Uniwersytet Mikołaja Kopernika w Toruniu
Wydział Filozofii i Nauk Społecznych
Katedra Kognitywistyki

Weronika Sójka
nr albumu: 303243

Praca licencjacka na kierunku kognitywistyka

Learning representation of the structured environments through a complex graphs inference

Opiekun pracy dyplomowej
dr Karolina Finc

Toruń, 2022

Abstract

Humans and animals succeed in searching through various complex environments. Back in the thirties, Tolman argued that such abilities arise from the construction of "cognitive maps" through exploration. Such maps enable active mapping of real-world environments to simplified cognitive representations. Although, this theory is widely accepted and studied, the way of how different cognitive maps are stored in the brain is still discussed. Some authors proposed that cognitive maps represent the environment as Euclidean space, but in my thesis I provide evidences that support the hypothesis that graph-theoretical approach provide a more robust framework that additionally allows for behavior flexibility. Here, I investigate a potential of reinforcement learning algorithms trained in navigation of different environments to construct an internal graph-based representation of these spaces. An ability to form a cognitive map in parallel with real-time exploration may underlie human ability to navigate in various spaces, represent the world structure and generalize to novel tasks.

Contents

1 Introduction

We learn from new experiences on a daily basis. Our cognition evolved to work with cognitively demanding problems in an efficient way. Some behavior does not require much conscious reasoning from us. We have developed sophisticated ways that help us to make quick inferences and decisions under novel environmental conditions. How do we represent complex tasks on abstract cognitive level and infer only important information for learning purposes?

The main interest of this thesis is focused on one particular type of mental representation observed for the first time during an active navigational task in rats (Blodgett, 1929; Tolman, 1948). Such a construct was defined as a cognitive map that enables the creation of a sufficient abstract representation of a real-world environment used to recall and decode important information about relative locations, their attributes, and relatedness. While facing a particular navigational dilemma, a person represents the underlying environment structure as a mental cognitive map and then performs various computations to identify a good spatial exploration plan. In an experimental setting, the goal is usually to get to a particular location to receive a reward, but in the real world, it can be any goal we can think of - from getting to a grocery shop to navigating in an airport. The whole process is happening on the abstract level of our cognition and it is oriented toward proper planning to find the best sequence of actions.

In the beginning, I will introduce the basis of the experimental approach developed to study the cognitive mapping phenomenon from the birth of the theory in 1948. Multiple brain evidence will be raised to support the theory with findings from neuroscientists studying the brain on the level of electrophysiological activity as well as structural and functional neuroimaging. The whole brain circuitry involved in navigation will be explained with an emphasis on current research that extends the cognitive map theory to a non-spatial domain that guides cognition towards generalization in various domains.

The next section is devoted to the computational approach that tries to build the internal model of the world using reinforcement learning, a branch of machine learning focused on optimization of the decision-making process (Sutton and Barto, 2018). As a potential base to test various strategies used by humans and machines to actively explore the world, its essence is described in detail before diving deeper into its applicability. I try to collate human and machine intelligence with each other to suggest a modeling approach in which an artificial agent builds an internal model of the environment through exploration. Further challenges in the sufficiency of various forms of such representation are discussed before the presentation of a possible solution in the following chapter.

The final section addresses the problem of the structure of cognitive map representation put in frames of graph-theoretical approach. Through navigation in complex environments, humans and animals create an internal graph-based representation of space that enables encoding the most important information about relatedness between multiple spatial locations. In parallel to an active exploration process in an unknown environment, a simplified representation of the world is learned by making inferences visualized in a form of a graph. This model is described in terms of cognitive mapping theory with an emphasis on neural correlates which encode topological information in the brain. Despite the focus on the neuroscientific basis of this approach, an artificial creation of a cognitive map in a form of a graph is also proposed. Finally, similarly to the cognitive map phenomenon itself, these assumptions are extended to a non-spatial domain, in particular to a social network structure.

2 Cognitive maps

Most social animals live in a homogeneous groups and possibly stable environments in which any changes cause initiation of defense mechanisms. It is caused chiefly, because of their inability to adapt to unknown environmental factors that could lead to security risks and death. Humans do not seem to have such difficulty, because the adaptation does not necessarily indicate any change in our behavior. One of the mechanisms supporting our orientation in new environments is the construction of simplified models of the world. Such a cognitive phenomenon was introduced as a *cognitive map* by Edward Tolman on a basis of his behavioral experiments on rats. He stated that: "learning consists, not in stimulus-response connections, but in the building up in the nervous system of sets which function like cognitive maps" (Tolman, 1948). Brain processes involved in rat exploration of the maze operate on the autonomous system which selects only relevant information that does not directly depend on receiving a reward. Such a system is meant to support memory processes, adaptation to novel situations, and guide behavior. However, the main reason for the existence of a cognitive map would be to allow for planning novel combination of routes that we have not traveled before.

Following the original idea, cognitive maps usually refer to the representation of space in the brain. Through the years scientists have studied this topic and constructed a more general understanding of cognitive mapping system. The original concept grew up to the idea supported by experimental results that humans and animals may form an internal representation of the spatial environment they were asked to navigate in. Nowadays, it has been proposed that a cognitive map may offer a mental tool for forming representations that enable encoding relevant information about various phenomena which does not have to be limited to the spatial domain (Behrens et al., 2018).

2.1 Tolman's experiments on rats

In 1948, Tolman wrote the most-cited article of his career in which he described the famous idea of cognitive maps. The concept came from his purely behavioral experiments conducted on rats, where he was studying how rodents learn to navigate through mazes. In the typical experimental setup, the rat was put at the entrance of the maze and then an animal could explore its structure to find food. Psychologists believed that such a navigational task involves strengthening or weakening simple stimulus-response associations of animal behavior. However, Tolman suggested that rats are doing more than that and they are able to learn comprehensive maps of their surroundings, which can guide goal-directed behavior, such as finding shortcuts to a reward (i.e., food).

Studies performed by Tolman were investigating multiple phenomena, out of which the one of them that postulated the creation of cognitive maps was the most significant. The most relevant experiments that Tolman was replicating, were started by (Blodgett,

1929) in which he was studying *latent learning*, the phenomenon of memorizing some information subconsciously without a directed exposure to behavioral reinforcement. Blodgett ran 3 groups of rats through a maze, as shown in Fig. 1. Results obtained from his experiments made it clear that rats' rate of learning was comprehensibly higher in groups that encountered delayed rewards (Fig. 2.2).

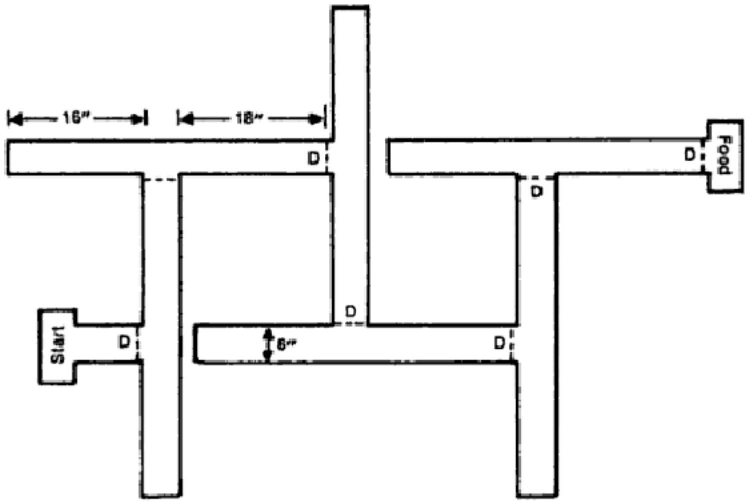

Figure 2.1: Experimental maze used for studying cognitive maps phenomenon
Maze in which rats were trained to find food. Adapted from Blodgett, 1929.

Based on these insights, Tolman engaged rats in maze exploration without any rewards. He hypothesized that rats gradually built up a map of the environment and could utilize it as soon as they were motivated to do so (in this case, because they were put in a maze hungry so they were motivated to look for food). He put hungry rats at the entrance of a maze and they were free to explore various paths to search for food. These conditions were repeated every 24 hours and within the next trials, rats tended to make fewer errors and spend less time between start and goal. Obtained results supported the idea that rodents were in fact able to learn the structure of a maze even in the absence of any reward. The main proof for this assumption was a sudden drop in the error score that appeared after previous exposure to unrewarded trials. It means that rats were subconsciously learning the cognitive map of a particular maze, and they used it to find the shortcut when the food reward appeared at a given location.

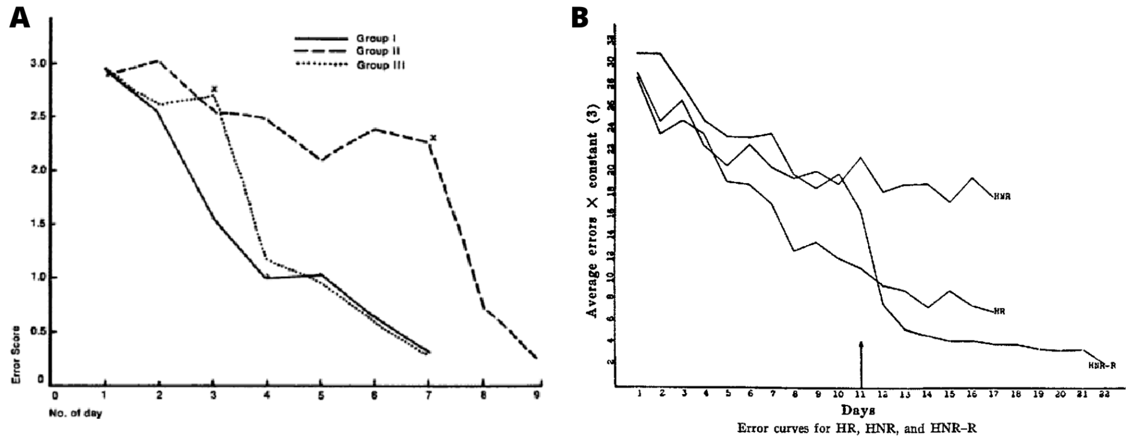

Figure 2.2: Performance in a maze by different rat groups
(A) Results obtained by Blodgett and his research group show the rate of learning in groups with different reward schema. Group I was a control group in which rats were running one trial a day and they found the food at the end of each. For the rest of the groups, rewards were introduced on different days (marked by a star on the graph). (B) Experiment repeated by Tolman and Honzik with a changed design. They used two control groups: the first one that never found food in the maze (HR), and the second one that found it throughout the experiment duration (HNR). The rats in the experimental group (HNR-R) started to find the food in a maze on the 11th day. Adapted from Tolman, 1948.

In another experiment run by Tolman, rats were put in the maze and as soon as they found food, a visual pattern appeared and it was followed by an electrical shock. Tolman noticed that one shock was enough for a rat to learn an association with a visual pattern. Even more interesting was an observation that, straight after an electrical shock, rats seemed to "look around". It was suggested that they might have looked for landmarks to find out what caused this unpleasant stimulus. Such behavior may be justified by a need to establish associations that an animal may learn to avoid in the future.

To summarize, Tolman and his fellow scientists were observing rats making flexible inferences in complex mazes. Results obtained from these experiments supported not only the existence of an internal model of the world in the rodents' brain but also how various sensory information impact its form.

2.2 Representation of space in the brain

After the birth of the concept of a cognitive map, spatial behaviors started to be studied by neuroscientists that tried to find accurate neural correlates of navigation through complex environments. Besides the observation of increased gray matter density along the long-axis of the hippocampus in a selected group of individuals that spend their daily lives navigating, e.g. taxi drivers (Maguire et al., 2000), most of the evidence for brain correlates of cognitive maps were found using techniques allowing direct recordings of neuronal activity in freely moving rodents. Representations of space were successfully found in brain areas reflected by the activity of specific neurons in the dorsal/posterior hippocampus and neighboring entorhinal cortex (O'Keefe and Nadel, 1978). Till then, this brain region was associated mainly with memory processes observable after dramatic memory impairment caused by the removal of the hippocampal formation (Scoville and Milner, 1957).

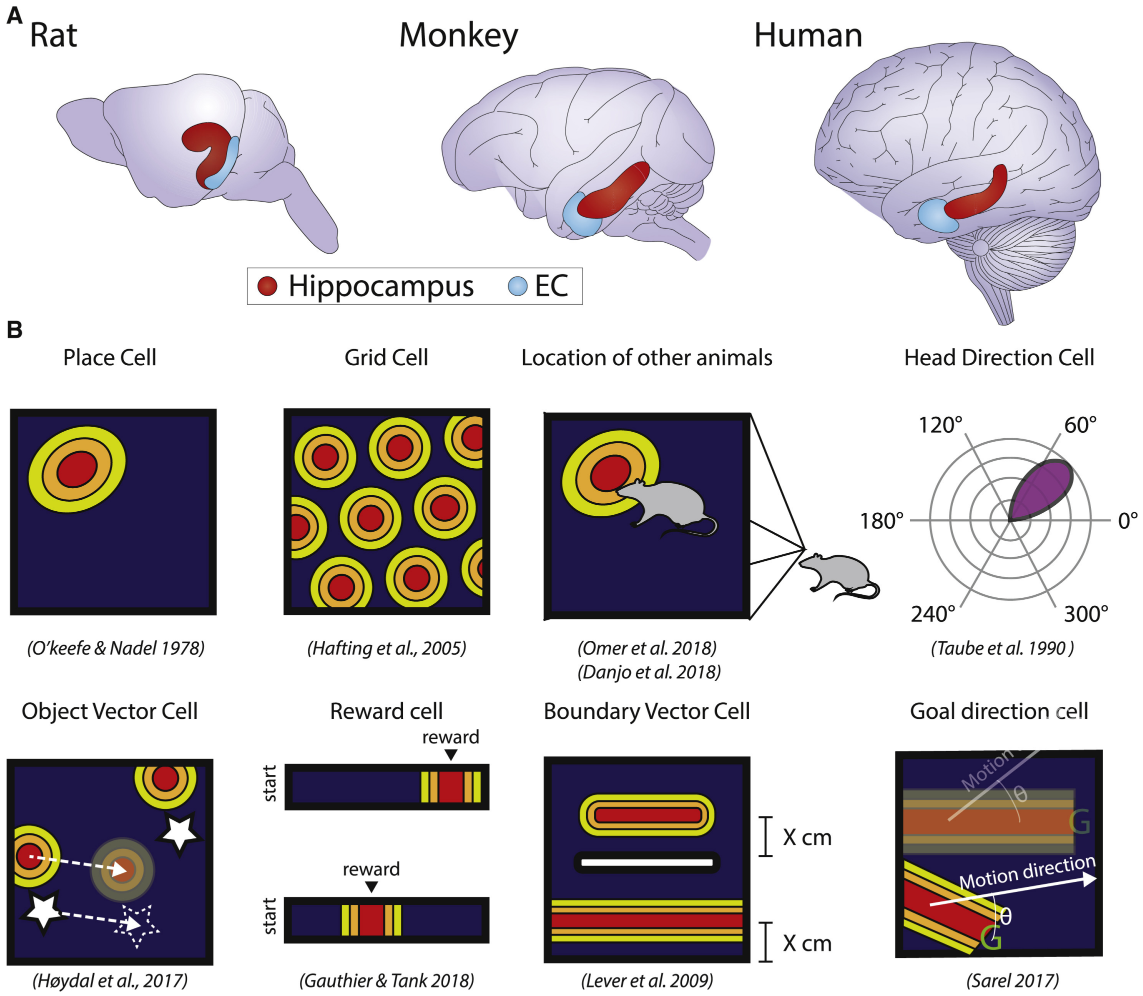

Figure 2.3: In search for neural correlates of cognitive maps in the brain (A) Anatomical location of the hippocampus and entorhinal cortex in the brain of different species. (B) Characteristics of the activity observed in hippocampal-entorhinal cortex cells together with corresponding dates and names of specific discoveries. Adapted from Behrens et al., 2018.

First experimental support for the hypothesis that the hippocampus functions as a spatial map were provided by the discovery of *place cells* (O'Keefe and Dostrovsky, 1971).

Results obtained from experiments conducted on rats with microelectrodes permanently attached to the skull had shown that there is a group of neurons firing when a rat is situated in a particular spot facing a specific direction. *Grid cells* were observed years later, and their discovery provided even deeper insight into cognitive maps mechanism (Hafting et al., 2005). In the original paper, it was postulated that the existence of grid cells determines whether the entorhinal cortex has a map-like structural organization. Based on neural activity recordings from rats running in enclosures, the authors have discovered that this type of cell is activated at multiple locations spanned together on a triangular grid. Their specific ability to picture topographic organization was then considered as a possible provider of an internal metric of space (Dang et al., 2021) to the brain. Grid cell firing patterns represent distances and angles between different spatial locations in the real environment. Both place and grid cells discoveries revolutionized the question asked originally by Tolman - how does the brain create an internal representation of the surroundings and how, based on this knowledge, humans can navigate through complex environments? Over the next years, a series of experiments led to the discovery of other cells in the hippocampal-entorhinal cortex (HEC) that are postulated to play an important role in mechanisms of navigation, such as boundary and object vector cells (Høydal et al., 2019; Lever et al., 2009), head and goal direction cells (Sarel et al., 2017; Taube et al., 1990), rewards cells (Gauthier and Tank, 2018) as well as cells encoding location in the space of other animals (Danjo et al., 2018; Omer et al., 2018). Altogether, these cells form circuitry that constitutes our current knowledge about something that we can call the brain's inner GPS.

It was also shown that neural activity in HEC can be modulated by the *theta* rhythm. Theta oscillations are one of the bands that can be seen in the electroencephalography signal, a brain imaging technique recording electrical activity through electrodes placed along the scalp. They occur in the human brain at a frequency between 7-12 Hz (O'Keefe and Recce, 1993). A quick change in place cell firing probabilities relates to theta rhythm and it is believed to stand for fast dynamics in spatial representation. Moreover, phase-amplitude coupling between theta and gamma (30-100 Hz) oscillations has been shown to accompany better working memory performance in both rats (Tort et al., 2009) and humans (Axmacher et al., 2010). Altogether, various brain cells encode the different types of information, and with neuronal activity, they constitute a framework that is believed to help animals and humans in navigation and orientation in space. Nevertheless, the brain is constituted of large networks of connections between multiple brain areas and thus, scientists look for a broader view into mechanisms that are involved in navigation.

2.3 Global coordinate frame

Since 1978, the dominant model in neuroscience relied on the assumption that cognitive maps are based on Euclidean geometry (O'Keefe and Nadel, 1978). Euclidean space is a space that we use in geometry and it consists of points and lines representing different dimensions. The primary property of mapping in Euclidean space is the ability to calculate distance and other metrics between two locations in space. Grid-cell-based metrics may serve such a general global frame for spatial cognition (Moser et al., 2008). As I mentioned before, grid cells are characterized by their hexagonal firing fields that provide a low-dimensional representation of space. Such a firing pattern is hypothesized not only to provide coordinates for space but also to perform path integration (Section 2.5) (Gorchetchnikov and Grossberg, 2007).

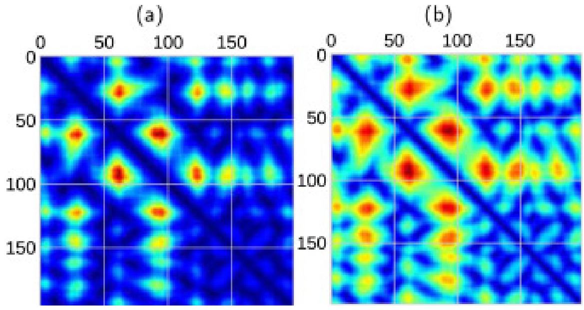

Figure 2.4: Distance metric calculated by grid cells
Figure adapted from Dang et al., 2021, in which researchers used parameters of grid cells to calculate the metric between every two positions from the selected trajectory, on (a) a distance matrix acquired from grid cells population and on (b) a Euclidean distance matrix.

However, assuming that grid cells provide a rigid metric representation as in Euclidean space would be a overinterpretation because while for instance, in 2-dimensional Euclidean space there is a $(0, 0)$ starting point, we do not have such a reference point in our brain. Recent experimental results successively revealed a more complex picture through investigating grid cell firing patterns. It was shown that such a unique pattern rescales and deforms in response to changes in the environment structure (Barry et al., 2007). It is noteworthy that when such a process happens for any environmental change, grid cells behave equally across all spaces. The grid cell fields rotate in their orientation but keep the same relative relation to each other. It suggests that they may provide us a sense of space while navigating, e.g., because we already have formed a globally consistent cognitive map for a given environment (Latuske et al., 2018). To visualize this problem, let's imagine that we are in some airport in a non-English speaking country that we have never seen before. However, we can orient ourselves in space, because we notice similarities in the airport structure, such as the check-in area, boarding gate, etc. It would not happen for places that we were never familiarized with, because for them we would not have a chance to build a cognitive map with established relations between specific locations.

To summarize, human efficiency in navigation may be enhanced by such a particularly characteristic firing pattern of grid cells that can help to constitute a consistent cognitive map. Each change observed in grid cell activations plays an important role in the real-time navigation experience that aims to help in orientation.

2.4 Non-spatial domain

It seems quite intuitive to think about learning cognitive maps (consisting of states, sequences, and transitions) in a physical space. This point of view is supported by the idea that physical locations may be encoded as vectors in the Cartesian (x, y) coordinates. However, real-world locations can be represented in our brain in a variety of ways (possibly also by capturing metric information that may be provided by grid cells encoding schema). Representing a spatial or more abstract structure does not have to preserve only metric information put together in a global coordinate system. The most important information that has to be maintained should try to reflect the general structure based on relations between multiple entities. Neuroscientific findings from research on patients with hippocampal lesions are supporting this more general idea of *cognitive mapping theory* (Buzsáki et al., 2022). Such impairment results in profound deficits in both spatial and non-spatial tasks (Keane et al., 2020; Ross and Eichenbaum, 2006) that do not depend on any specific metric.

Such a re-conceptualization of the cognitive map in terms of capturing connections between entities rather than concrete metric representation can be formalized by topology. A *topological space* enables us to define the structure of continuity, e.g. world that we live in. The most important property, being also one of the motivations to develop this concept, is that many geometrical problems do not depend on the structure of the objects involved, but on the way, these objects are put together. In topological terms, both square and the circle have many properties in common; i.e., they are both one dimensional and they enable to separate the space into the part inside and outside itself. However, they are not the same, because they are built from different configurations of the objects involved in their structure.

The idea that the human brain may form a common computational framework that governs the organization of both spatial and conceptual knowledge (Dehaene et al., 2022; Schiller et al., 2015) was raised by scientists quite recently. We live in a complex world in which such a consistent and systematic organization applied for different cognitive domains would support our flexibility in adaptation to problems that occur not only during spatial navigation (Behrens et al., 2018). It seems that navigation in the non-spatial domain could similarly involve the representation of abstract forms capturing relational knowledge. The structure of the real environment that we navigate can be in fact very similar to the organization of our social network or any other framework that consist of connections between elements within its structure. Altogether, it theoretically suggests that humans are using the same coding mechanisms that underlie both physical and abstract representations.

The flagship example which proposed that cognitive map may be used to encode non-spatial features investigated brain response to encoding an abstract multidimensional "bird space" (Constantinescu et al., 2016). In this study, human grid cells have been identified using functional magnetic resonance imaging (fMRI). Subjects were tested in the task of navigating abstract conceptual spaces in which they learned the visual stimuli and associated specific variations of an image of a bird according to two continuous dimensions: the lengths of the neck and legs (referred to as "bird space"). In each task trial, subjects watched a video of a bird morphing in the consistent ratio between two dimensions. To evaluate subjects' performance, they were instructed to imagine the outcome if the bird continued to morphe further and they had to choose an expected outcome between given options on the screen. Researches have found that subjects with greater hexagonal modulation of grid cells had more correct responses in total. This evidence suggests that grid cell codes known for involvement in spatial navigation may also be used in navigating through more abstract structures. However, these results are not profound, because of the nature of the fMRI signal that does not encode neuronal encoding directly, but only through correlation with blood-oxygen-level depending on the appearance of the stimulus. However, HEC may be a valuable conceptualizing tool for multidimensional cognitive maps in different contexts.

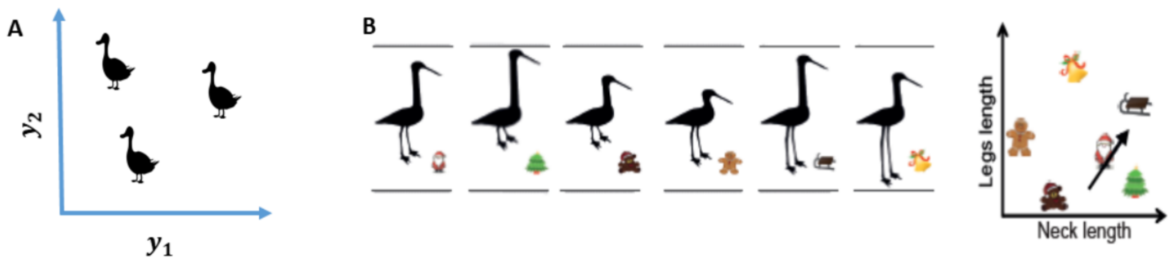

Figure 2.5: Abstract conceptual dimensions as the "bird space"
Experimental stimuli used by Constantinescu et al., 2016 in order to study the navigation through abstract knowledge. (A) Possible two-dimensional transformation of the bird, and (B) associated points learned in the training phase. Adapted from Anselmi et al., 2020

For the purpose of finding more insight into coding other abstract spaces, many other studies were conducted to study various phenomena, e.g. brain's perception of time. Instead of clock-face, the hippocampus may be considered as a general sequence generator that carries a content-limited structure (Buzsáki and Tingley, 2018). Navigation in a cognitive space does not have to be based on temporal dependence, but more probably on a sequence of events. We cannot assume that the brain changes with the things that we perceive as clockwise changes in time. These assumptions are supported by many experimental results; e.g. one study showing that hippocampal activity encodes information about specific letter sequences (Kalm et al., 2013) and not the time they were presented. Moreover, the subjective sensation of time occurs all over the animal world, and so far we did not see any rat with a watch on its wrist. It was also hypothesized that a cognitive map could be not only a system that enables navigation through both spatial and non-spatial spaces but a *skill space* itself (Schiller et al., 2015). There are studies indicating that the acquisition of abstract knowledge may involve structural changes in hippocampal gray matter. It was found that gray matter volume in the hippocampal formation was significantly correlated with years of piano tuning (Teki et al., 2012) or taxi driver experience (Woollett and Maguire, 2012). Similarly, social space incorporates a number of continuous dimensions, such as hierarchy, social status, and affiliation (i.e., closeness, intimacy, etc.). One of the most fundamental experiments where participants were asked to interact with other characters in a role-playing game proved that the hippocampus is also crucial for social cognition (Tavares et al., 2015). The response obtained from fMRI had shown that hippocampal activation keeps track of a vector metric indicating the location of characters in social space relative to the participant. What is more, relationship trajectories were stronger for participants who reported better social skills.

To sum up, the hippocampus (or rather the entire circuitry of HEC) supports experience-dependent mental navigation within various spaces built from sounds, letters, geographical locations, and many others. When it comes to comparison between dynamics of spatial and non-spatial behavior, reasoning in both domains follows similar computational principles (Wu et al., 2020). However, even if there are similar ways of generalizing in both domains, patterns of exploration may differ according to a given task.

2.5 Path integration

One of the basic characteristics of spatial behavior is path integration which enables us to realize when we are in the space that we traveled. When an animal (or human) is located somewhere in the environment, HEC may be able to calculate subsequent positions on the basis of how far and in what direction the animal moved. This idea was proposed for the first time by Darwin, 1873 that postulated that animals use self-motion cues to stay updated about their relative starting point position. Indeed, what path integration states is defined as the capacity to use self-motion cues to calculate an updated position by monitoring its trajectory in relation to a start location (Asem and Fortin, 2017). To make path integration work, an animal has to have an idea of its direction of motion, and such information is provided by a variety of cells that exist in HEC.

As introduced before, place cells are activated at specific locations in space and they are influenced by environmental cues (Burgess and O'Keefe, 1996). One example of an environmental cue that we use could be the name of the street put on the building, but for animals, it could be the position of the sun. There is still no consensus about how our location in space is represented in the brain. However, there are usually two conductive ideas that are distinguished when possible interactions between path integration and stored sensory information are taken into consideration, as shown in Fig. 5. In both cases, there is some metric information stored, and a direct path between two points can be calculated by subtraction of two vectors representing this places.

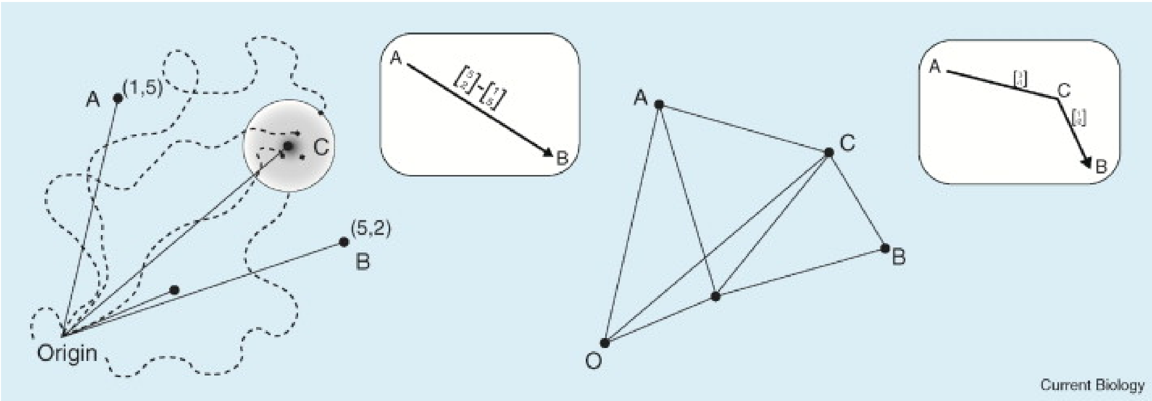

Figure 2.6: Two path integration systems
Left: path integration position stored in the Cartesian coordinates. Right: path integration as a graph representing the relation between connected places. Adapted from Collett and Graham, 2004.

Each place or physical location can be stored in memory together with visual or other sensory landmarks. Their existence is supported by the need of operating on egocentric coding information, which precision can be recalibrated when the position of an animal changes. The main concept of starting location stays the same, and animals should be able to keep track of the so-called *"home vector"*. During the whole process of integrating sensory information from the environment, inaccuracy to get back to a home vector increases with longer distances traveled from a starting point (Wehner and Rössler, 2013).

However, the situation changes critically when it comes to performing path integration inside a non-spatial domain. In this case, we would not accumulate self-motion information, but abstract movement throughout the topological structure itself, e.g. parent, grandfather, and so on in a family tree framework. What is crucially needed to perform a path integration within such a structure would be a consistent and compressed representation capturing all of the connections that imply what the relationship stands for.

2.6 Generalization

Fundamental frameworks of knowledge organization captured by a cognitive map can govern cognition that guides us directly to the main feature of biological intelligence, characterized mostly by its high flexibility. We do not exhaustively explore all of the possibilities in different tasks, but rather acquire the ability to generalize about stimuli based on observed similarities. With a minimized number of observations, we can rapidly adopt new behaviors and flexibly adjust them to specific goals.

Principles governing generalization were famously described by Shepard, 1987, who proposed a unified first law of psychology named *law of generalization*. Based on results achieved from his experiments using the stimulus-response paradigm on humans and animals (Shepard, 1958), he formulated a law stating that the probability of perceiving similarity or analogy between two items A_1 and A_2 is a negative exponential function of the metric distance $d(A_1, A_2)$ between them happening in the brain. The main implication of this theory is that similar stimuli are associated with similar responses, while at the same time when similarity decreases, then the response probability decays exponentially. Such cognitive information processing has been put into the frames of *mental time travel* phenomenon defined by Tulving, 2002 as the human ability to engage in active participation in the past or future imagined events. Such a mechanism requires generating event trajectories from possible environments and policies.

Altogether, generalization understood as the transfer of knowledge in different domains involves functional interaction between abstract and sensory knowledge as well as memory and imagination. However, it seems that the key is not only to transfer the information but also to generalize about sub-components from a finite number of observations about a specific task or environment. This has been an area of investigation for studies involving machine learning (ML) methods that train an artificial agent to learn an optimal solution for a given task through obtaining rewards. In the next chapter, I will discuss more in-depth a possible way of building a cognitive map representation through one of the most effective ML algorithms of reinforcement learning.

3 Building the model of the world with reinforcement learning

Understanding the brain's ability to learn effectively is one of the biggest challenges in neuroscience appearing all over the animal world. Even organisms deprived of the cerebral cortex (a part of the brain believed to be associated with higher cognition) can learn certain behavioral patterns, e.g. earthworms, fruit flies, and nematodes. One of the most studied and completely mapped brain connectome belongs to the nematode named *Caenorhabditis elegans* (Sasakura and Mori, 2013). Despite the detailed knowledge of the structure and connectedness of its nervous system, scientists are still unable to explain fundamental learning processes. *C. elegans* nervous system shares some features with the human brain, but it is extremely simple when compared to each other, because of a huge difference in the number of neurons, connectivity, and distance on the phylogenetic tree between these species (Strange, 2006). Therefore, it is difficult to find significant similarities in the complexity of the system, such as the brain in which different abilities emerge from the interaction with the environment on many levels. Additionally, there are many smaller concerns to consider, such as the ability of each organism to adapt to multiple environments. We are not able to find two organisms of one specie that have faced the same requirements of the ecosystem. Thus, it seems that the pure ability to generalize learned rules to novel environments is an essential feature of adaptive behavior in all species.

If not in the structures of the brain themselves, we can take a deeper look at the functionality of particular components of complex systems, by which we can also understand brains. A main feature of the brain as a complex system is the ability to generate other systems as complex as itself (Neumann, 1966). Essentially, both several hundred and several trillion neuronal systems can learn from interactions with various environments. For this reason, scientists started to describe the brain as a structure consisting of many networks which computationally model certain aspects of cognition. Starting from the previous century, brain abilities inspired new solutions developed in the field of machine learning (ML). Machine intelligence can partially simulate or even model efficiently how a particular property of the brain functions. The main feature that connects our brain with ML algorithms is the ability to carry out complex calculations that are aimed at optimization (Heilbron and Meyniel, 2019). In the case of machines, an optimization problem will be defined by a mathematical function, while for the brain of a simpler organism, this would be represented by specific searching behavior that will enable finding food and survival.

One of the properties of the brain is the ability to learn by searching for information from an experience gain. Searching was compared to the cognitive domain for the first time by James et al., 1890, sowing the seed of the reinforcement learning (RL) field that studies learning-through-searching abilities (Sutton and Barto, 2018). This approach is distinct from other ML approaches, because of the emphasis on learning from direct interactions with the environment without relying on exemplary supervision. At the outset, the RL algorithm does not have any specific world models or operating instructions, but at the same time, it knows what to do, due to defined possibilities of actions and overall goal (Mnih et al., 2016). On the one hand, such an algorithm goes through a simulated environment, and on the other hand, collects information about the effectiveness of taken actions. In this way, it becomes a self-critic aiming to maximize the final result that it can get.

The idea that we learn through searching by taking some actions seems to be one of the first that occurs when we think about the nature of learning. In Tolman's experiments, rats' behavior was guided by stimulus-response connections that enabled to uncover cognitive navigation abilities. The RL algorithm is based on similar assumptions, where an agent interacts with the environment and represents goal-directed learning while searching for the rewards. For an algorithm, the reward will be numerical, but in experiments involving rats, it will usually be represented by food.

3.1 Essence of reinforcement learning

RL uses a formally defined framework that describes the dynamics of the interactions taken by an agent with the environment represented as a set of states (Sutton and Barto, 2018), which dynamics are illustrated in Fig. 3.1.

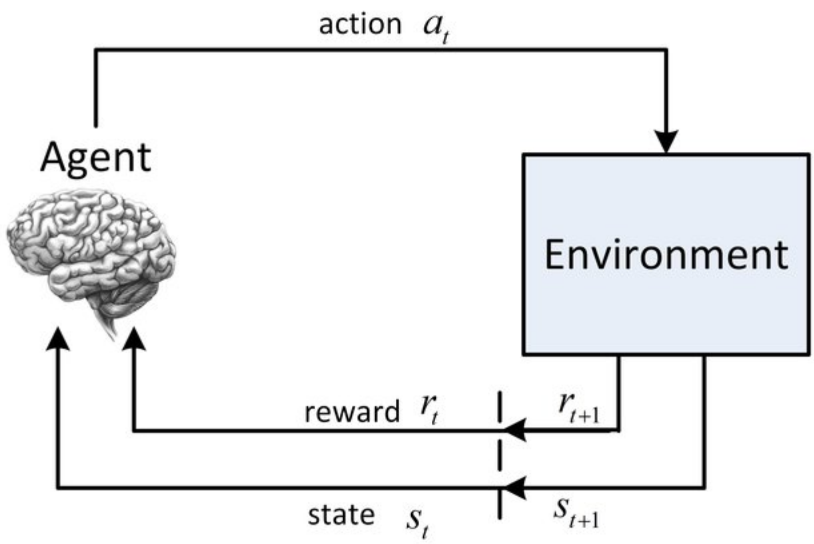

Figure 3.1: Interactions between agent and environment in RL
Reinforcement learning aims to figure out a policy that an agent (represented as a brain) will obtain through taking actions and receiving rewards at given time stamps. Adapted from Yu et al., 2019.

3.1.1 Agent

An agent is "the brain of an entire system" that can directly interact with the environment. At each discrete time step t, the agent executes some action a_t receiving at the same time a specific representation of the environment referred to as a state s_t. As a result of taking action, the agent gets a scalar reward r_t. Through such an iteration process, the agent can interpret information from the environment and learn. Indeed, the learning of an agent is based on trial and error to finally develop a policy that will allow it to solve the task as effectively as possible. Thus, the entire process is a time series that corresponds to the experience of an agent, which goal is to maximize the overall reward achieved.

3.1.2 Environment

The environment is the agent's world in which it lives and interacts, and it can be both more or less realistic (but it does not always have to be easy to visualize). In any case, the purpose is to create an environment such that an agent will be able to move between its parts. The agent interacts with the environment by performing some action, but at the same time, it cannot influence the rules or dynamics of the environment. When the agent performs some action a_t in the environment, it returns a new state s_{t+1}, making the agent move. The environment also sends a reward r_{t+1} that acts as feedback informing whether the action was good or bad.

In the case of having a fully observable environment, the agent can always determine the state of the world at any given point in time. It means that agent is aware of all states in the environment, e.g., in a chess game the agent would always able to infer position of all the pawns put on the board. However, an agent can also interact with a partially observable environment, which means that the agent can take an action only based on part of all possible observations. Such case can be represented by our egocentric visual scene which is limited to a particular view we observe at particular location in time.

3.1.3 Reward

The main purpose of RL is for the agent to learn an optimal policy that maximizes the "reward function". This type of learning is similar to *operant conditioning*, which states that human learning is guided by reward and punishment, both positive or negative (Watson, 1930). Biologically, human brains are hardwired to interpret signals such as pain or hunger as negative reinforcements, and signals like pleasure and food intake as positive reinforcements. The ability to tune our behavior by learning different types of associations seems to have a crucial role in goal-seeking behavior. And that is where the RL enters, as the only ML method that uses a reward hypothesis supporting the idea of learning associations between actions and states. On each time step, the environment sends to the agent a reward that can be compared to the agent's expected value of the action taken. By such feedback structure, reward prediction error can be calculated and it guides behavior that leads to reinforcing some actions while others can be extinguished. The agent aims to maximize the expected return G_t as a sum of rewards:

$$G_t = R_{t+1} + R_{t_2} + \ ... \ + R_T, \tag{1}$$

where T is a final time step. It is worth emphasizing that we are not insisting that the agent has to achieve the goal of maximum reward. Trying to maximize a quantity does not mean that the quantity has to be ever maximized, but rather close to the optimum.

3.1.4 Policy

Having already mentioned what the actions are, it is important to describe more of this foundation of interactions between agent and environment. The agent selects actions to maximize total future reward and the way of achieving the goal is guided by some policy (behavior of an agent). A policy is a simple mapping of action given some state represented as a function that defines the probability $\pi(a|s)$ over actions a for each state s at a particular time step t:

$$\pi(a|s) = P[a_t|s_t]. \tag{2}$$

This process includes some type of planning because actions may have long-term consequences and the reward can be delayed. Learning of policy had already appeared in Tolman, 1948 experiments, where he was observing how rats' behavior (in other words, their policy) changed when the rewards appeared in the maze. The information about how good or bad it is to be in a particular state can be provided by a value function, which enables predicting future rewards. The value of state s when the agent is following some policy π is denoted by $v_\pi(s)$, which is the expected return starting from a particular state s and following a policy π for the next actions until the agent reaches the terminal state. A state-value function can be represented as:

$$v_\pi(s) = E_\pi[G_t|s] = E_\pi\left[\sum_{k=0}^{\infty} R_{t+k+1}|s\right], \tag{3}$$

where E_π stands for an expected value of a random variable given that the agent is following policy π in many time steps and k denotes the number of time steps.

3.1.5 Model

Additionally, for planning some behavior guided by a policy π, agents may use a model of the environment. Algorithms that use models and planning strategies are called model-based as they aim to understand the world to create a model representing it (something that could be theoretically referred to as a cognitive map). On contrary, model-free agents use a simple trial-and-error learning strategy. The big advantage of model-free learning is that the algorithm may be easily transferred to another environment architecture.

According to Sutton and Barto, 2018, the distinction between model-free and model-based RL algorithms can be compared to automatic and goal-directed behavioral patterns. By automatic behavior, we can understand reflexes that are triggered by some stimuli and they do not involve processing information directly by the brain. When it comes to goal-oriented behavior, the agent is controlled by the knowledge of the value of goals and the relationship between actions and their consequences.

3.1.6 Markov decision process

RL framework includes a sense of cause and effect, as well as a sense of uncertainty in achieving goals in an unknown environment. Mathematically, using the properties of stochastic processes, it is possible to describe knowledge about past, present, and future states by a *Markov property*. It tells us that the probability of the next state given the previous one is the same as the probability of the next state given all of the previous states.

$$P[s_{t+1} \mid s_t] = P[s_{t+1} \mid s_1,\ ..., s_t] \tag{4}$$

The main idea behind this property states that the future is independent of history, given the present state. It makes the whole system work better because an agent stores only one state of information (a.k.a *Markov state*) which contains all of the necessary details obtained by experience.

In the case of such a defined system, both the environment and the task are also said to have the same property, which means that the environment response at $t + 1$ depends only on the state and action representation at t. Thus, the agent can predict the next state and expect the next reward given only the current state and action. The RL algorithm that satisfies the Markov property (Ross and Eichenbaum, 2006) is called a *Markov decision process* (MDP) and it provides framework for modeling decision-making problems. MDP is defined by its state and action sets and the dynamics of the environment. Altogether, the probability of each possible pair of the next state and reward given any state and action is denoted as:

$$p(s', r|s, a) = Pr\left[s_{t+1} = s', r_{t+1} = r \mid s_t = s, a_t = a\right] \tag{5}$$

This framework motivates the need for efficient learning strategies showing how optimal solutions could be obtained.

3.2 Human and machine intelligence

Various kinds and degrees of intelligent behavior occur in people, animals, and some machines. It is one of the most intensively studied topics, mostly to gain insight from human cognition to improve machine learning methods, as well as in the opposite direction - using advances in machine learning to understand human intelligence. We, as humans, learn through experience either directly or shared by others, but we are also using some innate knowledge learnt by evolution, such as language preadaptations. When in the case of machines, everything has to be learnt through experience in the form of working on data. However, it does not have to exclude the possibility that machines can simulate intelligence, because they still may be able to complete a given task in the presence of an unreliable and dynamic environment. It seems that it is just a matter of algorithms and adjusting machine actions.

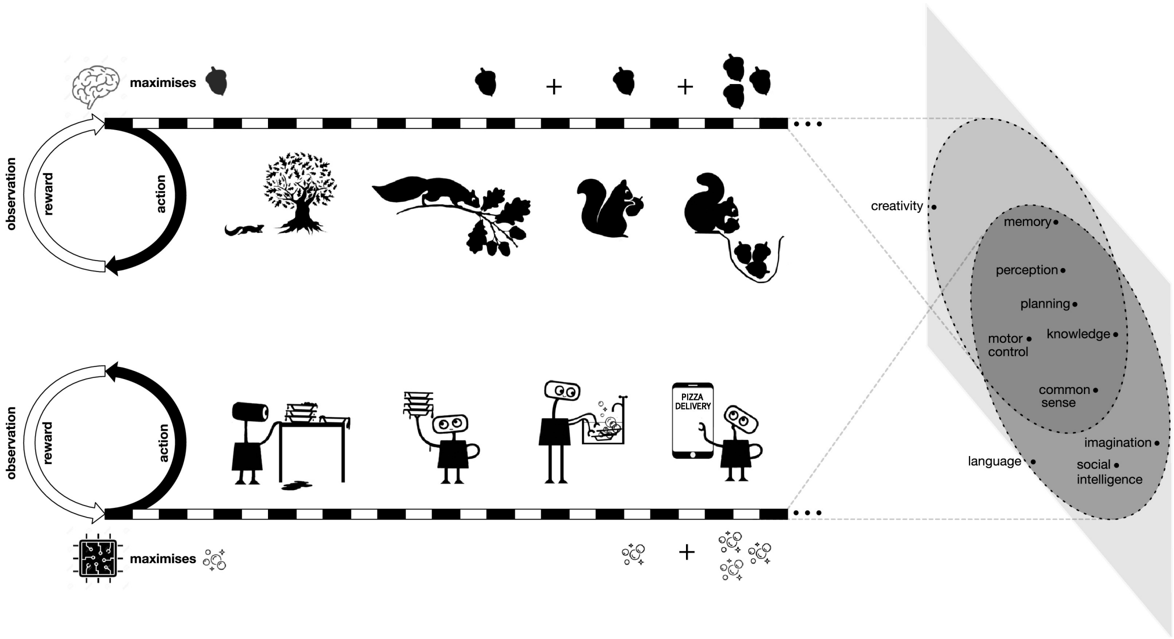

Figure 3.2: Biological and artificial example of goal-oriented behavior
Example of goal-seeking behavior represented by both biological (top row) and artificial system (bottom row). From the biological perspective, goal-oriented behavior can be understood in terms of maximizing the amount of food collected by a squirrel. While an artificial system represented by a robot has the goal of keeping the house clean. Both organisms have to use the entire repertoire of cognitive abilities that can be associated with a general understanding of intelligence, such as memory, perception, motor control, and so on. Adapted from Silver et al., 2021.

Let's get back to the topic of the RL framework I was describing in the previous section. Do we have some evidence for an analogous system that may occur in the human brain that could be simulated by RL? There is no doubt that both human and animal behavior can be characterized by decision-making processes by which actions can be selected in a specific environment. RL provides a framework within which fundamental questions of behavioral neuroscience can be analyzed. The generic objective of reward maximization drives behavior that exhibits goal-seeking studies in human and machine intelligence (Silver et al., 2021). The presence of a goal in the environment makes animals orient themselves toward it, which represents the so-called *beacon navigation* (Gallistel, 1990). Beacon navigation refers to situation in which the navigation is oriented toward a specific location, such as the lighthouse that informs about obstacles in the open water or port area. Early approaches in psychology, focused on a behavioral approach, fall under the category of conditioning. One of the suggested theories of operant conditioning described learning as a shift of behavior in response to rewards or punishments (Watson, 1930). However, late twentieth school of psychologists supported the idea of the model which put learning in frames of an active process of making predictions about immediate rewards guided by dopamine (Rescorla and Wagner, 1972). The model itself turned out quickly to be limited, because it did not differ much from conditioning idea despite adding a brain correlate. In practice, the difference between expected and received rewards is not crucial for learning processes to occur. Tolman himself criticized the behavioristic approach which stated that learning was guided only through reinforcement and punishment. He particularly emphasized the role of latent learning guided by many cognitive abilities, such as cognitive mapping. Nevertheless, the underlying principle of prediction error learning has not been fully rejected as one of the forms of goal-directed behavior enabling learning from our own mistakes in novel environments.

Experiments conducted on monkeys made it possible for scientists to interpret the phasic dopaminergic firing patterns as a possible reflection of prediction error signaling (Schultz, 1997) in the brain. Dopamine, a neuromodulatory molecule that functions as a neurotransmitter released by neurons to send signals to other nerve cells, is at the root of this phenomenon. It was shown that the activity of dopaminergic neurons corresponded to a temporal difference model that learns prediction errors. According to these observations, dopaminergic neurons could convey information about current significant events and the predictive value of this state - in such a way the whole circuitry of brain structures involved in dopamine transportation could compute reward prediction errors. The greater the signal (e.g. hunger), the greater the likelihood that behavioral sequences aiming at finding food will be executed by the organism despite obstacles that may arise. Based on these results, it is hypothesized that signals collected from dopaminergic neurons may play a role of an internal critic, which in case of artificial system could be represented as a gathered numerical reward (Niv, 2009).

In fMRI studies, particular brain regions correlated with prediction error were identified. Regions involved in both signaling prediction error and release of dopamine were located in the ventral tegmental area (VTA) and substantia nigra (SNc). However, some independent (or at least not directly dependent on dopamine release) structures were also observed. Together with VTA and SNc, they construct a so-called *mesocorticolimbic circuit* which is known also as a reward system (Arias-Carrión et al., 2010). This interconnected system is particularly involved in incentive salience, i.e. desire or craving for some reward, motivation, and associative learning and it is responsible for reward-related cognition. Achieving some reward motivates an organism and induces approaching behavior.

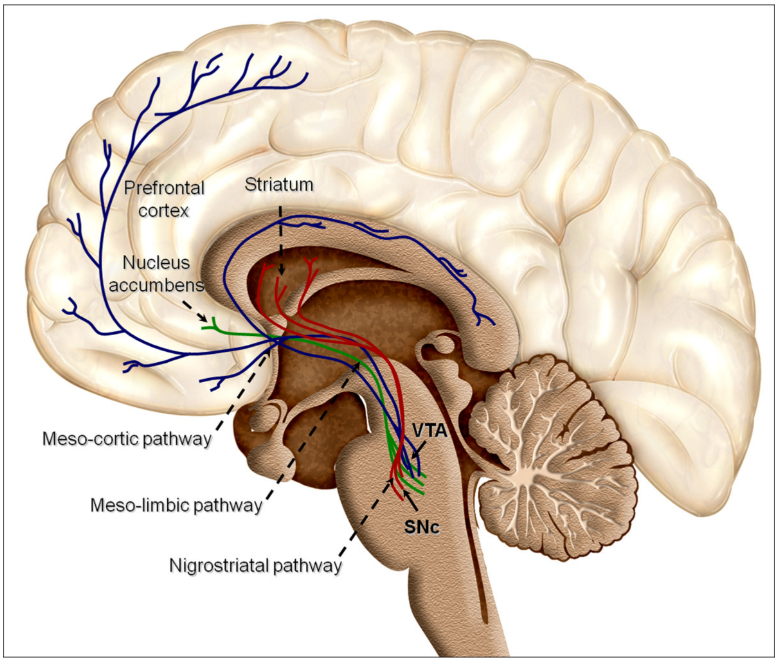

Figure 3.3: Reward system in the human brain
Brain structures constructing the *mesocorticolimbic circuit* are involved in goal-seeking behavior connected to reward presence. Projections of dopaminergic neurons playing a significant role here are concentrated mainly in midbrain structures, such as SNc and VTA. These projections go to the striatum, the dorsal, and central prefrontal cortex. Different functional connections within the network correspond to specific properties of the entire system. Adapted from Arias-Carrión et al., 2010.

Neural correlates of learning through searching in novel environments have been associated mostly with the brain reward system. However, precise search algorithms and the level at which they would operate are still unknown. There is no doubt that both humans and animals are performing some theoretical predictions on internal models of the environment. It is believed that the human orbitofrontal cortex (OFC) represents a cognitive map of a specific task space and enables to perform operations in the current state at some time point. According to experiments conducted by (Schuck et al., 2016), fMRI activity in human OFC contains information about a specific location in a mental map of a subject. This brain structure activity, representing a cognitive map, can be investigated further for possible insight into specific information encoded for maximum effectivity in a task. Nevertheless, the whole learning process may be even more complex, because it can depend not only on making predictions, but also on applying different exploration strategies, and so on.

3.3 Exploration-exploitation dilemma

Another key to efficient learning is to balance the two competing goals of exploration and exploitation. An agent navigating in space, just like a human who explores the external environment, encounters a *exploration-exploitation dilemma*. This problem can be described by considering whether we should make the decision that seems to be optimal for now (with the assumption that the current knowledge is reliable enough) or whether we should make a decision that seems to be suboptimal now, keeping in mind that gathering more information can help us to improve. We must try a variety of actions and progressively favor those that appear to be the best. In RL to maximize long-term rewards, an agent that gathers more information about the environment has to face the expense of temporarily choosing less rewarding actions. Exploration refers to trying actions that improve the model, whereas exploitation means behaving optimally given the current state of knowledge. In a stochastic task, each action must be tried many times to gain a reliable estimate of its expected reward.

To find the optimal strategy under the uncertainty of the whole environment, we need to accept some trade-offs. Exploratory behavior is needed especially to build an internal model of the world, which we can understand as a cognitive map. Information about the structure of the environment is needed for both a biological and an artificial agent, and it is guided by curiosity (Schulz and Gershman, 2019). In this case, not only the reward itself matters, but also builds up a map of the environment in which the exploration takes place. Getting to know such a model of the environment enables better prediction of possible rewards (Peterson and Verstynen, 2021). Curiosity, then, can be satisfied by exploration, to improve later exploitation and get better results.

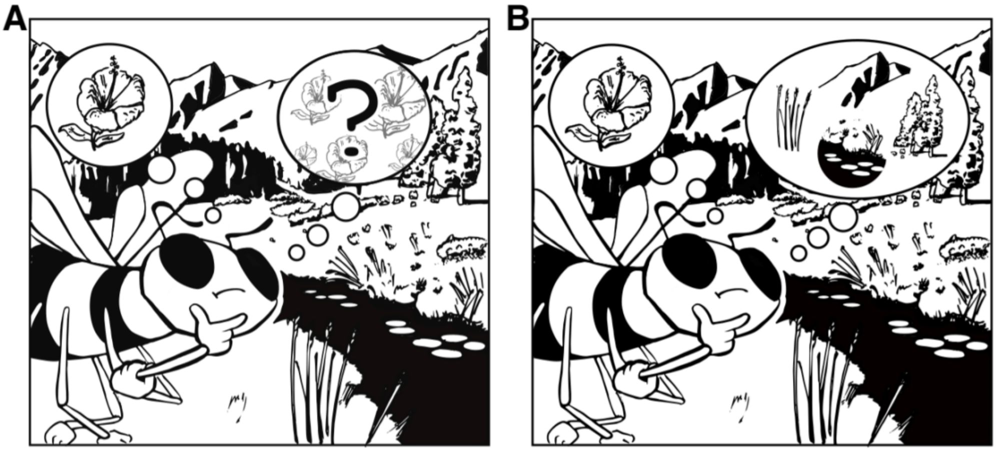

Figure 3.4: Different views on exploration-exploitation dilemma
In the figure, a bee is faced with the dilemma of how to get to the plant to find a portion of food. (A) From a classical point of view, a bee is faced with the choice of picking a known flower or exploring and looking for new flowers. In both cases, it aims to maximize the chances of getting as much food as possible. (B) A different perspective on the same dilemma points out that the bee aim to maximize the chances of finding food trough building a model of the world from which the best route can be chosen. Construction of cognitive map becomes a key goal for a bee to find maximal reward. Adapted from Peterson and Verstynen, 2021.

In RL, the exploration-exploitation dilemma is limited to the aim of maximizing the overall reward signal repeatedly through selecting an action, as well as observing, and collecting the resulting reward. We want an agent to explore the environment, but not so much that its performance is greatly decreased. Since optimal solutions are generally unavailable, any other solution to the exploration-exploitation dilemma in different environments is likewise non-existent. At the same time, how to reconcile this problem with the fact that mainly such exploration will be effective, but only for one environment? Focusing on a single aspect of learning makes it hard to compare and integrate the results to find information that can be generalized. Recent study has shown that people use the structure of presented decision-making task to generalize, explore and plan Ludwig et al., 2022. The larger the environment is, the correlation between specific parts of the underlying structure becomes more valuable for the subject to operate on. It seems that structured environment makes the exploration more organized and thus, easier to work with. At the same time, when there was no regularities presented, exploration became random, and more sophisticated planning was needed.

However, it is still unknown, how people balance the dual demands of exploration and exploitation. Most surely, no solution is both perfect and practical for this type of dilemma. It is difficult to study abstract strategies undertaken by humans in various types of tasks. However, it seems particularly important for a subject to have the freedom of choosing how much to explore and when to start to exploit. This criterion cannot always be satisfied, especially if people perform decision-making in a limited number of trials.

3.4 Reinforcement learning applied for spatial navigation

Spatial behavior usually takes place in the structured environment where an agent looks for intersections near associated items. Application of RL formalism allows to model learning as searching through simulation of exploration in a maze (or labyrinth) within its confusing structure. Such navigation task happens through exploring complex branching and making decisions of paths and directions to take, starting from the entrance to some goal position. A 2-dimensional grid maze can be defined to run a classical RL algorithm.

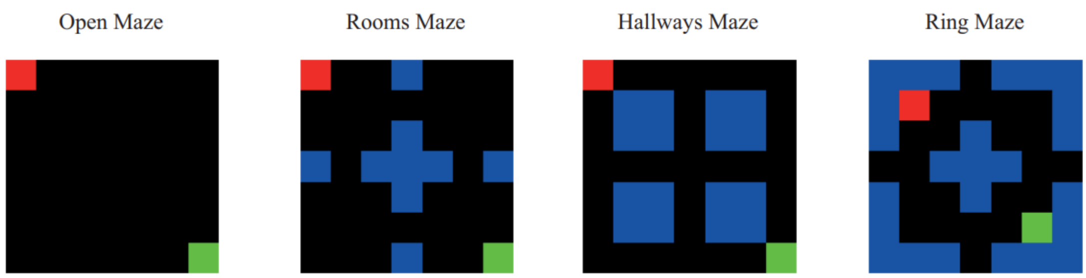

Figure 3.5: Different topographies of grid maze environments
From left an open maze, four rooms maze (used in many classic RL problems to study hierarchical learning (Botvinick and Weinstein, 2014)), hallways maze, and ring maze. Red: agent position. Green: goal position. Blue: wall positions. Adapted from A. W. Juliani, 2020.

The only thing that an agent can do in such an environment is pure exploration through trials and errors. At the first stage of exploration, a route toward the goal must be found for further goal-oriented exploration to happen. An agent has an infinite number of possibilities to take, but his job is such a search through the environment that will enable him to find the shortest or the most optimal way to reach the desired state. However, how does the agent know when it reaches the goal? Simply, when it arrives at the desired goal grid square it gains the biggest reward. The desired policy requires finding the shortest path between two points on - start and goal position - which can be acquired using Dijkstra's algorithm solver. Dijkstra et al., 1959 developed an algorithm that represents the environment in the form of a graph and looks for the shortest distance between the starting node and the final destination node. To visualize the whole process, Dijkstra's algorithm was applied to an artificial maze built using mazelab package (Zuo, 2018) - code and visualization available on github.com/wsojka00/graph_navigation. In a complex environment, the chances of finding such combination on the first iteration are rather impossible. In a hypothetical agent brain, at the first stage of exploration, the grid cell firing pattern would be random and determined by environmental cues. However, finding this path is desired by the RL agent that aims to build a sufficiently extensive model that enables it to generalize and exploit what it learned. At the same time, the model should form a universal consistent map that would allow capturing the most important information encountered by the agent during the exploration.

However, how does an agent guide its learning? Human exploration in a real-world environment is taking place on the abstraction level as a cognitive map, likewise in an artificial agent it can be modeled by a policy and value function using MDP. The process of abstraction and the form it takes has not yet been standardized and throughout the decades, various algorithms with different assumptions about the agent's state of mind were proposed; e.g., temporal difference learning (Russek et al., 2017) or successor representation (Stachenfeld et al., 2017). Evaluation of specific algorithms enables the observation of different navigational strategies that arise from abstractions of the underlying complexity of environment-agent systems.

Agent model-based representation of the world captures low-level knowledge about the structure of the preserved information gained during exploration. The ability to screen the most relevant information collected at each state is the case of preserving only this information that is comprised with the maximum ignorance of irrelevant details. Each agent, both artificial and biological, represents only a specific configuration of the world since we cannot store the entire "world state" in our memory. Given a high-dimensional observation of the world, only a small subset of information is important for navigational purposes. Such sparsity observed in such tasks is referred to as *curse of dimensionality* (Bellman, 1966), because of its countless dimensions of real-world unable to process efficiently. RL provides the crucial advantage of a formalization of narrowing down all the information to optimize. In this case, a cognitive map learned by an agent must contain only information that is relevant to downstream behavioral tasks given to an agent. That is why it has to be optimal to capture the complexity of the maze and enable sufficient navigation towards the goal. Moreover, such a map has to have the possibility to be generalized to flexibly behave when facing different challenges. An agent needs to "be aware" of the possible ways of interacting with the environment, understand its characteristics and associate past actions with future rewards. The remaining question is, how such a structure can be captured in a simple form preserving only the most important information in a simple form? I will try to answer this question in the following section.

4 Graph-based mental representations

According to *representationalism*, a philosophical theory of perception, scientists hypothesize that our brain perceives information in a simplified form and creates a mental representation based on real objects observed all around us. Humans' capacity to learn and represent multiple problems have to be balanced between the complexity of the mental representation and its utility for future usage (Ho et al., 2022). The cognitive maps phenomenon aims to capture compact and sparse mental representations of the environment structure used for navigation purposes. In the case of encoding information from space, experimental work has brought a lot of insight into possible brain regions and cellular activity that underlie such neural encoding systems. However, there is still a bridge between how neurons represent information, what computations are made by specific cognitive systems, and how we are able to flexibly adapt our behavior to a wide range of cognitive tasks. One of the major challenges is to explain how humans are able to infer the structure of complex environments to create simplified mental representations that capture the most relevant information. Structure learning has been proposed to rely on human's ability to learn the networks connecting discrete items and events in the environment (Lynn and Bassett, 2020). Such an approach enables to generalize of cognitive maps to many different domains that represent any kind of an organized structure; e.g. language with specific letter combinations, and syllables, as well as social networks, web of concepts, and many others. Learning specific structures of networks supports performance in a wide range of human daily challenges that help to perform abstract reasoning, flexibly adapt to changing environments (not only physical ones), and enable planning of the actions.

Studying the process of building relational knowledge can be informative in the context of how humans actually perform many of our cognitive abilities and reason about multiple network structures present all around us. People's ability to learn networks can be explained as an ability to uncover relational knowledge between single entities in sequence. Researchers have established that human behavior and cognition may depend on such temporarily evolving global topologies of networks in many environments. Given such experimental indications, it has been proposed that humans are able to acquire simplified mental representations through graph learning. The field of graph learning is highly interdisciplinary and aims to build computational techniques from network science, information theory, and statistical learning to provide models that can help in understanding human cognition on multiple levels of their lives. In recent years, graph-structured data has become a ubiquitous and universal language for describing complex systems from different domains. Within this particular structure, due to interactions between items, a variety of dynamics emerge and create regularities or patterns.

The brain itself can be also understood as a complex system, in which both structural and functional properties have features of complex networks Bullmore and Sporns, 2009.

Structural brain networks encode anatomical histology of connectivity between specific regions, while a functional brain network represents correlation of brain activity as a time series. Both of them can be acquired more or less precisely using different neuroimaging techniques and they can be analyzed with a graph-theoretical approach. These networks are based on topological connections and the brain's ability to propagate information through these structures. Such abilities to perform different inferences on graphs might have co-evolved with the human species' growing network size and diversity of topologies. Nevertheless, since topological networks can be observed within the structure and functionality of the brain, do mental representations also follow similar assumptions? The physiological and the computational mechanisms by which a cognitive map comes into existence remain vague. However, important insights into its structure can be obtained through geometric and topological constructions.

In this chapter, I will focus on putting most of the assumptions that scientists came up with into a common topological framework captured by graph theory. I refer to experimental results suggesting that different brain regions capture egocentric geometric sensory inputs and transfer them to cognitive maps enabling topological storage of information from an allocentric perspective. Such a topological approach implies having general knowledge about the structure of an environment and being able to infer it on the abstract level. Starting from a basic introduction to graph theory and the graph learning field, I propose putting cognitive map phenomena in graph form. Moreover, I provide experimental evidence of the brain's ability to store and process topological knowledge. In the end, I try to extend possible applications of such information processing from spatial environments to less precisely defined navigational tasks, like navigation in social networks.

4.1 Graph learning

The field of studying graphs - graph theory - has its beginning in the 18th century in the merchant city of Königsberg (present Saint Petersburg). Seven bridges throughout the Königsberg river were built to help in communication between traders. This arrangement of bridges gave birth to a puzzle called the *Königsberg Bridge problem* - can one walk across all seven bridges and never cross the same one twice? Despite many attempts, such a combination of paths was not easy to find. Leonhard Euler also searched for an answer to this question, but he approached this problem from a unique perspective. He constructed a graph that consists of 4 objects (nodes) corresponding to a patch of land along with a set of interactions (edges) for the bridges. He defined the very first property of graphs and proved that there was no path that would cross all the bridges without revisiting one of them twice because the underlying graph has more than 2 connections with an odd number of edges. This puzzle is relatively trivial, but it gave rise to a field of mathematics that studies the complex properties of graphs.

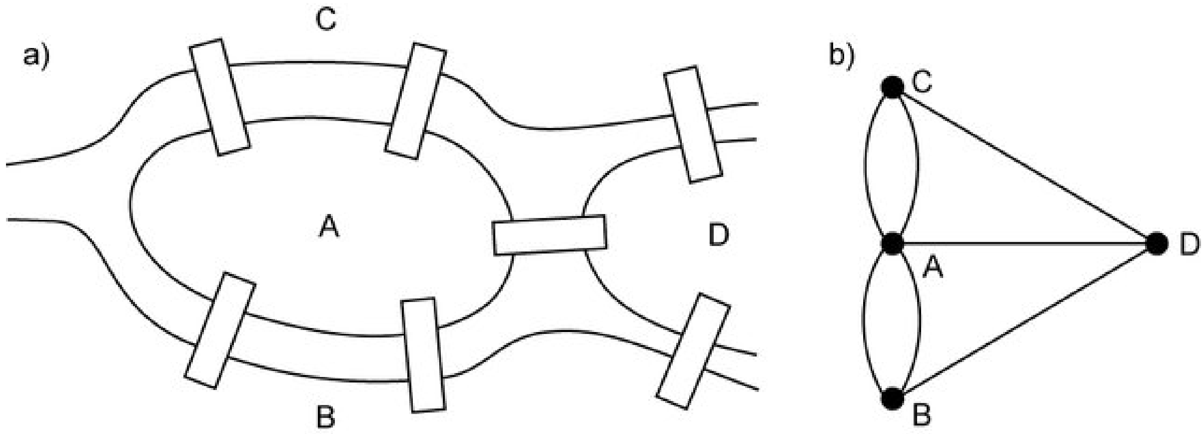

Figure 4.1: The Bridges of Königsberg
Depiction of the problem solved by Euler in (a) a realistic scene and (b) represented as a graph. Adapted from Boguslawski et al., 2011.

The power of graph formalism lies in the focus on relations between objects, which I refer to as topological relationships (Di Felice et al., 2009) that describe the position of objects embedded in space. Moreover, not only the graph formalism itself may be used to represent different networks, but also it has been observed that some problems are easier to understand when represented as a graph. Additionally, some properties encoded in a graph structure may limit or enhance the dynamics of the information propagation within the system. Recently, a new interdisciplinary field of graph learning has been trying to explain how humans infer and represent complex relationships in the form of graphs. In real-world large structures, multiple embedded subgraphs with their own properties may arise and help to understand the dynamics of the entire system. One of the most important properties that seems helpful in understanding the underlying structure of a given task is an innate ability of *statistical learning* (Daikoku and Yumoto, 2017). We are able to infer statistical structure embedded in a presented sequence of events of a novel set of stimuli. Graphs are the perfect tool for such a task because they simply encode the variations in frequencies of transitions between two nodes from continuous streams of data (Lynn and Bassett, 2020). The transition probability from node i to neighboring node j can be calculated as follows:

$$p_{ik} = \frac{1}{k_i},$$

where k_i is the degree of node i; in other words, it is the number of connections with this particular node j. It seems trivial, but the ability to infer such information enables to reveal of useful statistical information about the topological relationships of the environment. It was experimentally shown that humans anticipate more probable transitions and they are sensitive to variations of local properties of networks (Lynn et al., 2020).

In addition to the local properties of nodes and edges, graphs also yield large-scale characteristics of their structural representations. One of the most universal examples is *graph clustering* (the tendency of having a common neighbor between a pair of nodes) which gives rise to *modular structure* (tightly connected modules of nodes) that can be observed all around us. Even neuroimaging studies have shown that the network's modularity impacts our neural activity (Schapiro et al., 2013) and helps in building coherent mental representations of the environment (Karuza et al., 2017). To understand how various factors affect our cognition, scientists have proposed a model that creates graph representation changing relative to the form of a function lynn2020abstract. An exemplary function $f(t)$ puts the weights on the transitions of a given distance between nodes ij and estimates the form of the true underlying graph structure. If the transition distance of length in an internal network representation of a human would be equal to 1 (Fig. 4.2A), then the unbiased estimate of the true transition structure P can be revealed (similarly to most probable maximum-likelihood estimation). Although, if transitions over distances are combined (Fig. 4.2B), large-scale features of the graph are represented within its entire form. In the case of equally weighted transition distances (Fig. 4.2C), we lose all information and uncover a graph of meaningless connections. All in all, adapting transitions to distances that can be applied by people, enables detection of the global structure of the network and make coherent representations of the underlying structure.

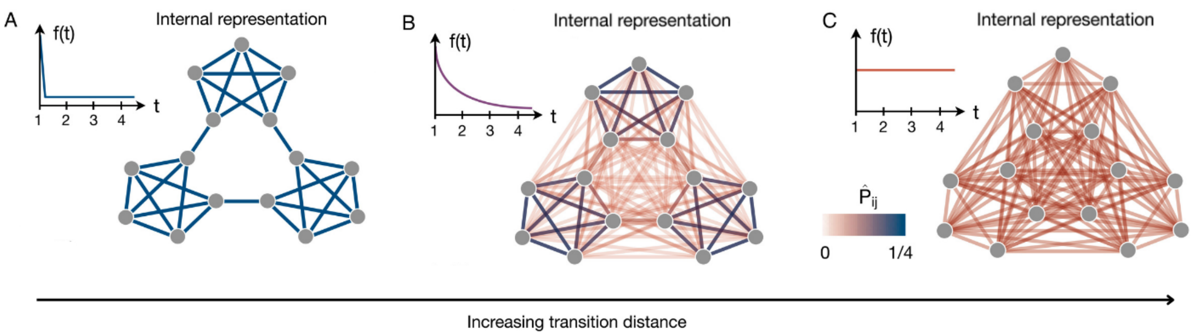

Figure 4.2: Decoding internal network representations
Visualization of the emergence of a network representation $\hat{P}$ (right) with respect to different forms of weight function $f(t)$ (left) for learners that consider only transitions of length one (A), longer distances depending on higher-order features of network structure (B) and for those that equally weight transitions distances (C). $\hat{P}_{ij}$ estimates equally transition probabilities about the true values of P_{ij}. Adapted from Lynn and Bassett, 2020.

To summarize, graphs provide an elegant theoretical structure and offer strong mathematical foundations that we can build to represent and analyze real-world phenomena of complex systems. Graphs enable us to take out the essence of the environmental structure based on connectivity between entities, while then memory processes may enable us to reconstruct specific contextual information. We are throwing away extra information so that we will be able to store in our memory simplified representations enabling a quick inference.

4.2 Cognitive map as a graph

Since we already have some intuitions about how graphs could be learned and which features could be encoded, we can move to the next question: how do we actually infer such representations throughout the navigation in the real world? It is believed that a cognitive map is a temporary, simplified model of a specific task environment that includes only relevant information for one's current goals and circumstances. The most important information that has to be preserved in such a simplified mental model has to be both useful and simple, including details that are necessary to solve the problem at hand. This capacity of storing and recalling low-level information about some structure can be acquired by a representation based on relatedness, where locations are connected in a directional and motion effort-based manner (Hamilton, 2020). In such a setting, we are ignoring irrelevant information, i.e., where the buildings are, the geographical layout, and similar. After some experience in creating cognitive maps (e.g., spatial ones) that reproduce observed environmental patterns, we may be able to anticipate and generalize about other possible configurations by probabilistic structural inference (Sharma et al., 2021).

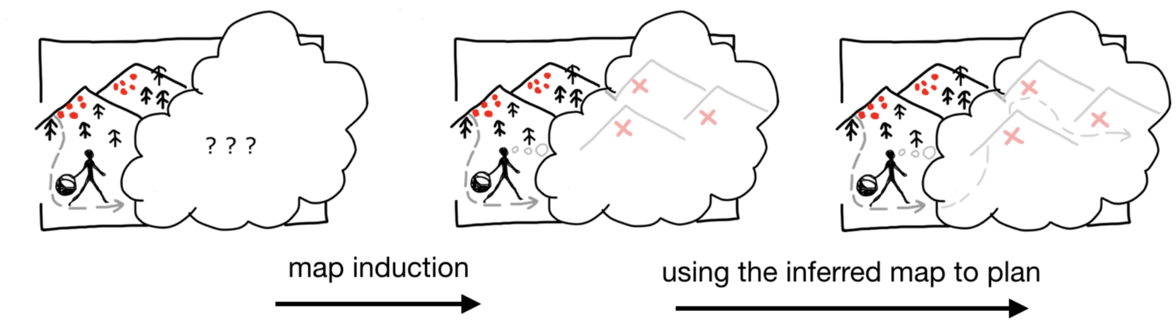

Figure 4.3: Structural inference of unknown places

Map induction based on path observations of similar environments used to minimize exploration and generalize about possible actions. Adapted from Sharma et al., 2021.

As I tried to show in the previous section, graphs are a flexible tool for representing a different kinds of problems, but the remaining question would be: how do we know which graphs to build? A key difficulty is that these maps are collective, abstract phenomena that cannot be reduced to a simple combination of inputs provided by individual neurons. An individual, sensory observation cannot define a state, since two identical observations can have different consequences. While single reflection seems to be not enough to infer latent state representations, sequences of observations are. Therefore, instead of considering single observations, we look at relevant for us sequences of events that we can refer to as states, similarly to the RL approach. The graph-based approach tries to model the environmental states followed by taken actions as nodes and spatial constraints between them as edges. However, in the standard approach, the size of the map expands quickly while an agent explores the world. Information storage has to be optimized such that the layout of the environment will still be preserved. From a graph perspective, we would be ignoring redundant nodes and manipulating only important content.

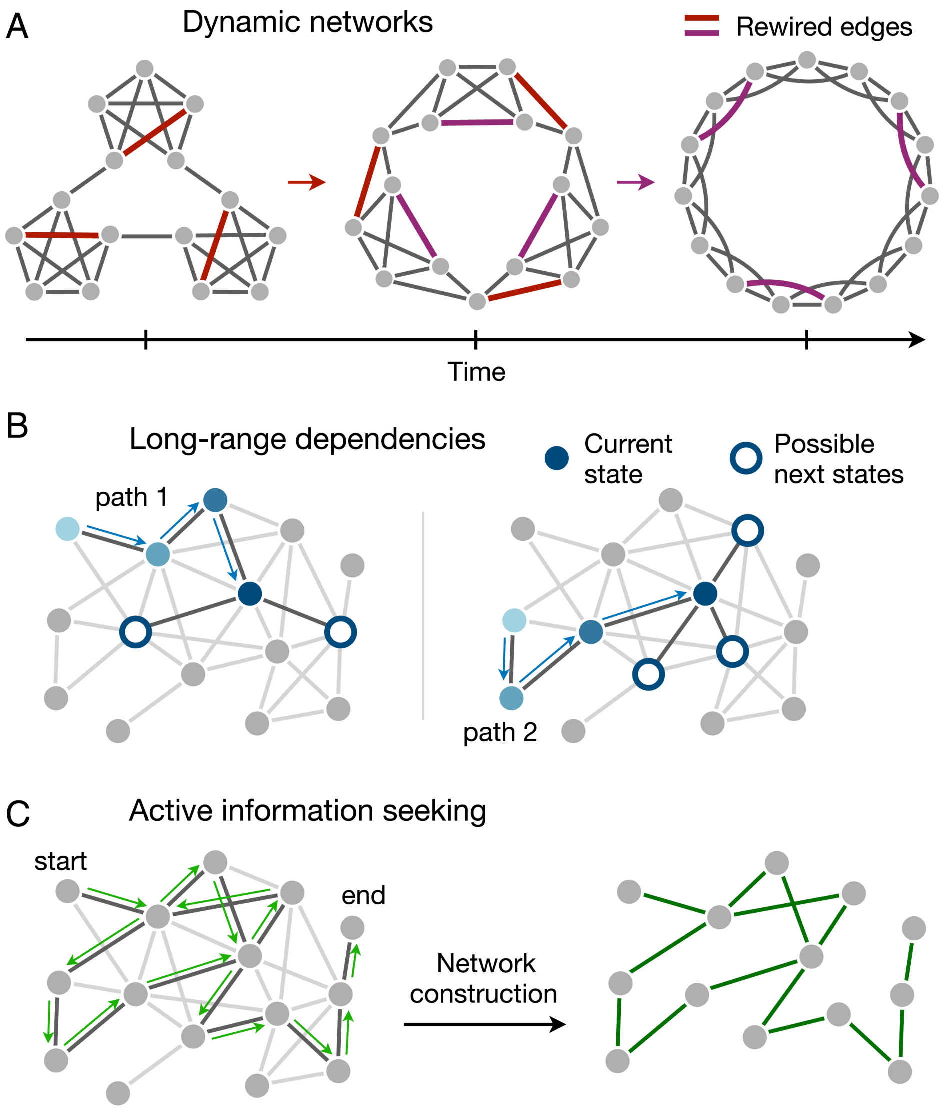

Figure 4.4: Dynamics of real-world graphs
Humans are able to detect and encode different dynamical features of an expanding network (A) over time, as well as notice the dependencies between sequences of stimuli (B). While only some information are recalled from memory, one would expect that humans do not bring back an entire network to perform a given task, but rather subgraphs with relevant information for a given task recalled thanks to active information seeking process (C). Adapted from Lynn and Bassett, 2020.

As an example of a corresponding real-world phenomenon, consider London taxi drivers. They are famously known for their outstanding spatial memory and the ability to quickly plan new routes with the shortest path within an entire network of runs. After a long period of training that enables them to get the license, they have to learn about 26 thousand roads and other thousands of points of interest around London with an emphasis on important touristic spots (Griesbauer et al., 2022). At the beginning of the training stage, future taxi drivers are not expected to learn a topographical map, but rather how specific areas locally link to each other. Acquiring the topological structure of the town is meant to avoid misconceptions about the city's geography that could lead to mistakes in route planning. In such a case, the city network has to be captured by a directed graph, because of a one-way road system or other restrictions. It is easy to think about such representation (also on a mental level) because we can encode all relevant information in a much-simplified version of the real-world problem. Although, even such a form of cognitive graph expands to a much more complicated complex network with time and experience. Within its structure, we could distinguish a more regular network of streets that can be easier to learn in a shorter period of time, but also more complex and irregular ones that need practice in order to be learned. However, after repeated exposure to the road configurations, the taxi drivers become fluent and find shortcuts with fewer errors.

Figure 4.5: Visualization of part of London routes network
Due to the relatedness of specific locations, all together they create a network within which we can observe the variety of complexity and density of routes (a). In an area of center of London (c) we can observe visible irregularity (e), while closer to the border area of Six Mile Radius (b) we observe much simplex and regular structure (d). Adapted from Griesbauer et al., 2022.

Subdivision of the city into districts of clusters is also very helpful in the distinction of the entire structure. Thus, taxi drivers can infer that the shortest path between district X to Z leads through district Y, because X and Z are not directly connected. London taxi drivers are particularly interesting because they are obliged to learn highly specific knowledge about small districts of the city. While compared to other big cities, such as Paris or Madrid - drivers there are usually required only to learn the base network that covers the major part of the surrounding of the city center (Griesbauer et al., 2022). These spatial abilities are very helpful in reasoning and route planning, but these elaborate cognitive maps in humans are being replaced by simple-to-acquire GPS. It seems that our flexibility in interacting with various environments can be lost out.

4.3 Encoding topological knowledge in the brain

The topological model of spatial learning rests on the insight that the hippocampus produces a topological representation of spatial environments filled with geometric details encoded by the brain. Spatial knowledge contains local information produced by a variety of information; e.g. location from place cells or spatial layout from grid cells. A graph structure can store a large-scale map of the environment created out of sequential information. The big advantage of such representation is its stability when the metrical features of the agent movements change (Dabaghian et al., 2014).

However, how such a map representation of the external world could be encoded with the emphasis on relationships between entities in the observed environment? Based on a spectrum of cell types observed experimentally during navigation, a *conjunctive encoding theory of cognitive map construction* from geometric perception is postulated by Zeng, Si, and Feng, 2022. According to the theory, different brain cells encode geometrical information of important environmental entities (e.g., boundaries, centers, landmarks of the environment) and constitute computational building blocks in transforming the egocentric information sensed by an animal into allocentric cognitive map representations. It is worthwhile to note that such a model for generous representation has not yet been found. Although, it has been supported by artificial experiments conducted with real neural signal data from specific cell types characterized by their peculiar activation patterns. The theory focuses on boundary cells encoding relative distance metric to animal's self-position and center bearing cell model vectors (LaChance et al., 2019) with information about relative self-position to the geometric center. By the subtraction of boundary cells vectors $\vec{EP}$ and center-bearing cells model vectors $\vec{EO}$, authors define a new cell type called geometric cells referred to as $\vec{OP}$.

Geometry cells aim to encode the geometry of the space relative to the center of the environment, independently of an animal's position and head rotation. The authors assume that this type of cells should be present in the postrhinal cortex (POR) that was found to support the transformation of egocentric representation into an allocentric spatial map (LaChance et al., 2019). This theoretical investigation of geometry cells aims to holistically characterize the computational mechanisms that support the construction of cognitive map representation in a topological manner.

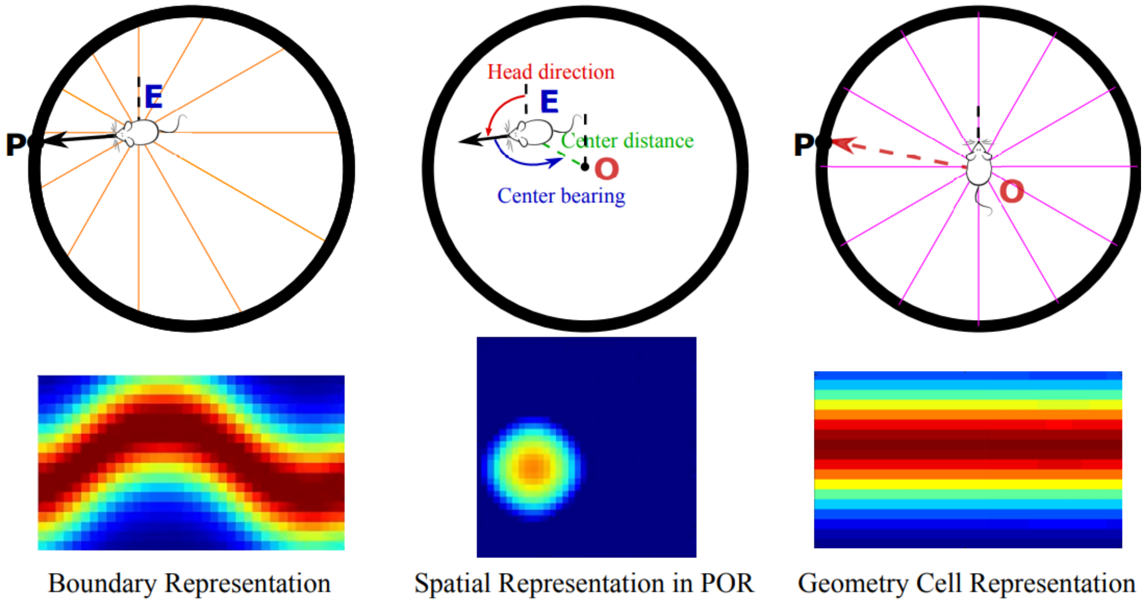

Figure 4.6: Formation of geometry cells representation
Geometry cell representation (right) emerges from the subtraction of vectors of boundary representation (left) and center-bearing spatial representation vectors in POR (middle). The conjunctive center bearing cell model proposed by LaChance et al., 2019 encodes the egocentric bearing and the animal's allocentric head direction, as well as the distance of the geometric center. Adapted from Zeng, Si, and Feng, 2022.

Each geometry unit of the cell can be referred to by its coordinates (θ, r), where θ corresponds to allocentric head direction and r to the distance to the boundary. The population firing pattern activity of geometry cells that characterizes the local space representation depends on the orientation, distance traveled, and head rotation. The obtained geometrical activity provides pure allocentric representations of high-level scenes and supports the quick formation of a cognitive map capturing the spatial structure of complex environments. These ways of transforming information from specific cell activity throughout the course of navigation support topological decoding of the underlying layout of various environments. Nevertheless, some metric information is still stored and used to encode the self-location of the animal in a polar coordinate system.

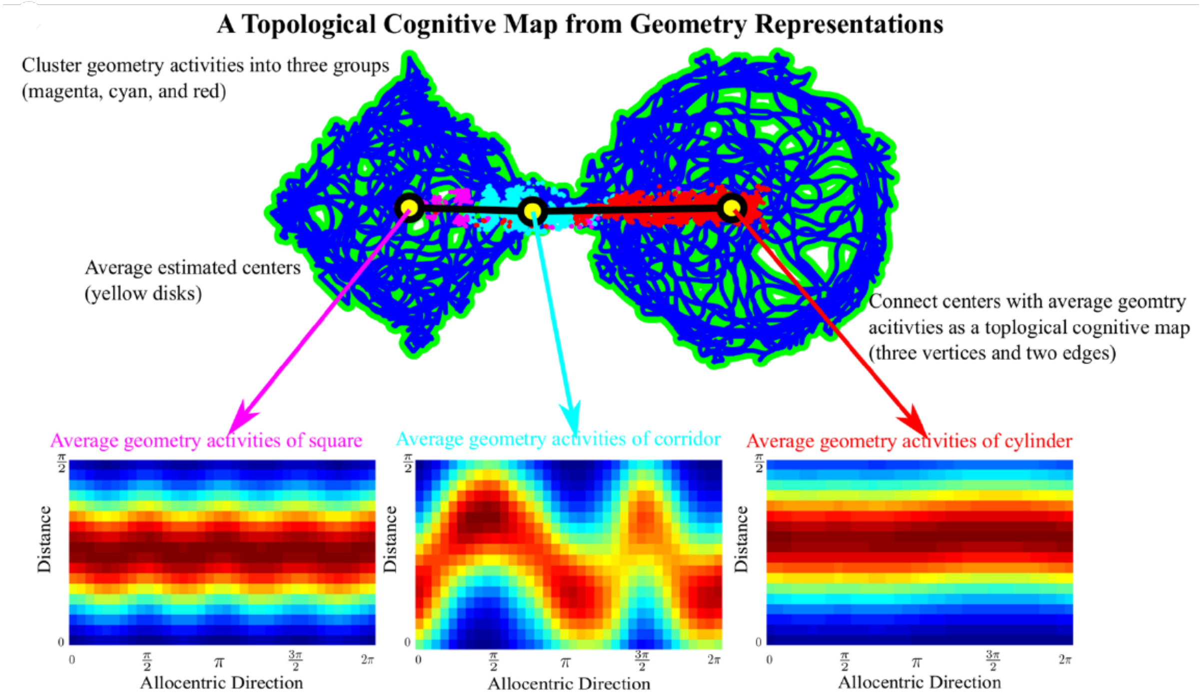

Figure 4.7: Compact topological cognitive map capturing spatial layout
In convex mazes of different shapes, the population activities of geometry units express uniform firing activity patterns reflecting layouts of the mazes. The neural activity of geometry units depends on the size and shape of the environment, e.g. concave polygon, circle, etc., that cause the spatial tuning of geometry cells during the navigation. This kind of change in neural activity could support the topological representation of the environment that is composed of multiple subregions of a different shapes. Adapted from Zeng, Si, and Feng, 2022.

However, the model proposed by Zeng, Si, and Feng, 2022 is not excluding the possibility of the co-existence of multiple processes operating on many levels with various cell representations. The remaining question is, how such parallel streams of spatial information processing could cooperate? Grid cells are known for their regular hexagonal firing patterns spanning the whole environment and providing a universal spatial metric system that is elastic and adaptive to local sensory cues (Rosay et al., 2019). They allow the identification of an animal's position in space by storing and integrating information about location, distance, and direction. However, the goal of such a navigation system is to create a globally coherent map that would provide mental representation covering environment structure, similarly to the model described by Zeng, Si, and Feng, 2022. Recent experiments have shown that regular and coherent grid firing patterns support navigation (Bush et al., 2015), where particular activations are self-correcting themselves throughout the exploration (Carpenter et al., 2015). Thus, it seems like a biological answer to the exploration-exploitation dilemma that aims to build a globally consistent cognitive map that will enable an agent to find the best route (Peterson and Verstynen, 2021). Nevertheless, the construction of such representation may require long-term exploration and getting feedback from other brain structures, e.g. hippocampus may play the role of map calibration correction mechanism (Zeng, Si, and Li, 2022). When such a globally coherent grid cells firing pattern will be put together with encoding geometry metrics of spatial locations, the unique and complex universal cognitive map could emerge. Such a globally coherent representation was proposed to exist due to the path integration system determining the structure of the cognitive map (stored in a graph) through calibration when an agent is revisiting locations. In the very first stage of exploration of new surroundings, we get a lot of noisy information which has to be filtered out and organized based on its importance for a given task. The interplay between an active process of navigation and long-term cognitive map representation plays an important role in encoding the dynamics and relevance of spatial information. Globally stable patterns capturing spatial layout enable future usage of information in an optimized manner for planning actions in the future. Such a process is similar to a well-known mechanism of hippocampal replay that involves recalling previous experiences to consolidate memories from the past (Derdikman and Moser, 2010).

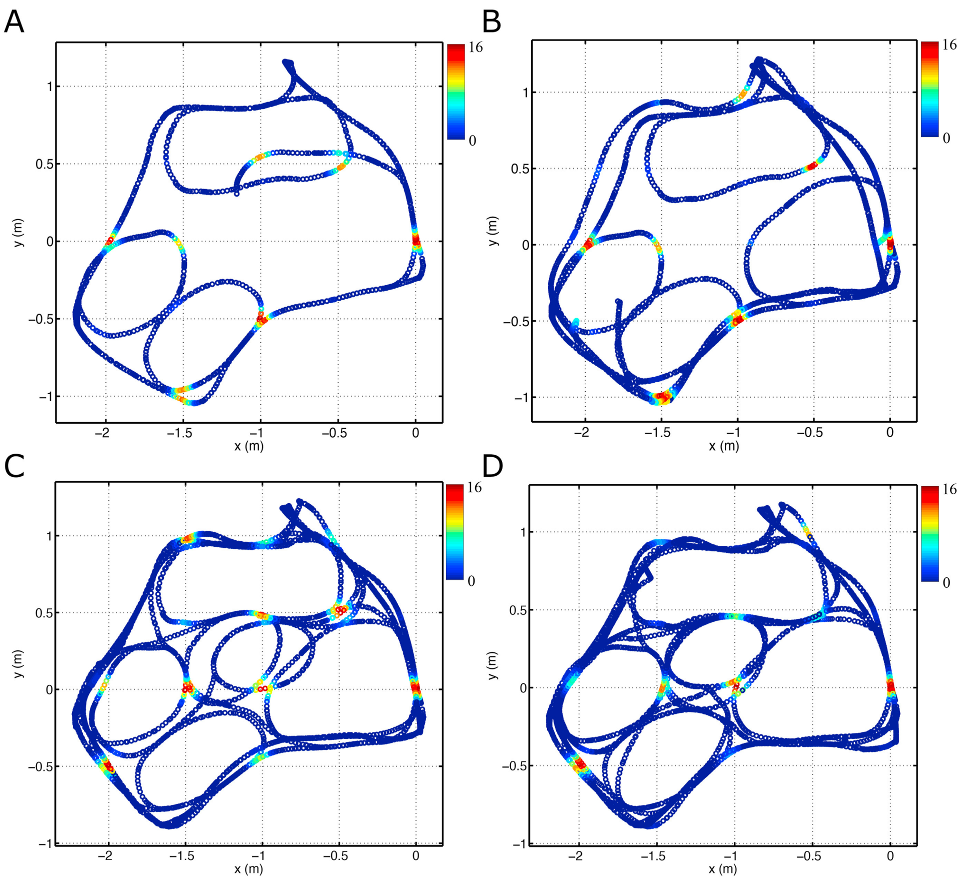

Figure 4.8: Development of globally coherent grid representation
Firing map of one grid cell for following time intervals of the exploration (A-D). At a certain exploration stage, a coherent map capturing the entire structure is created and it can further form a cognitive map through which we can plan the most optimal route to the reward. Adapted from Zeng, Si, and Li, 2022.

A highly organized network structure of the brain is based on the wiring architecture developed to process information efficiently, leading to subsequent cognition (McCulloch, 1944). Thus, the topological organization serves not only as an arbitrary way of organizing the information but also as a complex computational mechanism that helps to handle high-dimensional problems.

4.4 Artificial cognitive graph

The approach that enables us to test how such a topological cognitive map could be created online involves artificial agency. Navigation of a RL agent enable assessing and expanding their cognitive map of the maze in which sequences depends only on the current stimulus according to Markov property. The entire problem of capturing information in a form of mental representation can be compared to encoding low-dimensional vectors that summarize sensory observations and transform them into a graph. Preserved graph structure may be projected into a latent space, where geometric relations correspond to these from the real environment (Hamilton, 2020). Such an approach constitutes the encoder-decoder framework that is based on two key components, in which the first of them is an *encoder* that maps each node in the graph to a low-dimensional vector representation. While the second component, a *decoder*, uses this representation to reconstruct necessary information for each node's neighborhood based on the topological arrangement of the real environment. Such architecture is suitable for sequence learning problems such as learning a cognitive map.

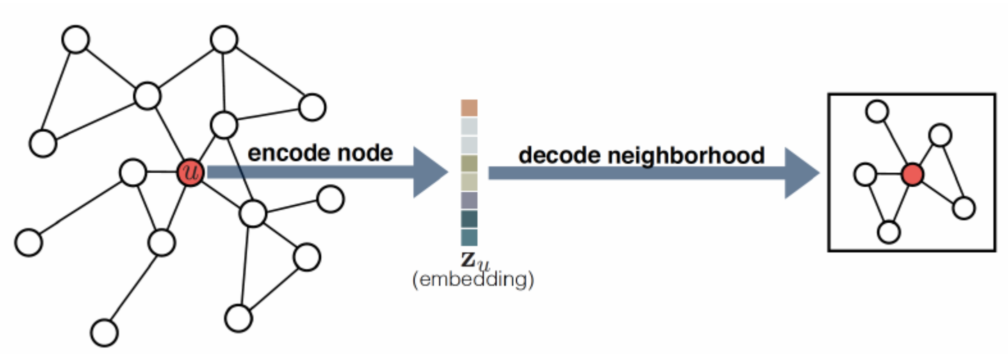

Figure 4.9: Encoder-decoder approach of graph representation learning
The encoder maps all nodes u to a low-dimensional vector embedding z_u that can be used by the decoder to reconstruct certain graph statistics or its local structures according to specific goals and situational circumstances. Adapted from Hamilton, 2020.

We would aim for the reconstruction of the relationship between nodes and vector embeddings with a minimal reconstruction loss. However, we cannot expect the encoded structure to be rich enough after the very first exposition to a novel environment. It becomes more and more coherent and detailed throughout our experience enabling us to become familiarized with its spatial layout. Formally, a computational framework that would encode the graph structure of a given high-dimensional environment will differ based on the underlying topological structure. Some examples are presented in Figure 4.10, where different mazes are presented with corresponding graph representation. These graphs could potentially stand for the mental representation of specific spatial layouts even for different scales of given environments. Based on these graph inferences, agents would be able to plan movement trajectories and navigate to goal positions in an optimized way.

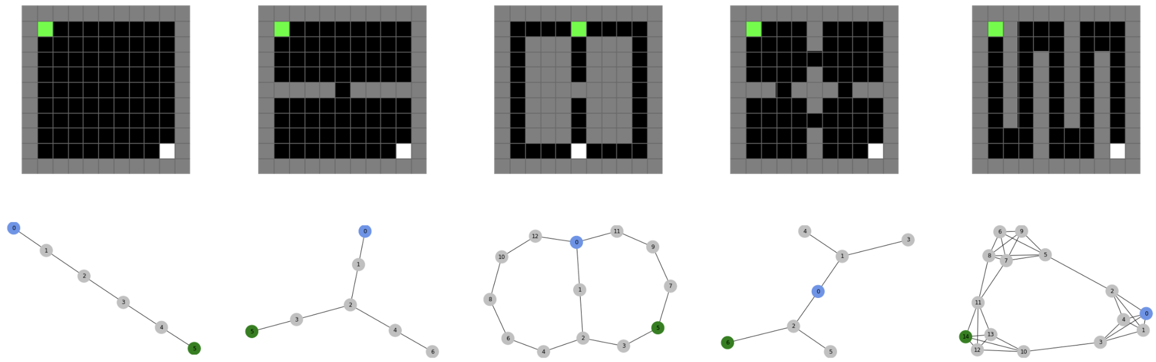

Figure 4.10: Different mazes with corresponding graph structure
A set of mazes inspired by experiments studying decision-making problems with corresponding grid (top row) and graph (bottom row) structure representation. Adapted from A. Juliani et al., 2022.

For the purpose of topological cognitive map inference, I will focus on model simulations made on geometry cells (Zeng, Si, and Feng, 2022) activity, where the artificial rodent-like agent was trained to navigate in a real-world environment. While a virtual animal was exploring, it calculated the environmental boundaries and head direction positions. By combining these two vectors, an agent was estimating the centers of the local fields used as a reference point for the creation of the nodes. Throughout the navigation, it inferred the topological map of the virtual office that contained only the centers of local spaces and their connections. Such a simplified representation described the most important information about relations to other places in the environment and thus, it was computationally efficient for action planning and generalization.

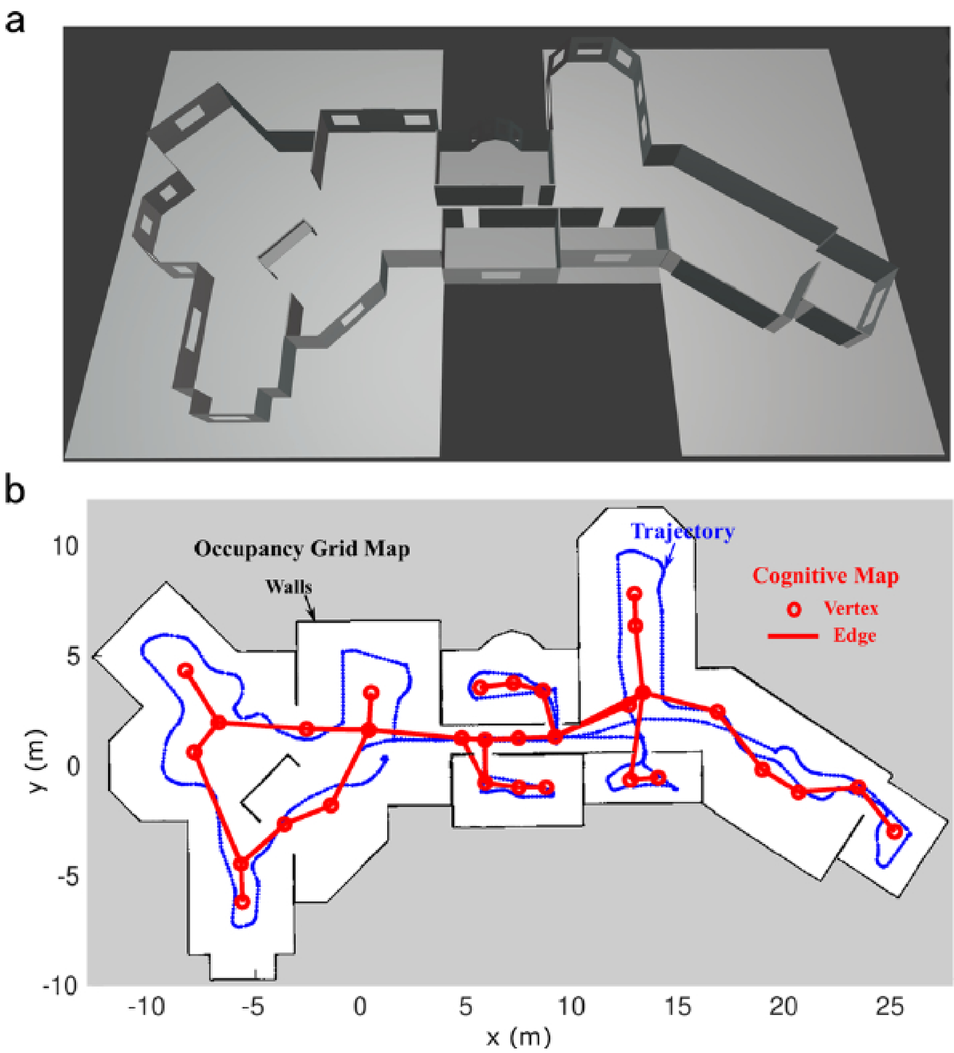

Figure 4.11: Spatial representation captured by a graph
(A) Realistic virtual office environment used for exploration by a virtual animal. (B) The corresponding topological cognitive map inferred by the animal through the exploration is nested inside a graph structure. Nodes represent latent state representation encoded in a sequential manner. Adapted from Zeng, Si, and Feng, 2022.

A new connection could be established if the distance from the features of the current local field to the features of the previous one was larger than a certain threshold. In Figure 4.11, a cognitive map is constituted from 31 nodes and 33 edges, obtained from the actual trajectories of the animal agent that characterize the local geometry of the environment. Such representation is compact and sparse since the number of nodes is just a piece out of the 1491 points traveled by the agent in this task.

It is possible that the map of the environment grows over time or changes in the connections with other nodes. The global optimization of the cognitive map is crucial for the purpose of resizing and properly encoding its structure. Memory replay (rethinking about encountered events) of this information enables the map learning to reorganize and remain informative.

4.5 Cognitive social topologies

I have already mentioned the importance of the topological structure of the environment in a navigational task. Through exploration, we are able to infer such information and create a coherent mental representation. However, what if we encounter a similar structure in another domain, such as connections between people in a social network? A social network is a theoretical construct that captures the relations between people, organizations, or entire societies that can be characterized by some features. As an example of a highly socialized species, it can be assumed that our brain established some similar ways to deal with such organized structures. Since such networks may be very large and it is impossible to track every relation by hand, our brain should be able to infer only some information unrevealing the most important characteristics. We know that people are able to create cognitive maps about arbitrary, abstract phenomena (Section 2.4), but how does it apply to the social domain?

The study of dynamics of cognitive social structures (CSS) with various topologies have been interesting for many scientists and economists for years (Brands, 2013). Various social networks were examined to find the impact of CSS on the level of an individual, as well as the entire group dynamics. One of the very basic organizational structures that can be observed in multiple societies is a hierarchy, where individuals are arranged according to their status. Usually, such a structure emerges quickly without a need for external integration due to an innate preference for hierarchical relationships (Zitek and Tiedens, 2012). People find it more intuitive to adapt themselves to a hierarchical network, where particular power dynamics are defined. Experiments have shown that even the human brain encodes the hierarchical status of a given person within a known network (Parkinson et al., 2017). Moreover, individuals with higher status tended to accurately predict the assessment of other people from the hierarchy. Such a hierarchical division of a network seems to be very important for many environmental constraints because it enables rapid adaptation to novel situations. Another important structural characteristic of a network is its clustered organization that divides a network into subregions. One study has created two environments for people coming to the laboratory that were asked to study some pieces of information and then discuss them with each other during the dyadic conversational phase (Coman et al., 2016). These environments were arranged to separate individuals with a different degree, whereas in a clustered condition the degree of separation was bigger than in a non-clustered one. The results have shown that different network topological structures impacted the collective memories, such that in a condition with less clustering (smaller degree of separation between people) memories were recalled more often and thus, remembered better.

Beyond the level of an individual, some studies were interested in the phenomenon of collective intelligence that could possibly guide the human species towards the ability to solve diverse problems (Woolley et al., 2010). Some outcomes that emerge from the cooperation between members of a particular group are above the abilities of the most intelligent member. For the purpose of studying these dependencies, six-person collaboration networks were asked to discover different hierarchical levels of chemical substances in a virtual task (Derex and Boyd, 2016). These groups were working together simultaneously (in a full connectivity condition) or in pairs, where members of these pairs were exchanged twice throughout the experiment (partial connectivity condition). Such an experimental arrangement led to the possibility of investigating the role of the topological organization of networks in collective outcomes. The results have shown that participants working with each other separately achieved further levels of the hierarchical invention in a given task, while the members of the group collaborating with each other simultaneously were trapped in local optimum. It seems that the community can benefit from collaboration, but the composition and diversity within its structure can enhance expected outcomes.

Similar to changes in topological relationships of the environment we navigate in, people also seem to react to changes in relatedness to other structures such as CSS. However, it still does not explain the perception of a network that could create something like a social cognitive map. Although, we could assume that the underlying structure would be based on individuals with some affiliation interacting with each other.

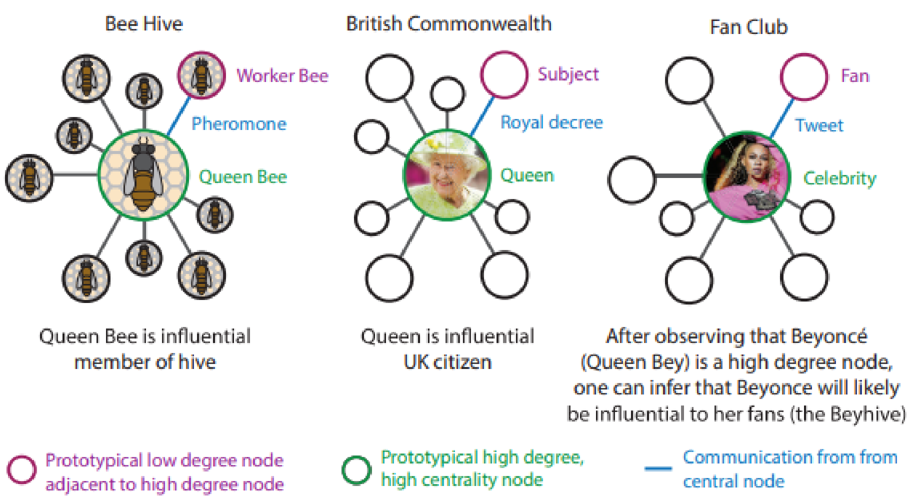

Figure 4.12: Example topologies of social networks
Social networks can represent not only the connections between individuals but also their role based on the influence relative to others (node degree refers to a number of connections to other nodes in the network). Adapted from Behrens et al., 2018.

Cognitive map concept applied for a social domain can thus be very powerful in shaping how people represent and reason about individuals from a social network. The dynamics between two people are sufficient to infer their roles, intentions, and relatedness. Additionally, the graph approach makes it very efficient to generalize structural analogies that can be found in other domains of cognition, such as navigation.

5 Summary

The ability to navigate within a structured environment has been an interest of neuroscientists both from human and animal perspective. The mechanisms involved in exploration behavior is not guided just by trial-and-error search, but also more sophisticated actions involving planning, motor control, imagination and knowledge generalization. The role of cognition in this process was defined by Tolman who proposed cognitive map hypothesis. He noticed that rats involved in spatial exploration were able to infer an unified representation of the maze environment. He stated that such ability could not be reduced to simple associations between stimulus and reinforcement, because it occurred before the presentation of reward. This theory has received a strong support from experimental studies which identified the entire spectrum of brain cells active during the navigation. All of them seem to encode only part of the information unified in a hippocampal-entorhinal circuitry. Some of them (e.g., reward cells and boundary vector cells) fire only when particular object is encoded, while others (e.g., grid, head direction and place cells) adjust their firing pattern according to observed changes in the location or the environment.

The basic intuitions about such system are based on navigation through spatial environments and building map-based mental representations. Simple analogy to well-known topographic maps and characteristic firing pattern of grid cells supported the idea that cognitive map is based on Euclidean geometry. However, the assumption that our brain encoded the world within strict metric system seems inefficient. Cognitive maps could be encoded much more efficiently if they capture the relatedness between relevant locations in space. Such arrangement can be formalised with graph-theoretical approach, where cognitive map consist of transitions between nodes through the edges. Additionally, such representation would not be limited to the spatial task of navigation within an environment, but also within social network, language and similar.

This thesis contributes to recent scientific investigations in representationalism theory of perception, where cognitive maps are believed to form mental representation of various environments in which an agent is navigating. The main goal is to find the best route through exploration of the environment. The entire process can be formalised with reinforcement learning, self-teaching machine learning algorithm that looks for an optimal solution. Despite the behavioral approach, the underlying cognitive map representation is being learned. I tried to prove that cognitive map is constructed from relational knowledge between entities encoded by brain cells involved in navigation. These assumptions are put together with the help of graph learning with an emphasis on spatial cognition. However, the theory finds support in computational neuroscience which points at possible brain mechanisms encoding information in a topological manner.

It is worth mentioning that I focused mostly on the limited role of hippocampus and entorhinal cortex, but nothing in the brain operate in isolation. This circuitry is just a part of bigger neuronal network which supports behavior and cognition through a neuronal paths processing various information. Also, our cognition can be influenced by different factors, such as concentration of neuromodulators in particular brain areas. The main limitation in cognitive map research is caused by unprecise neuroimaging techniques which does not clearly encode navigational mechanisms. Most of the evidences come from multi-site recordings of neuronal populations in simpler species, such as rodents. To investigate human spatial navigation, scientist rely mostly on computational models based on assumptions coming from animal data. As long as we will only approximate the role of spatial cognition, we will not find any consensus about cognitive mapping phenomenon.

Mental mapping of the environment is an important ability that enable to perform and plan autonomous actions, as well as behave flexibly. Graph learning field applied for this cognitive domain captures complex information in a simple format which can guide us during real-time experience. The unique ability to orient ourselves in complex topological structures can be then applied for various non-spatial domains.

6 Bibliography

Anselmi, F., Murray, M. M., & Franceschiello, B. (2020). A computational model for grid maps in neural populations. *Journal of computational neuroscience*, *48*(2), 149–159.

Arias-Carrión, O., Stamelou, M., Murillo-Rodrıéguez, E., Menéndez-González, M., & Pöppel, E. (2010). Dopaminergic reward system: A short integrative review. *International archives of medicine*, *3*(1), 1–6.

Asem, J. S., & Fortin, N. J. (2017). Memory for space, time, and episodes.

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences*, *107*(7), 3228–3233.

Barry, C., Hayman, R., Burgess, N., & Jeffery, K. J. (2007). Experience-dependent rescaling of entorhinal grids. *Nature neuroscience*, *10*(6), 682–684.

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509.

Bellman, R. (1966). Dynamic programming. *Science*, *153*(3731), 34–37.

Blodgett, H. C. (1929). The effect of the introduction of reward upon the maze performance of rats. *University of California publications in psychology.*

Boguslawski, P., Gold, C. M., & Ledoux, H. (2011). Modelling and analysing 3d buildings with a primal/dual data structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(2), 188–197.

Botvinick, M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130480.

Brands, R. A. (2013). Cognitive social structures in social network research: A review. *Journal of Organizational Behavior*, *34*(S1), S82–S103.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, *10*(3), 186–198.

Burgess, N., & O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus*, *6*(6), 749–762.

Bush, D., Barry, C., Manson, D., & Burgess, N. (2015). Using grid cells for navigation. *Neuron*, *87*(3), 507–520.

Buzsáki, G., & Tingley, D. (2018). Space and time: The hippocampus as a sequence generator. *Trends in cognitive sciences*, *22*(10), 853–869.

Buzsáki, G., McKenzie, S., & Davachi, L. (2022). Neurophysiology of remembering. *Annual review of psychology*, *73*, 187–215.

Carpenter, F., Manson, D., Jeffery, K., Burgess, N., & Barry, C. (2015). Grid cells form a global representation of connected environments. *Current Biology*, *25*(9), 1176–1182.

Collett, T. S., & Graham, P. (2004). Animal navigation: Path integration, visual landmarks and cognitive maps. *Current Biology*, *14*(12), R475–R477.

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, *113*(29), 8171–8176.

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468.

Dabaghian, Y., Brandt, V. L., & Frank, L. M. (2014). Reconceiving the hippocampal map as a topological template. *Elife*, *3*, e03476.

Daikoku, T., & Yumoto, M. (2017). Single, but not dual, attention facilitates statistical learning of two concurrent auditory sequences. *Scientific Reports*, *7*(1), 1–10.

Dang, S., Wu, Y., Yan, R., & Tang, H. (2021). Why grid cells function as a metric for space. *Neural Networks*, *142*, 128–137.

Danjo, T., Toyoizumi, T., & Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. *Science*, *359*(6372), 213–218.

Darwin, C. (1873). Origin of certain instincts. *Nature*, *7*, 417–418.

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences.*

Derdikman, D., & Moser, M.-B. (2010). A dual role for hippocampal replay. *Neuron*, *65*(5), 582–584.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987.

Di Felice, P., Clementini, E., Hinterberger, H., Domingo-Ferrer, J., Kashyap, V., Khatri, V., Snodgrass, R., Terenziani, P., Koubarakis, M., Zhang, Y., et al. (2009). Topological relationships.

Dijkstra, E. W. et al. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, *1*(1), 269–271.

Gallistel, C. R. (1990). *The organization of learning.* The MIT Press.

Gauthier, J. L., & Tank, D. W. (2018). A dedicated population for reward coding in the hippocampus. *Neuron*, *99*(1), 179–193.

Gorchetchnikov, A., & Grossberg, S. (2007). Space, time and learning in the hippocampus: How fine spatial and temporal scales are expanded into population codes for behavioral control. *Neural Networks*, *20*(2), 182–193.

Griesbauer, E.-M., Manley, E., Wiener, J. M., & Spiers, H. J. (2022). London taxi drivers: A review of neurocognitive studies and an exploration of how they build their cognitive map of london. *Hippocampus*, *32*(1), 3–20.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806.

Hamilton, W. L. (2020). Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning*, *14*(3), 1–159.

Heilbron, M., & Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS computational biology*, *15*(4), e1006972.

Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., & Griffiths, T. L. (2022). People construct simplified mental representations to plan. *Nature*, 1–8.

Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M.-B., & Moser, E. I. (2019). Object-vector coding in the medial entorhinal cortex. *Nature*, *568*(7752), 400–404.

James, W., Burkhardt, F., Bowers, F., & Skrupskelis, I. K. (1890). *The principles of psychology* (Vol. 1). Macmillan London.

Juliani, A., Barnett, S., Davis, B., Sereno, M., & Momennejad, I. (2022). Neuro-nav: A library for neurally-plausible reinforcement learning. *The 5th Multidisciplinary Conference on Reinforcement Learning and Decision Making.*

Juliani, A. W. (2020). *Learning and acting with predictive cognitive maps* (Doctoral dissertation). University of Oregon.

Kalm, K., Davis, M. H., & Norris, D. (2013). Individual sequence representations in the medial temporal lobe. *Journal of cognitive neuroscience*, *25*(7), 1111–1121.

Karuza, E. A., Kahn, A. E., Thompson-Schill, S. L., & Bassett, D. S. (2017). Process reveals structure: How a network is traversed mediates expectations about its architecture. *Scientific reports*, *7*(1), 1–9.

Keane, M. M., Bousquet, K., Wank, A., & Verfaellie, M. (2020). Relational processing in the semantic domain is impaired in medial temporal lobe amnesia. *Journal of neuropsychology*, *14*(3), 416–430.

LaChance, P. A., Todd, T. P., & Taube, J. S. (2019). A sense of space in postrhinal cortex. *Science*, *365*(6449), eaax4192.

Latuske, P., Kornienko, O., Kohler, L., & Allen, K. (2018). Hippocampal remapping and its entorhinal origin. *Frontiers in behavioral neuroscience*, *11*, 253.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, *29*(31), 9771–9777.

Ludwig, T., Wu, C. M., & Schulz, E. (2022). Connecting exploration, generalization, and planning in correlated trees. *City*, *1*, 9.

Lynn, C. W., & Bassett, D. S. (2020). How humans learn and represent networks. *Proceedings of the National Academy of Sciences*, *117*(47), 29407–29415.

Lynn, C. W., Kahn, A. E., Nyema, N., & Bassett, D. S. (2020). Abstract representations of events arise from mental errors in learning and memory. *Nature communications*, *11*(1), 1–12.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, *97*(8), 4398–4403.

McCulloch, W. S. (1944). The functional organization of the cerebral cortex. *Physiological reviews*, *24*(3), 390–407.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *International conference on machine learning*, 1928–1937.

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, *31*, 69–89.

Neumann, J. v. (1966). Theory of self-reproducing automata. *Edited by Arthur W. Burks*.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford university press.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, *3*(3), 317–330.

Omer, D. B., Maimon, S. R., Las, L., & Ulanovsky, N. (2018). Social place-cells in the bat hippocampus. *Science*, *359*(6372), 218–224.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2017). Spontaneous neural encoding of social network position. *Nature Human Behaviour*, *1*(5), 1–7.

Peterson, E. J., & Verstynen, T. D. (2021). Curiosity eliminates the exploration-exploitation dilemma. *bioRxiv*, 671362.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, 64–99.

Rosay, S., Weber, S., & Mulas, M. (2019). Modeling grid fields instead of modeling grid cells. *Journal of Computational Neuroscience*, *47*(1), 43–60.

Ross, R. S., & Eichenbaum, H. (2006). Dynamics of hippocampal and cortical activation during consolidation of a nonspatial memory. *Journal of Neuroscience*, *26*(18), 4852–4859.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS computational biology*, *13*(9), e1005768.

Sarel, A., Finkelstein, A., Las, L., & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, *355*(6321), 176–180.

Sasakura, H., & Mori, I. (2013). Behavioral plasticity, learning, and memory in c. elegans. *Current opinion in neurobiology*, *23*(1), 92–99.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature neuroscience*, *16*(4), 486–492.

Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath, C. (2015). Memory and space: Towards an understanding of the cognitive map. *Journal of Neuroscience*, *35*(41), 13904–13911.

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, *91*(6), 1402–1412.

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current opinion in neurobiology*, *7*(2), 191–197.

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current opinion in neurobiology*, *55*, 7–14.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.

Sharma, S., Curtis, A., Kryven, M., Tenenbaum, J., & Fiete, I. (2021). Map induction: Compositional spatial submap learning for efficient exploration in novel environments. *arXiv preprint arXiv:2110.12301*.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of experimental psychology*, *55*(6), 509.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, 103535.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, *20*(11), 1643–1653.

Strange, K. (2006). An overview of c. elegans biology. *C. elegans*, 1–11.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, *10*(2), 420–435.

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A map for social navigation in the human brain. *Neuron*, *87*(1), 231–243.

Teki, S., Kumar, S., von Kriegstein, K., Stewart, L., Lyness, C. R., Moore, B. C., Capleton, B., & Griffiths, T. D. (2012). Navigating the auditory scene: An expert role for the hippocampus. *Journal of neuroscience*, *32*(35), 12251–12257.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

Tort, A. B., Komorowski, R. W., Manns, J. R., Kopell, N. J., & Eichenbaum, H. (2009). Theta–gamma coupling increases during the learning of item–context associations. *Proceedings of the National Academy of Sciences*, *106*(49), 20942–20947.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology*, *53*(1), 1–25.

Watson, J. B. (1930). Behaviorism.

Wehner, R., & Rössler, W. (2013). Bounded plasticity in the desert ant's navigational tool kit. *Handbook of behavioral neuroscience* (pp. 514–529). Elsevier.

Woollett, K., & Maguire, E. A. (2012). Exploring anterograde associative memory in london taxi drivers. *Neuroreport*, *23*(15), 885.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, *330*(6004), 686–688.

Wu, C. M., Schulz, E., Garvert, M. M., Meder, B., & Schuck, N. W. (2020). Similarities and differences in spatial and non-spatial cognitive maps. *PLoS computational biology*, *16*(9), e1008149.

Yu, Y., Wang, T., & Liew, S. C. (2019). Deep-reinforcement learning multiple access for heterogeneous wireless networks. *IEEE Journal on Selected Areas in Communications*, *37*(6), 1277–1290.

Zeng, T., Si, B., & Feng, J. (2022). A theory of geometry representations for spatial navigation. *Progress in Neurobiology*, 102228.

Zeng, T., Si, B., & Li, X. (2022). Entorhinal-hippocampal interactions lead to globally coherent representations of space. *Current Research in Neurobiology*, 100035.

Zitek, E. M., & Tiedens, L. Z. (2012). The fluency of social hierarchy: The ease with which hierarchical relationships are seen, remembered, learned, and liked. *Journal of personality and social psychology*, *102*(1), 98.

Zuo, X. (2018). Mazelab: A customizable framework to create maze and gridworld environments.

7 Appendix

List of Figures