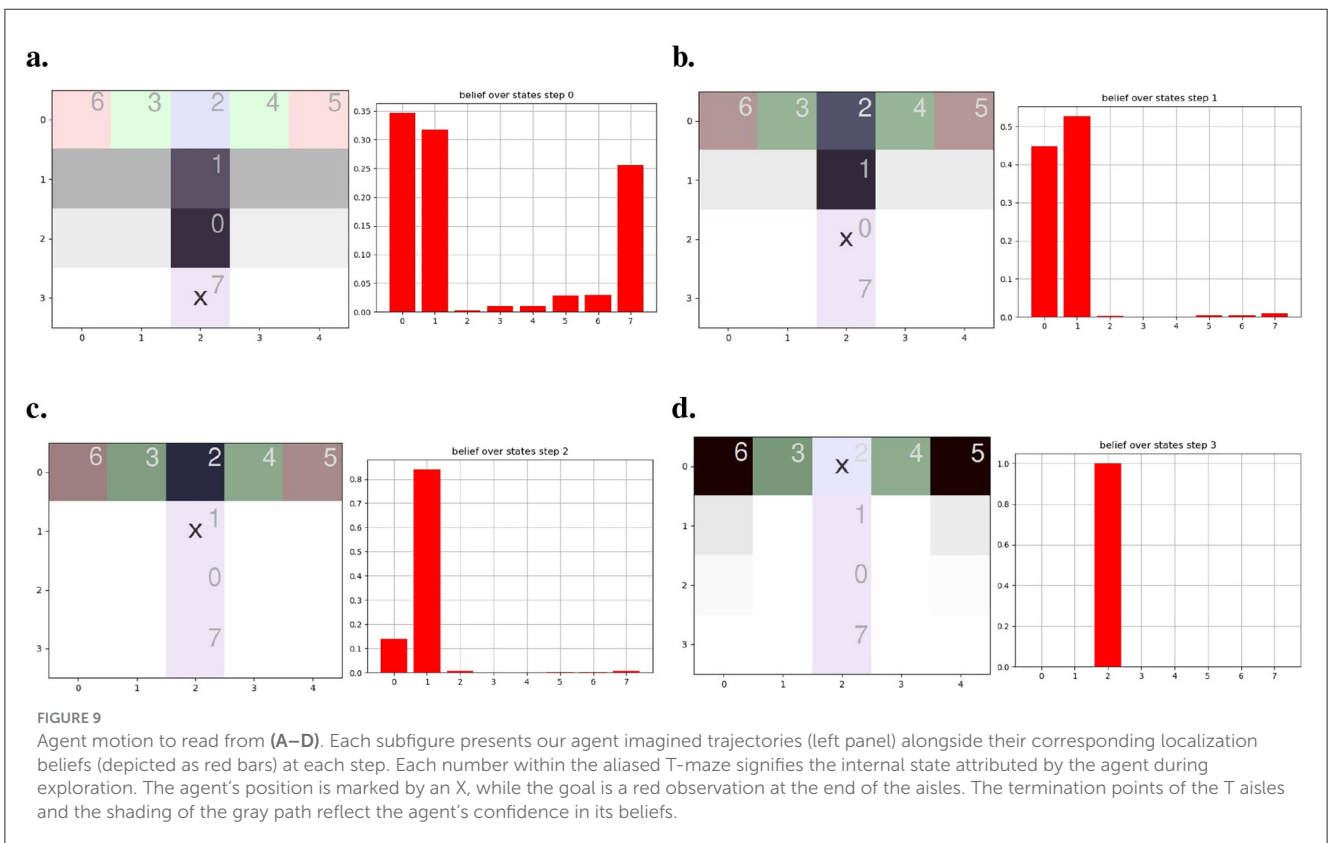
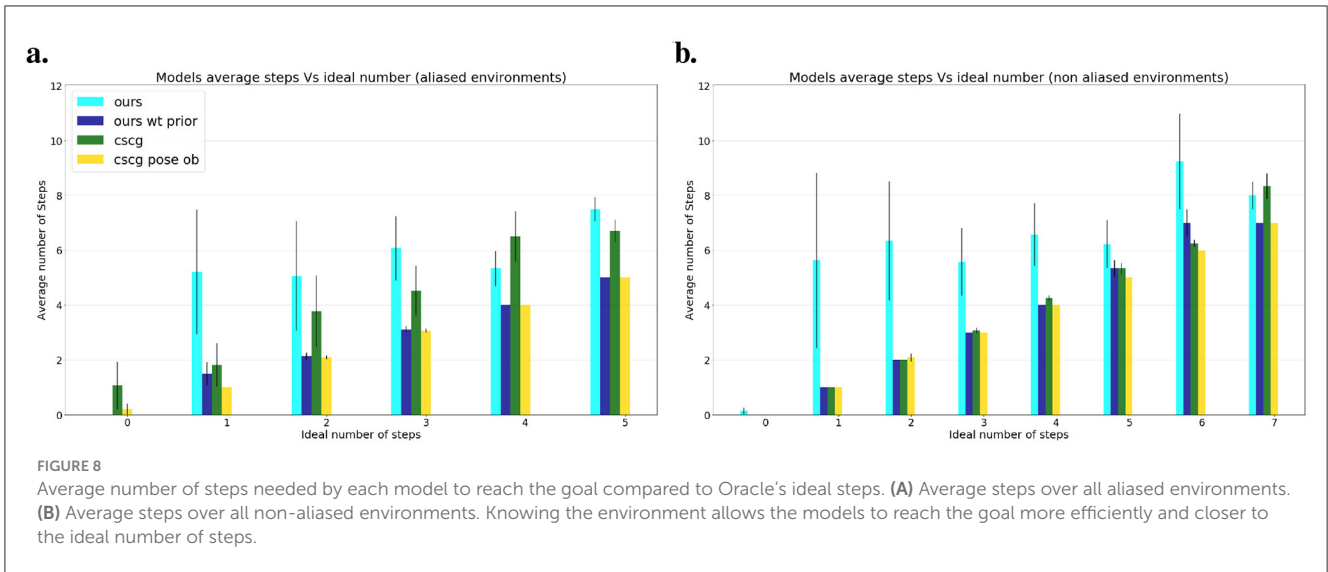
Our second test to assert the robustness of our model to dynamic environments entails replicating the Tolman maze experiment outlined in Tolman (1930). This maze configuration, illustrated in Figure 11A, involves placing starving rats at a starting point (marked by a red triangle) with food positioned at the opposite end of the map. Three routes are available to reach the food, with Route 1 being the shortest and Route 3 the longest. However, these routes can be blocked at points A and B, where the blockages affect the accessibility of Routes 1 and 2. In our case, the agent cannot observe obstacles from a distance; it has to be in contact with them from an adjacent room to learn about it. Each colored grid in Figure 11B represents an independent closed room, and the white box represents an obstacle disposed at positions A and B. Thus, we rely on the "insight" (Aru et al., 2023) of our agent and its ability to update its beliefs rather than its perceptual capabilities to reach the goal.

This "insight" of our agent is based on two key parameters: its planning ability, determining how far ahead it can imagine, and cognitive map plasticity, dictating how well it can adjust past beliefs to new evidence. We allowed the agent to predict action consequences up to 14 steps ahead; thus from any obstacle, the agent could imagine reaching the goal anyway.

The model is provided with a red visual observation as its preferred prior throughout the entire experiment with a weight on EFE utility term of 2 (which results in the agent favoring reaching the objective over exploring). The 10 agents consistently start at the bottom of the maze as depicted in Figure 11A red triangle. The agents followed three series of 12 runs in the Tolman maze with no obstacle (Figure 11B), then with an obstacle at position A (Figure 11B) and finally at position B (Figure 11B).

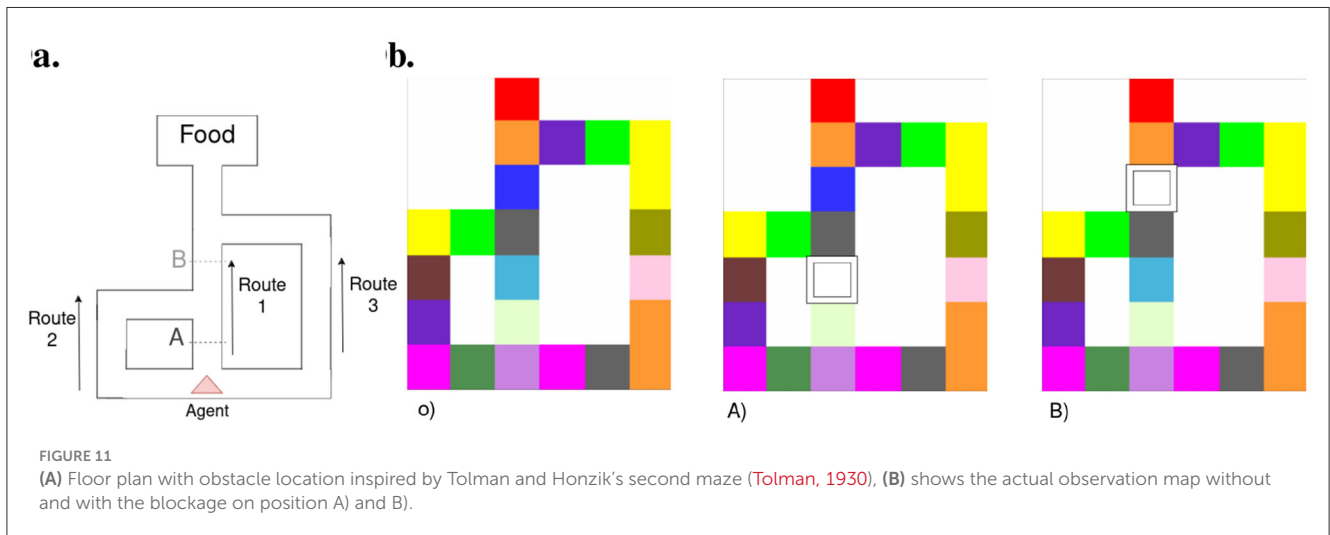
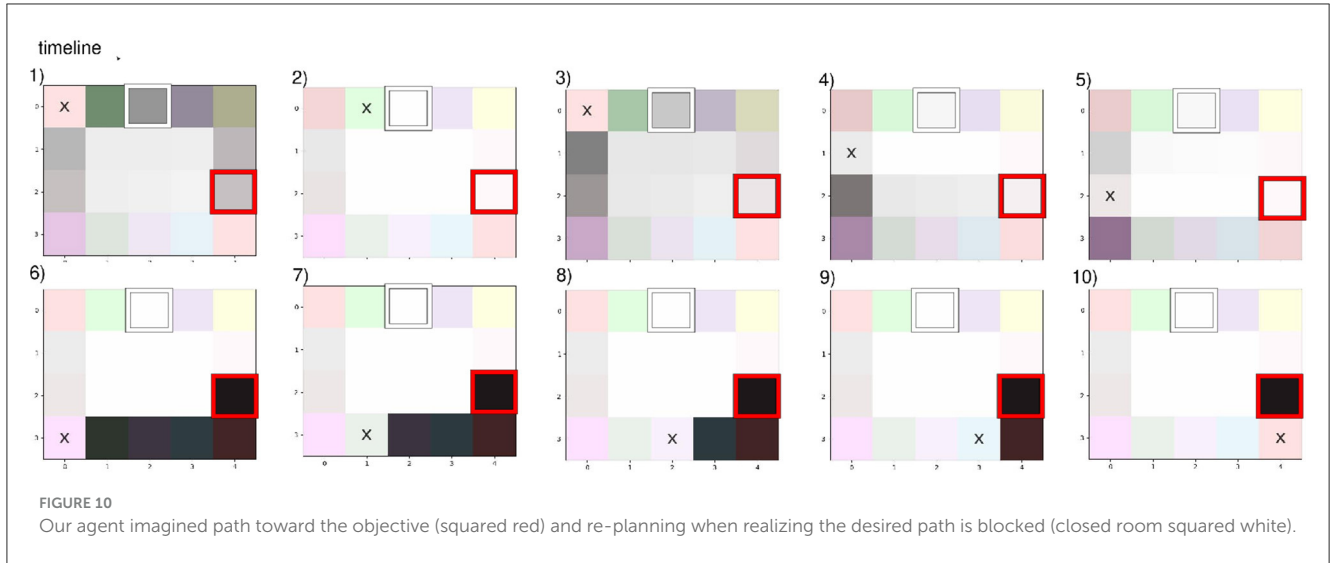
Initially, the agent has no prior knowledge of the environment and observations. Every 20 steps, the agent is "kidnapped" and repositioned at the starting point without being informed of this relocation. The agent must infer and correct its belief about its location based on subsequent observations. Figure 12 shows the path choice frequency over all agents given an obstacle compared to Martinet and Arleo (2010) results on 100 animats reproducing Tolman's insights.

During the first 12 runs, the maze contains no obstacles, and the fastest route to the goal is path 1. Our agent typically



explores all available paths at least once, following AIF framework as understanding the environment helps minimize free energy. Although the agent alternates between the three paths to balance between exploration and exploitative behavior, path 1 is generally preferred (almost 50% of the time) as we can see in the path count Table 2, recording all the attempts to reach the goal after experimenting with the blockage in path 1, if any. It closely aligns with what we would expect from a hungry rat as shown in Figure 12 first column; it resembles Tolman's expectations of rat behavior. Our agent tends to alternate between taking the

quickest path and taking another alternative path, with a clear regularity. Extending the number of runs would show path 1 being taken 50% of the time and paths 2 and 3 25% of the time each. The reason why it is not exactly 50% of the time is because the agent takes a few runs to explore the environment instead of always reaching the goal. This alternate path selection is linked to the balance chosen between exploitation (the preference of being at the goal) and exploration (the preference of learning the environment). When a path is considered better understood than the others, the agent counterbalances by re-exploring



those other paths. Let us remember that the agent can predict reaching the goal from those paths as well, even if it takes more steps.

An obstacle is introduced at position A for the next series of 12 runs. The agent continues to navigate using the same memory, which has accumulated experience from the previous runs in an obstacle-free environment. Path 2 becomes the quickest route. After encountering the obstacle a few times, the agent updates its internal map and shifts its preference to path 2, though it still periodically checks paths 1 and 3 and reaffirms that path 1 remains blocked. Trials consider all the tentative to reach the goal after experimenting with the blockage in path 1. Contrary to *Martinet and Arleo (2010)* results, our agents show a more divided comportment with path 2 being preferred but not completely neglecting path 3. Finally, the obstacle is moved to position B, and the experiment continues for 12 more runs. Initially, the agents often attempt to traverse path 2, but after further updating its model to account for the new blockage, they predominantly switch to path 3 while periodically checking path 2 again.

These results demonstrate our agent's robustness in handling kidnapping scenarios and its ability to update its internal cognitive map when new evidence contradicts previously well-verified beliefs. The model exhibits a clear and adaptable behavior, akin to rats navigating a maze. By adjusting the agent's prediction horizon, the weighting of the utility term in the expected free energy, and the stochasticity in path selection (testing values defined in *Appendix 2.1*), we can control the navigation behavior of our agent, favoring exploration or exploitation. These variations in behavior are easily interpretable through the AIF framework our model uses, avoiding the opacity often associated with black-box models.

5 Discussion

This study proposes a novel high-level abstraction model grounded in biologically inspired principles, aiming to replicate key aspects of animal navigation strategies (*Zhao, 2018; Balaguer et al., 2016*). Integrating a dynamic cognitive graph alongside

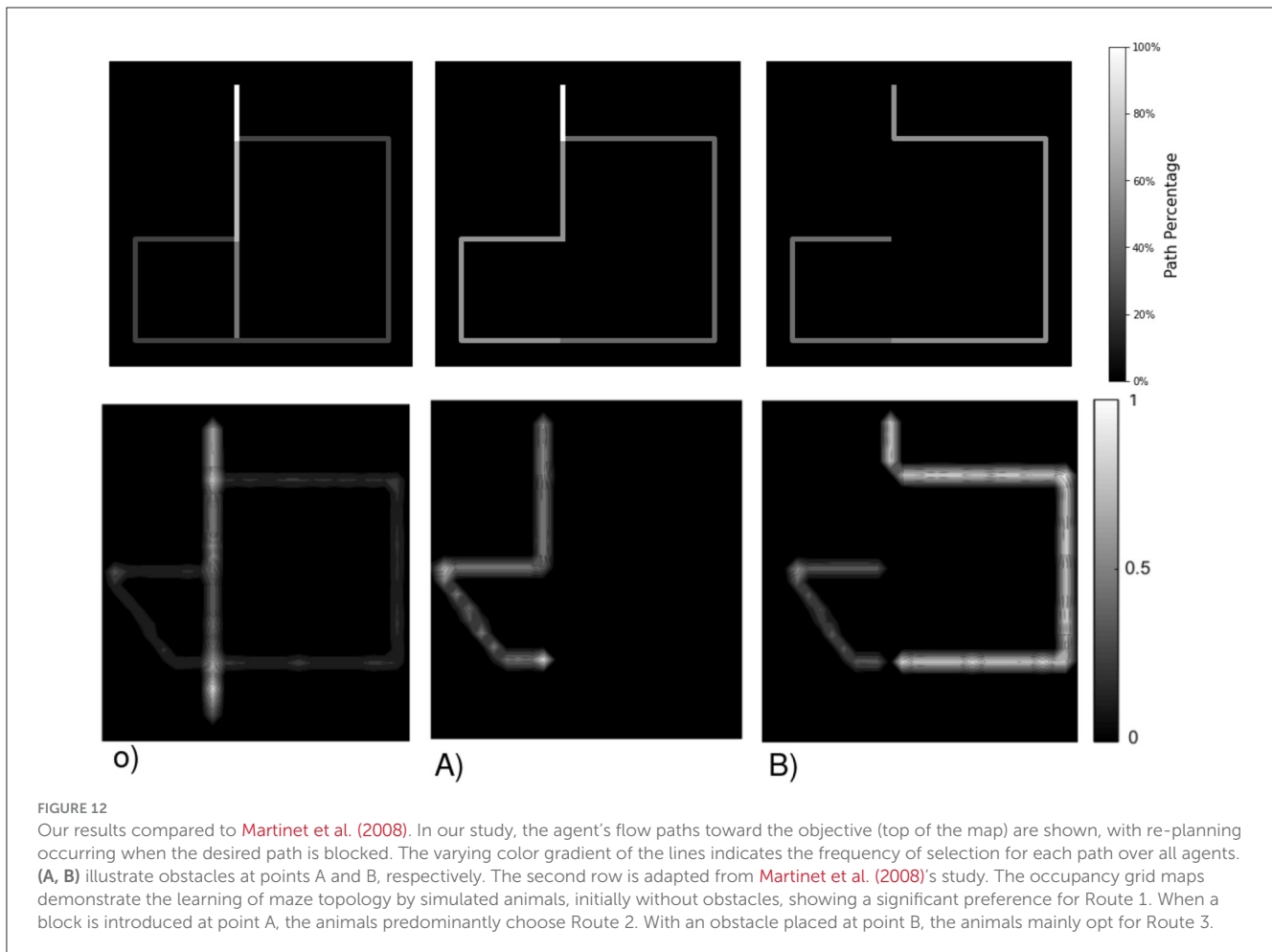


TABLE 2 Path count over all 10 agents when they reach the goal considering the obstacle position.

	Path 1	Path 2	Path 3
No obstacle	48	27	28
Obstacle A	0	70	49
Obstacle B	0	36	48

Trials, where the objective is not reached under 20 steps (exploration phases), are not counted.

internal positioning within an active inference framework is central to our approach. This novel combination enables our model to dynamically expand its cognitive map upon prediction while navigating any ambiguous mini-grid environment with or without prior, mirroring the adaptive learning and efficient exploration abilities observed in animals ([Tyukin et al., 2021](#); [Rosenberg et al., 2021](#)). Comparative experiments with the Clone-Structured Graph (CSCG) model ([George et al., 2021](#)) underscore the superiority of our approach in learning environment structures with minimal data and without preparatory knowledge of specific observation and state-space dimensions.

Moreover, we demonstrated the model's ability to adapt its cognitive map to new evidence and reach preferred observations in various situations. The presented model can represent large or complex environments such as warehouses or houses, without significant memory or storage demands for the cognitive map itself due to its matrix-based structure. However, as the agent attempts to make long-term predictions—projecting further steps into the future—computational requirements for processing power and memory increase. In this study, we implement up to 13-step predictions without issue due to the efficiency of our policy design; achieving predictions beyond 10 steps is notable for this type of generative model. Looking ahead, it could be interesting to investigate the impact of a perfect memory on future policies and exploration efficiency in addition to measuring the impact of belief certitude on the graph dynamic adaptation. Furthermore, splitting the data by unsupervised clustering ([Asano et al., 2020](#)), or by using the model's prediction error to chunk the observations into separate locations ([Verbelen et al., 2022](#)) would offer a closer approximation to animal behavior and allow open space exploration. This could extend this model capacity into real-world scenarios, such as StreetLearn ([Mirowski et al., 2019](#)) based on Google map observations or simulated realistic environments such as Habitat ([Savva et al., 2019](#)). The real world presents additional

challenges, such as segmenting large open spaces into manageable information chunks as realized by Hwang et al. (2024), efficiently processing visual data or other sensory inputs (with a method such as Kerbl et al., 2023), and distinguishing new observations from known ones based on memory. Moreover, navigating around local static and dynamic obstacles effectively is essential. Williams et al. (2017) propose a possible solution to move between locations while avoiding obstacles. Each of these challenges can be addressed individually and integrate the mentioned solutions to enhance our model's performance without altering the core navigation strategy. Finally, developing this high-level abstraction model with a hierarchical model (de Tinguy et al., 2024) would allow the model to reason over different levels of abstraction temporally and spatially extending its navigation to local in-room navigation and long-term planning.

Data availability statement

The code generated for this study can be found in the high-level nav planning: https://github.com/my-name-is-D/high_level_nav_planning/tree/main.

Author contributions

DT: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Conceptualization. TV: Writing – review & editing, Visualization, Validation, Supervision, Conceptualization. BD: Writing – review & editing, Supervision, Funding acquisition.

References

- Ali, M., Jardali, H., Roy, N., and Liu, L. (2023). "Autonomous navigation, mapping and exploration with Gaussian processes," in *Proceedings of the Robotics: Science and Systems (RSS)*, Daegu, Republic of Korea. doi: 10.15607/RSS.2023.XIX.104
- American-Psychological-Association (2024). *Cognitive map*. Available at: <https://dictionary.apa.org/cognitive-map> (accessed June 25, 2024).
- Aru, J., Drüke, M., Pikamäe, J., and Larkum, M. E. (2023). Mental navigation and the neural mechanisms of insight. *Trends Neurosci.* 46, 100–109. doi: 10.1016/j.tins.2022.11.002
- Asano, Y., Rupprecht, C., and Vedaldi, A. (2020). "Self-labelling via simultaneous clustering and representation learning," in *International Conference on Learning Representations*.
- Balaguer, J., Spiers, H., Hassabis, D., and Summerfield, C. (2016). Neural mechanisms of hierarchical planning in a virtual subway network. *Neuron* 90, 893–903. doi: 10.1016/j.neuron.2016.03.037
- Bush, D., Barry, C., Manson, D., and Burgess, N. (2015). Using grid cells for navigation. *Neuron* 87, 507–520. doi: 10.1016/j.neuron.2015.07.006
- Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., and Tardós, J. D. (2020). ORB-SLAM3: an accurate open-source library for visual, visual-inertial and multi-map SLAM. *CoRR, abs/2007.11898*. doi: 10.1109/TRO.2021.3075644
- Chaplot, D. S., Gandhi, D., Gupta, S., Gupta, A., and Salakhutdinov, R. (2020a). "Learning to explore using active neural slam," in *International Conference on Learning Representations (ICLR)*.
- Chaplot, D. S., Salakhutdinov, R., Gupta, A., and Gupta, S. (2020b). Neural topological SLAM for visual navigation. *CoRR, abs/2005.12256*.
- Chevalier-Boisvert, M., Willems, L., and Pal, S. (2018). *Minimalistic gridworld environment for openai gym*. Available at: <https://github.com/maximecb/gym-minigrid> (accessed October, 2022).
- de Maele, T. V., Dhoedt, B., Verbelen, T., and Pezzulo, G. (2023). "Integrating cognitive map learning and active inference for planning in ambiguous environments," in *International Workshop on Active Inference* (Cham: Springer Nature Switzerland), 204–217. doi: 10.1007/978-3-031-47958-8_13
- de Tinguy, D., Van de Maele, T., Verbelen, T., and Dhoedt, B. (2024). Spatial and temporal hierarchy for autonomous navigation using active inference in minigrid environment. *Entropy* 26:83. doi: 10.3390/e26010083
- Dedieu, A., Lehrach, W., Zhou, G., George, D., and Lzaro-Gredilla, M. (2024). Learning cognitive maps from transformer representations for efficient planning in partially observed environments. *arXiv preprint arXiv:2401.05946*.
- Edvardsen, V., Bicanski, A., and Burgess, N. (2019). Navigating with grid and place cells in cluttered environments. *Hippocampus* 30, 220–232. doi: 10.1002/hipo.23147
- Eichenbaum, H. (2015). The hippocampus as a cognitive map – of social space. *Neuron* 87, 9–11. doi: 10.1016/j.neuron.2015.06.013
- Epstein, R., Patai, E. Z., Julian, J., and Spiers, H. (2017). The cognitive map in humans: spatial navigation and beyond. *Nat. Neurosci.* 20, 1504–1513. doi: 10.1038/nn.4656
- Foo, P., Warren, W., Duchon, A., and Tarr, M. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of experimental psychology. Learn. Memory Cogn.* 31, 195–215. doi: 10.1037/0278-7393.31.2.195
- Friston, K. (2013). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Doherty, J. O., and Pezzulo, G. (2016). Active inference and learning. *Neurosci. Biobehav. Rev.* 68, 862–879. doi: 10.1016/j.neubiorev.2016.06.022

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research received funding from the Flemish Government under the "Onder-zoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fncom.2024.1498160/full#supplementary-material>

- Friston, K., Moran, R. J., Nagai, Y., Taniguchi, T., Gomi, H., and Tenenbaum, J. (2021). World model learning and inference. *Neural Netw.* 144, 573–590. doi: 10.1016/j.neunet.2021.09.011
- Friston, K., Parr, T., and Zeidman, P. (2019). Bayesian model reduction. *arXiv preprint arXiv:1805.07092*.
- Friston, K. J., Costa, L. D., Tschantz, A., Kiefer, A., Salvatori, T., Neacsu, V., et al. (2023). Supervised structure learning. *Biol. Psychol.* 193:108891. doi: 10.1016/j.biopsycho.2024.108891
- George, D., Rikhye, R., Gothoskar, N., Guntupalli, J. S., Dedieu, A., and Lzaro-Gredilla, M. (2021). Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nat. Commun.* 12:2392. doi: 10.1038/s41467-021-22559-5
- Gershman, S. (2017). Deconstructing the human algorithms for exploration. *Cognition* 173, 34–42. doi: 10.1016/j.cognition.2017.12.014
- Guntupalli, J. S., Raju, R., Kushagra, S., Wendelken, C., Sawyer, D., Deshpande, I., et al. (2023). Graph schemas as abstractions for transfer learning, inference, and planning. *arXiv preprint arXiv:2302.07350*.
- Gupta, S., Davidson, J., Levine, S., Sukthankar, R., and Malik, J. (2017). Cognitive mapping and planning for visual navigation. *CoRR, abs/1702.03920*.
- Ha, D., and Schmidhuber, J. (2018). World models. *CoRR, abs/1803.10122*.
- Hwang, J., Hong, Z.-W., Chen, E., Boopathy, A., Agrawal, P., and Fiete, I. (2024). Grid cell-inspired fragmentation and recall for efficient map building. *arXiv preprint arXiv:2307.05793*.
- Kanungo, T. (1999). “UMDHMM: hidden Markov model toolkit,” in *Extended Finite State Models of Language*, ed. A. Kornai (Cambridge, UK: Cambridge University Press).
- Kaplan, R., and Friston, K. (2017). Planning and navigation as active inference. *Biol. Cyber.* 112, 323–343. doi: 10.1101/230599
- Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42:139. doi: 10.1145/3592433
- Lajoie, P., Hu, S., Beltrame, G., and Carlone, L. (2018). Modeling perceptual aliasing in SLAM via discrete-continuous graphical models. *CoRR, abs/1810.11692*.
- Levine, S., and Shah, D. (2022). Learning robotic navigation from experience: principles, methods and recent results. *Philos. Trans. R. Soc.* 378:20210447. doi: 10.1098/rstb.2021.0447
- Martinet, L.-E., and Arleo, A. (2010). “A cortical column model for multiscale spatial planning,” in *Proceedings of the 11th International Conference on Simulation of Adaptive Behavior: From Animals to Animats, SAB'10* (Berlin, Heidelberg: Springer-Verlag), 347–358. doi: 10.1007/978-3-642-15193-4_33
- Martinet, L.-E., Passot, J.-B., Fouque, B., Meyer, J.-A., and Arleo, A. (2008). “Map-based spatial navigation: A cortical column model for action planning,” in *Spatial Cognition VI. Learning, Reasoning, and Talking about Space. Spatial Cognition 2008* (Berlin, Heidelberg: Springer-Verlag), 39–55. doi: 10.1007/978-3-540-87601-4_6
- Mirowski, P., Banki-Horvath, A., Anderson, K., Teplyashin, D., Hermann, K. M., Malinowski, M., et al. (2019). The streetlearn environment and dataset. *CoRR, abs/1903.01292*.
- Mirowski, P., Pascanu, R., Viola, F., Soyer, H., Ballard, A. J., Banino, A., et al. (2016). Learning to navigate in complex environments. *CoRR, abs/1611.03673*.
- Mohamed, I. S., Xu, J., Sukhatme, G. S., and Liu, L. (2023). Towards efficient mppi trajectory generation with unscented guidance: U-MPPi control strategy. *arXiv preprint arXiv:2306.12369*.
- Mohamed, I. S., Yin, K., and Liu, L. (2022). Autonomous navigation of AGVS in unknown cluttered environments: log-MPPi control strategy. *IEEE Robot. Autom. Lett.* 7, 10240–10247. doi: 10.1109/LRA.2022.3192772
- Neacsu, V., Mirza, M. B., Adams, R. A., and Friston, K. J. (2022). Structure learning enhances concept formation in synthetic active inference agents. *PLoS ONE* 17, 1–34. doi: 10.1371/journal.pone.0277199
- Parr, T., Pezzulo, G., and Friston, K. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. London: The MIT Press. doi: 10.7551/mitpress/12441.001.0001
- Peer, M., Brunec, I. K., Newcombe, N. S., and Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends Cogn. Sci.* 25, 37–54. doi: 10.1016/j.tics.2020.10.004
- Placed, J. A., Strader, J., Carrillo, H., Atanasov, N., Indelman, V., Carlone, L., et al. (2023). A survey on active simultaneous localization and mapping: state of the art and new frontiers. *IEEE Trans. Robot.* 39, 1686–1705. doi: 10.1109/TRO.2023.3248510
- Raju, R. V., Guntupalli, J. S., Zhou, G., Lzaro-Gredilla, M., and George, D. (2022). Space is a latent sequence: structured sequence learning as a unified theory of representation in the hippocampus. *arXiv preprint arXiv:2212.01508*.
- Ritter, S., Faulkner, R., Sartran, L., Santoro, A., Botvinick, M., and Raposo, D. (2021). Rapid task-solving in novel environments. *arXiv preprint arXiv:2006.03662*.
- Rosenberg, M., Zhang, T., Perona, P., and Meister, M. (2021). Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *Elife* 10:e66175. doi: 10.7554/eLife.66175
- Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., et al. (2019). “Habitat: a platform for embodied AI research,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9339–9347. doi: 10.1109/ICCV.2019.00943
- Schwartenbeck, P., Passerker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., and Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *Elife* 8:e41703. doi: 10.7554/eLife.41703
- Smith, R., Friston, K. J., and Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *J. Math. Psychol.* 107:102632. doi: 10.1016/j.jmp.2021.102632
- Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., and Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science* 322, 1865–1868. doi: 10.1126/science.1166466
- Stachenfeld, K. L., Botvinick, M. M., and Gershman, S. J. (2016). The hippocampus as a predictive map. *Nat. Neurosci.* 20, 1643–1653. doi: 10.1101/097170
- Tolman, E. C., and H. C. (1930). “insight” in rats. *Univ. Calif. Publ. Psychol.* 4, 215–232.
- Tomov, M. S., Yagati, S., Kumar, A., Yang, W., and Gershman, S. J. (2018). Discovery of hierarchical representations for efficient planning. *PLoS Comput. Biol.* 16:e1007594. doi: 10.1371/journal.pcbi.1007594
- Tyukin, I. Y., Gorban, A. N., Alkhudaydi, M. H., and Zhou, Q. (2021). Demystification of few-shot and one-shot learning. *CoRR, abs/2104.12174*.
- Verbelen, T., de Tinguy, D., Mazzaglia, P., Catal, O., and Safron, A. (2022). “Chunking space and time with information geometry,” in *Proceedings of the Thirty-Sixth Conference on Neural Information Processing Systems (NeurIPS 2022), Information-Theoretic Principles in Cognitive Systems Workshop*, 6.
- Whittington, J. C. R., McCaffary, D., Bakermans, J. J. W., and Behrens, T. E. J. (2022). How to build a cognitive map. *Nat. Neurosci.* 25, 1257–1272. doi: 10.1038/s41593-022-01153-y
- Williams, G., Drews, P., Goldfain, B., Rehg, J. M., and Theodorou, E. A. (2017). Information theoretic model predictive control: theory and applications to autonomous driving. *IEEE Trans. Robot.* 34, 1603–1622. doi: 10.1109/TRO.2018.2865891
- Zeng, T., and Si, B. (2019). A brain-inspired compact cognitive mapping system. *CoRR, abs/1910.03913*.
- Zhao, M. (2018). Human spatial representation: What we cannot learn from the studies of rodent navigation. *J. Neurophysiol.* 120, 2453–2465. doi: 10.1152/jn.00781.2017