AMERICAN PSYCHOLOGICAL ASSOCIATION

Decision

ISSN: 2325-9965

2024, Vol. 11, No. 4, 537–556
https://doi.org/10.1037/dec0000249

Cognitive Graphs: Representational Substrates for Planning

Jungsun Yoo[1], Elizabeth R. Chrastil[1, 2, 3], and Aaron M. Bornstein[1, 2]

[1] Department of Cognitive Sciences, University of California, Irvine
[2] Center for the Neurobiology of Learning and Memory, University of California, Irvine
[3] Department of Neurobiology and Behavior, University of California, Irvine

Making plans for upcoming actions is a computationally demanding process. To mitigate these demands, individuals can build extensive internal models of their environment—states, actions, and their sequential relationships—that allow for plans to be developed with minimal computational costs. Initially, these models reflect elaborate networks of learned associative relationships, which can be used to generate plans for reward through more iterative computations such as trajectory sampling. After sufficient experience, compressed forms of these models can efficiently capture long-range sequential structure, allowing them to be used for rapid planning even in pursuit of novel or changing rewards. Here, we review recent work on the multitude of representations that can support different forms of planning. We discuss how *cognitive graphs*, a framework with roots in both cognitive psychology and computer science, can provide a unifying view of these representations and their relationships to one another. Conceptualizing internal models as forms of graphs situates them on a spectrum where different kinds of structured sequences can be queried to support both planning and the formation of iteratively more compressed predictive representations. We discuss how each of these kinds of cognitive graphs are created during learning, and used to transfer and generalize knowledge across environments. Taken together, this review highlights the significant impact that the various associative structures of memory have on planning.

Keywords: cognitive graph, planning, reinforcement learning

Planning is a common, and complex, form of decision making. It requires both representing actions, along with their precedents and consequences, and sequencing them appropriately. This process consists of *offline* stages, during which predictive representations of the environment are formed and refined, and *online* stages, where extant representations of relevant past experience are interrogated, and their predictions for the outcomes of planned choices are arbitrated.

The term "representation" used here refers to how elements of a given decision problem are encoded in memory and associated with each other. Choices are strongly influenced by the format in which the decision elements are arranged when presented to an individual—for instance, risk attitudes often vary considerably when options are presented as explicit frequencies, rather than summary probabilities (Kahneman & Tversky, 1979). More recently, researchers have begun to systematically explore

This article was published Online First August 29, 2024.
Jungsun Yoo https://orcid.org/0000-0001-8341-9226
Elizabeth R. Chrastil https://orcid.org/0000-0003-2544-0152
Aaron M. Bornstein https://orcid.org/0000-0001-6251-6000
The authors have no conflicts of interest to disclose. The authors thank Bruce McNaughton and Jeffrey L. Krichmar for insightful discussions. This work was supported by the National Institute of Aging (Grant R21AG072673) to Aaron M. Bornstein and the National Institute of Neurological Disorders and Stroke (Grant R01NS119468) to Elizabeth R. Chrastil.
Jungsun Yoo played a lead role in investigation and writing–original draft and an equal role in writing–review and editing. Elizabeth R. Chrastil played a supporting role in conceptualization, project administration, supervision, writing–original draft, and writing–review and editing. Aaron M. Bornstein played a supporting role in conceptualization, project, writing–original draft, and writing–review and editing and a lead role in project administration and supervision.
Correspondence concerning this article should be addressed to Jungsun Yoo, Department of Cognitive Sciences, University of California, Irvine, 3151 Social Science Plaza A, Irvine, CA 92697-5100, United States. Email: jungsuy@uci.edu

how choices depend on the ways in which decision-relevant information, such as state spaces, are represented *internally* by the decision-maker (Doya et al., 2002; S. Wang et al., 2022). This work shows that the choice of internal representations can have a similarly dramatic influence on the outcome of a decision. For example, individuals who remember their local environment as a series of routes they have taken ("egocentric" representation) may be unlikely to try a novel route in the face of a detour, unlike individuals who have integrated their experiences to form a maplike ("allocentric") summary of the environment (Chrastil & Warren, 2014). This example highlights how internal, unlike external, representations, can be transformed from one format into another given sufficient experience and/or time; critically, this process can yield many intermediate formats, where some information is retained and other information is lost.

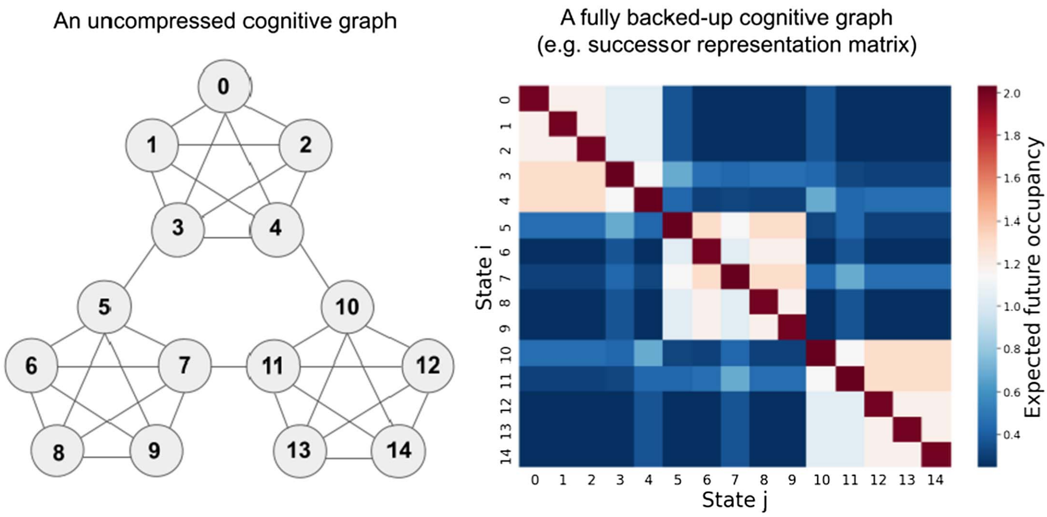
In the case of planning, representation format is critical in part because much planning occurs ahead of time, by constructing a semiflexible *policy* that establishes the rules by which sequences of actions are to be taken. In these cases, the selection of which *kind* of internal model provides the state space over which the policy is defined, and thus critically determines the actions ultimately taken (Ho et al., 2022). The importance of representational format in planning is further underscored by its role in *transfer learning*, which requires first identifying similar situations from the past and subsequently selecting the relevant aspects of that previously learned structure. When experience in related environments is extensive, allowing the agent to infer common latent structure, one could apply compact, "maplike" representations that allow for efficient planning with minimal error (Geerts et al., 2022; Whittington et al., 2020). However, as the overlap between well-learned settings and the current environment decreases, one must rely on approximations to identify relevant instances of previous experiences with the current or similar environments (Zhao et al., 2022). Internal simulations informed by these sorts of instance samples can be used for iterative, vicarious evaluation of decision problems that not only informs the decision at hand but allows the agent to accelerate the inference of a more general latent structure (George et al., 2021).

An implication of this *representation-centric* view of planning is that a key problem for agents to solve is how to summarize the available experience in a way that best supports efficient and effective planning and transfer learning. The type of summary representation best suited to each situation thus depends on the complexity of the environment, the amount of experience the agent has in it, and the time and computational resources available to evaluate candidate policies; these quantities are often dynamic or not known ahead of time, thus licensing the agent to maintain multiple representations that can be leveraged to different degrees in different settings (Doya et al., 2002; S. Wang et al., 2022).

We propose that these many distinct forms of internal representations—associative relationships—can be fruitfully understood as types of graphs (Butts, 2009). Here, environmental states are represented as nodes and the transitions between them are shown as various types of edges (Lynn et al., 2020; Schapiro et al., 2013), depending on the information available (Chrastil & Warren, 2014). The edges could be either unidirectional when describing causality or irreversible transitions or could be bidirectional when these conditions are not assumed. For instance, a decision tree is a specific example of a graph that encodes sequential, or unidirectional, relationships between states (Bertsekas, 2012). Formalizing these structures as graphs can allow researchers to formally connect seemingly disparate types of planning, to reason about their related algorithmic and implementational properties (Zhang et al., 2021), and to determine how and which information is transferred ("consolidated") from one format to another, for example, during sleep (Feld et al., 2022).

Graph-like forms are useful in planning range between extremes—at the one end, sets of instances of individual pairwise associations; at the other end, compact, long-range multistep contingencies—with many points along the spectrum between these (Chrastil & Warren, 2014, Figure 1). Recent work supports the simultaneous creation and updating of multiple graph-like knowledge structures in support of planning. These internal models are distinguished by their content, format, and also in what they entail for the dynamics of their learning and use in deliberative decision-making (Bornstein & Daw, 2012, 2013; Doya et al., 2002; Smith & Graybiel, 2013; Tambini et al., 2023). Below, we review findings that suggest that they influence the behavior in accordance with their suitability to the task at hand and that the apparent shift in behavioral control from one form to other is characterized by the transformation of information between representational formats, with attendant trade-offs in function and fidelity.

Figure 1
A Graphical Illustration of the Two Extremes of Representations as a Function of Compression: An Uncompressed Cognitive Graph (Left) and a Fully Backed-Up Cognitive Graph (Successor Representation Used as an Example Here; Right)



Note. Left: Numbers indicate labels of a node, or a discrete state, in a cognitive graph. Edges between two nodes depict the transition between two states. Right: A backed-up version of a cognitive graph that fully captures future trajectories from a given node (state). Row identification numbers (ID; *i*) indicate the current state, and column IDs (*j*) indicate the successor state. The values in the matrix represent the expected general future occupancy of *j* from *i* and are color-coded for visualization. Note that while future occupancy statistics preserve the coarse community structure, route information is diminished (e.g., the adjoining gateway nodes, such as 3 and 11 for the cluster of States 5–9, are only slightly distinguished). See the online article for the color version of this figure.

Theoretical (Weber & Johnson, 2006) and empirical (Otto et al., 2022; Palminteri et al., 2015; Wu et al., 2021) work supports the idea that the evaluation of an option depends in part on how that option is remembered—for instance, if it is remembered as part of a set of related options, with ranked preferences within that set (e.g., a favored restaurant among those of similar cuisine), or if its visible features are associated with other latent features (e.g., a food attribute linked to allergic reactions). Foundational work has demonstrated that successive memory retrievals are related to the underlying associative structure of memory (Howard & Kahana, 2002), supporting a form of trajectory sampling (Gershman & Daw, 2017; S. Wang et al., 2022), and that the content of extended memory retrieval at the time of choice has a meaningful influence on preferences (Bornstein & Norman, 2017). Taken together, this work supports a critical influence on the choice of the associative structure of memory. Therefore, it is important to understand the different forms this structure can take and to identify commonalities and points of divergence relevant to choice behavior.

Cognitive Graphs

These associative structures can be understood as forms of cognitive graphs, which range from "uncompressed" to "compressed" (Figure 1). The most uncompressed form, in which states are encoded as experienced sequences with minimal latent structure inference, conceptually aligns with previous articulations of "cognitive graphs" (Chrastil & Warren, 2014; George et al., 2021; Muller et al., 1996), and that is proposed to support types of model-based reinforcement learning (Daw et al., 2005; Gershman & Daw, 2017; Lengyel & Dayan, 2007). A cognitive graph can be characterized as a directed graph (Muller et al., 1996), with nodes representing states and edges indicating state transitions. These edges may be labeled, augmenting the topology with local metric information

(Chrastil & Warren, 2014; Warren, 2019). They may also be weighted, reflecting the transition probability between states (George et al., 2021; Natarajan & Kolobov, 2022; Sutton & Barto, 2018). A cognitive graph is formed through learning how different sequences of state transitions connect at intersections (Stiso et al., 2022), enabling agents to flexibly navigate conceptual and spatial networks by recombining the segments in novel ways (Mark et al., 2020; Peer et al., 2021; Warren, 2019). Additionally, their abstract nature supports counterfactual simulations and generalizations to novel environments, thereby accelerating the learning process (G. Zhu et al., 2020). Though the entire continuum of representations is graph-structured, we will, for clarity, refer to the most uncompressed extreme form as "full" or "flexible" graphs and the most compact representations as "backed-up" or "compressed" predictive representations.

At the other end of the spectrum, backed-up, predictive representations contain information that is fully predictive of the N-step consequence of taking a given action a in the current state s (Figure 1, right). To elaborate, a standard model of choice describes preferences between options as formed after a unitary expected value is computed by combining the reward distributions implied by each options' features (Rangel et al., 2008). These values—both unitary and the components—can be represented in different ways, each of which has different implications for the preference construction process. Backed-up representations enable fast, cheap evaluation of N-step plan outcomes, using an operation akin to matrix multiplication (though the neural instantiation of this process has yet to be fully described; Gershman, 2018, and may be approximated by sampling; Gershman et al., 2012). For example, model-free reinforcement learning of action values (Sutton & Barto, 2018) captures this unitary value as a recency-weighted average of the discounted total reward obtained in past episodes where the agent took the given action in the given state. Here, the outcome values of multistep actions are mediated by a discount factor, δ, applied at each update operation. An alternative approach to constructing unitary values is to use a backed-up representation of the discounted N-step state occupancy alone, irrespective of the reward obtained, which allows decoupling the environmental state dynamics, which may be more stable, from reward contingencies that may fluctuate more often or be entirely trial-unique. Such *successor representations* (or their mirror, predecessor representations; Jeong et al., 2022) compress occupancy of sequences following or preceding a given state (Dayan, 1993), which can be used to derive biological cell response types matching those observed in subfields of the hippocampal formation (Stachenfeld et al., 2017). There are several related formats that differ in what information is included in the backed-up representation, such as successor features (Barreto et al., 2017)—which generalize the state-space learning approach to a space over option dimensions (e.g., desirability for food)—and first-occupancy representations (Moskovitz et al., 2021) that only consider the first-time visits to each state. Inspired by the need to bridge the gap between behavioral economics and reinforcement learning, a novel state representation named λ representation (λR) incorporates the concept of diminishing marginal utility (Moss, 1984) into reinforcement learning by discounting multiple visits to a state. Using a discount parameter λ, this formalization puts successor representations ($\lambda = 1$, no discounting of multiple visits) and first-occupancy representations ($\lambda = 0$) in a continuum (Moskovitz et al., 2023).

Between the extremes of compressed versus flexible-model representations, the cognitive graphs with intermediate modes of approximation can also be identified. We described above how the discount factor allows the successor representation to be parametrically distinguished from the outcomes of Monte Carlo trajectory sampling from a full model. Another axis along which these representations can vary in their approximation of the full environment dynamics is the degree to which they reflect hierarchical structure. For example, agents may cluster or abstract related states as intermediate "subgoals" that exist in multiple levels hierarchically to plan efficiently (Noh et al., 2023; Tomov et al., 2020). Compression can also occur by compressing *actions or policies* into higher level actions, referred to as option or skill discovery (Sutton et al., 1999). Automated discovery of options at multiple levels has facilitated learning in artificial agents (R. Fox et al., 2017). Likewise, humans appear to adopt policy compression to balance cognitive costs and maximize reward (Lai et al., 2022; Lai & Gershman, 2021). Similar to this, extracting a causal relationship between events at various levels of granularity could be seen as an abstraction or compression of the environment (Kinney & Lombrozo, 2023a, 2023b).

In this article, we initially delve into the differences between planning predicated on the flexible recombination of action sequences and planning employing compressed representations. Subsequently, we propose cognitive graphs as a potentially unifying framework supporting both decisions based on sampling potential future sequences and decisions based on full long-run occupancy statistics. We discuss how these functions require key mechanisms—in particular, merging disjoint sequences and splitting aliased states—offered by some implementations of the cognitive graphs. We conclude with a discussion of further research directions, in particular understanding how the spectrum of forms of cognitive graph may support distinct control strategies. This, in turn, could potentially clarify the differential use of types of control in different stages of learning.

Planning as a Function of Representational Compactness

Planning in a Markov Decision Process

For simplicity, planning is often conceptualized within the context of a Markov decision process. In a classical Markov decision process, the environment in which an agent plan is characterized as a tuple of $\langle S, A, T, R, \pi, \gamma \rangle$, where S is a finite and discrete state space that comprised states, and A is a set of actions that can be executed in each state $s \in S$. γ, within the half-open interval of (0, 1], refers to the discount factor that represents how future rewards are valued in comparison to immediate rewards. The models consist of two functions, where $T(s, a, s')$ is the transition function for each $s \in S$ and $a \in A$, and $R(s, a, s')$ is a reward function that provides the immediate reward or value obtained after taking action a in state s and transitioning to state s'. π refers to the policy or the probability distribution over potential actions that the agent may choose to take at a certain state. We assume that an agent starts from an initial state s_0 and executes a sequence of possible actions in the successor states (s') up to a terminal or goal state $s \in S_G$. The agent's goal in planning is to learn and execute actions based on an optimal solution, or policy, that maximizes the cumulative value from an initial state to a goal state.

One of the most crucial components for successful planning is having an accurate internal model of the environment because the model is used for simulating or predicting behavior; inaccurate models could entail incorrect predictions and thereby result in a chain of suboptimal actions (Talvitie, 2017). It is also important to adopt the most suitable models for each specific context, given that the optimal type of model to use may vary depending on the relationship of the model to the environment—for instance, whether the model is known with certainty to correspond exactly to the environment (Jiang et al., 2015). Below, we delve into the kind of model utilization that may be optimal in scenarios where agents are still in the preliminary phases of environment interaction (the Planning With Uncompressed Representations: Sampling Instances section) or in circumstances where they possess sufficient experience for compression of representations to occur (the Planning With Backed-Up State/Action Sequences section).

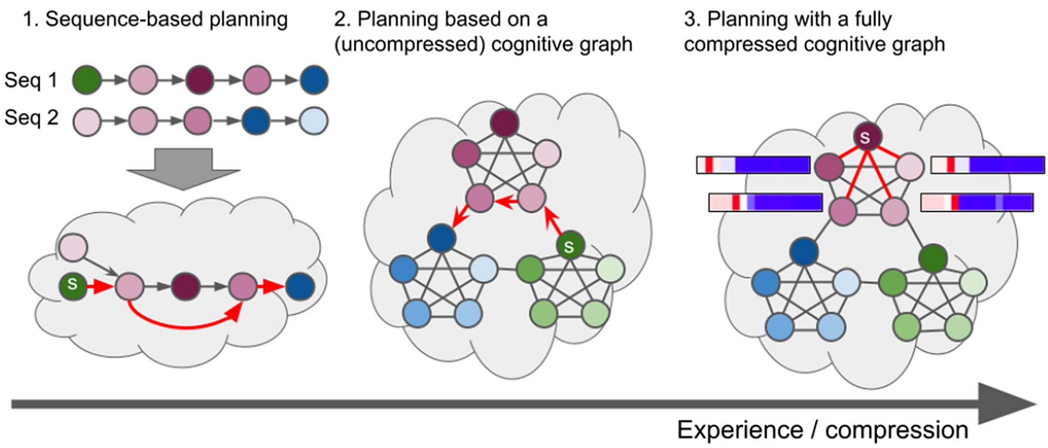
Planning With Learned Cognitive Graphs: Uncompressed Versus Backed-Up

Models in planning capture statistical regularities of the environment, and could be either given a priori or learned from experience. If an agent has full information about the transition structure of the environment, then the agent is able to plan even without experience. This is conceptually relevant to a classical control problem or search algorithm (Korf, 1987): for example, for the game tic-tac-toe, an agent can be endowed with a complete model (or a human can be verbally instructed about the rules of the game). Given this starting point, the player can construct a treelike graph of possible future states and actions, and perform a search to find the optimal decision (Sriram et al., 2009). However, in more naturalistic contexts, the dynamics of the environment are unknown to us initially, and our internal models develop and change with our experience with the environment (Lengyel & Dayan, 2007; Schrittwieser et al., 2020). We confine further discussion to these latter, learned models of the environment.

Cognitive graphs at different levels of compression could serve as models that can support different forms of planning (Figure 2). Raw, uncompressed cognitive graphs support planning via iterative sampling of subsequent states or actions from a given state, or node. Here, individual instances or nodes have minimal information about other nodes, thus making it crucial to traverse graphs based on the

Figure 2

A Graphical Illustration of Planning Based on the Suggested Spectrum of Representations—Specifically, the Degree to Which They Are Precompiled—As a Function of Experience



Note. Nodes indicated with "s" represent starting points. Arrows, or directed edges, describe possible plans for the agent in the starting node to execute. The gray clouds represent the form of model the agent is using to plan in a given phase. 1. Sequence-based planning: This phase represents the early learning phase where an agent has not yet constructed an integrated model of the environmental dynamics. Here, agents are assumed to plan based on sampling instances of previously experienced trajectories. First, two sample trajectories are shown here, labeled "seq 1" and "seq 2." From these two trajectories, an agent is able to create a combined representation and plan efficiently with it (e.g., taking the shortcut as seen in the red arrows). 2. Planning based on an uncompressed cognitive graph: After a few experiences, an agent is able to build cognitive graphs by conjoining past trajectories. Agents are thought to iteratively sample the next actions based on the cognitive graph. As agents gain more experience, simultaneously, a "diffusion-like" process is thought to take place such that information about neighboring nodes is integrated into each node. 3. Planning with a fully compressed cognitive graph: After sufficient experience, a compressed trajectory from a given node to each other node is available in a summary format. Rows of such a *successor representation* are shown here. The availability of information in this representational format allows agents to plan for novel or changing rewards in a statistically efficient manner. See the online article for the color version of this figure.

relationship between nodes, or edges. Thus, in this form of planning, the sampling algorithm is critical.

At the other extreme, actions and states in a cognitive graph are fully backed-up—for instance, successor or predecessor representations. In the successor representation, each node–state contains the expected future state occupancy given a current state and according to a given policy (Dayan, 1993); these can be thought of as integrated trajectories sampled from the current state. Conversely, predecessor representations can be thought of as fully bootstrapped versions of eligibility traces, a memory-like mechanism that assigns credit to past states and actions from a given state (Bailey & Mattar, 2022; Sutton & Barto, 2018). Predecessor representations could be seen as a hindsight version of successor representations in that it bootstraps the possible trajectories *that could have led to* a current state. Whether directed forward or backward, once these bootstrapped representations converge, the expected cumulative reward or credit can be efficiently computed for planning, just by taking the product of the representation and a separate reward function. With fully backed-up cognitive graphs, the need for edge-based sequences, or sampling successor states, becomes negligible.

Planning With Uncompressed Representations: Sampling Instances

During the early stages of learning an environment, planning could be facilitated by instance-based methods instead of relying on an explicit model or rule-based methods. Take, for example, the task of choosing a restaurant to dine in an area that one has just moved to and thus has limited experience with. It might be more effective to decide based on a few recent dining experiences rather than attempting to decide based on a general summary of what little experience one has. A model of decision making under uncertainty captures this intuition as *case-based* decision theory

(Gilboa & Schmeidler, 1995), which suggests that to make decisions under uncertainty, people rely on the memory of similar cases that had worked well in the past.

This nonparametric, kernel-based method offers several cognitive advantages that could potentially bolster decision efficiency and provides a better account of human decision making compared to rule-based methods. For instance, a small group of samples reduces memory load (G. Barron & Erev, 2003), simplifies the decision rule (Fiedler, 2000; Hertwig & Pleskac, 2010), facilitates generalization to unseen observations (H. C. Barron et al., 2013; Wimmer & Shohamy, 2012), and reduces time (C. R. Fox & Hadar, 2006; Hertwig et al., 2004). Also, instance sampling has been shown to be a superior explanation of decision behavior in several laboratory tasks (Bhatia, 2014; Bornstein et al., 2017; Hotaling & Kellen, 2022; S. Wang et al., 2022; Zhao et al., 2022). For example, in a repeated decision task, the individuals appear to rely on small numbers of samples of recent experiences. When intermittent reminder probes were added to the task, manipulating the apparent recency of past experiences, these probes had a significant effect on a subsequent choice (Bornstein et al., 2017). Critically, the choice of instances is sensitive to current task demands: Recent experiences may be more likely to be sampled in an environment that does not have an explicit structure, but the introduction of a periodic structure can lead to more adaptive sampling of relevant events (Plonsky et al., 2015).

An example of case-based decision theory applied to reinforcement learning is episodic control (Dayan, 2008; Lengyel & Dayan, 2007). Episodic control enables the agents to make an informed guess about the value of unseen states by averaging the values of the stored past instances that are most similar to the current state. The kernel-based nonparametric approach that underlies case-based decision theory improves the sample efficiency since the same amount of observations could be used to inform estimates about a greater number of states, compared to classic reinforcement learning, as well as providing a method for generalization, which is particularly important in continuous state spaces (Bhui, 2018; Gershman & Daw, 2017). Simulation results show that this advantage renders episodic control superior to model-based or model-free control during initial learning stages, as it accelerates the learning process under a low-data limit compared to other control methods (Blundell et al., 2016; Lengyel & Dayan, 2007). One drawback of episodic control is that the search process could be inefficient as the number of episodes stored increases. For scalability, neural episodic control uses deep learning methods to embed the keys of each state into a fixed-length vector (Pritzel et al., 2017). Embedded inputs are then fed into a differentiable neural dictionary or a learnable episodic memory system that maps keys to values. The final value of an observation is obtained by the weighted sum of the values in the differentiable neural dictionary, where the weights are computed by the similarity between the current key and the keys of states in the memory system.

Sampling-based accounts of human multistep planning have also provided descriptive value in at least two aspects. First, an extension of decision field theory toward the realm of planning—named decision field theory-planning (Hotaling, 2020)—can explain human planning behavior better than backward induction, at least in situations where multistep plans contend with high-payoff variability. Here, the unreliability of experience may be a critical factor favoring this instance-based approach. In large, continuous, and highly uncertain environments an agent would require unrealistically extensive experience to develop stable, compressed, and predictive representations. Silver and Veness (2010) show that in these environments, asymptotically optimal plans can be constructed using Monte Carlo trajectory sampling over an iteratively updated internal model. A second advantage of representing the full, uncompressed, model of the state space, with all its intermediate states is that it supports effective exploration strategies: in particular, one can perform “far” jumps across state spaces to distal, weakly connected nodes (J. Zhu et al., 2018); the resulting “Lévy flight” behavior matches observations of biological agents exploring novel environments (Hunt et al., 2021) and performing memory search (Rhodes & Turvey, 2007).

Research has shown that in an environment based on graph-like reward structures (e.g., subway maps), people leverage learned graph structure to guide sampling-based decisions (Wu et al., 2019, 2021). Nevertheless, whether people are able to *spontaneously* construct cognitive graphs from sequential experiences in graph-like structures and still leverage this to guide decisions has not yet been directly investigated.

Planning With Backed-Up State/Action Sequences

Earlier, we introduced the concept of backed-up representations as a way of incrementally learning compact summaries of multistep contingencies. Successor representations have been devised to balance the possible computational intractability of fully model-based methods and the inflexibility of computationally cheap model-free methods, providing a robust solution to this problem (Dayan, 1993). These compressed, predictive representations summarize the expected future occupancy of the successor states from a current state given a policy. Using successor representations compresses the multistep planning process into a single-step process since long-range outcomes of all the possible future trajectories are considered at once (Dayan, 1993). This not only reduces computational complexity but also facilitates generalization and learning when adapting to variable reward contingencies. Empirical evidence from studies conducted on humans (Momennejad et al., 2017) and artificial agents (Barreto et al., 2017) suggests that using transition dynamics compressed in a successor representation leads to faster adaptation to value-function changes because only the reward function requires relearning, thus significantly enhancing learning efficiency.

Another example of compressing sequences of observation, or states, is seen in robust predictable control (Eysenbach et al., 2021). This algorithm is explicitly encouraged to find a compressed policy by penalizing complexity, which is operationally defined as the amount of information needed from observations for a policy to make decisions. The intuition behind this is that agents will rely less on gathering information from observations as they become better at predicting the future accurately. Agents trained on compressed policies are less susceptible to unknown or missing observations (i.e., perturbations) since compressed policies have been trained to use fewer bits of information per observation. This leads to improved *open-loop control*—producing a plan of action sequences at the beginning and executing it without checking the progress along the way.

In sum, compressed representations lower the cost of planning by reducing complexity at the representational level. This kind of representation also fosters open-loop planning by enabling the execution of action sequences as a single operation (Eysenbach et al., 2021). This could be efficient in environments where transition dynamics are relatively well-known and unchanging. On the other hand, when models of the environment have not been fully developed yet, instance-based control can be useful. In particular, sampling trajectories of instances to preserve the sequential nature of experiences provides a method with less complexity and greater scalability, while still maintaining high performance. In the following section, we discuss how graph-structured representations can improve trajectory-based planning.

Ways in Which Uncompressed Cognitive Graphs Could Facilitate Planning

Recent approaches transform planning into a graph-search problem (Liu et al., 2020; Savinov et al., 2018). One study leveraged graph-based representations to identify the landmarks or subgoals in latent graphs and then performed graph search on the nodes (Zhang et al., 2021). Here, edges between the nodes are weighted with "reachability" between nodes, making it as a form of a labeled graph. In the domain of spatial navigation, algorithms construct graphs based on subgoals and then plan based on the constructed graphs for efficiency (Bagaria et al., 2021).

It has also been found that people spontaneously construct graph-like representations when observing a sequence of events, where these latent graphs could be either correlational (Kahn et al., 2018; Rmus et al., 2022; Schapiro et al., 2013; Solomon et al., 2019, undirected graphs) or causal (Gopnik et al., 2004; Gopnik & Schulz, 2004; Sommerville & Woodward, 2005a, 2005b, directed graphs). Furthermore, people have been shown to be able to capture the topological structure of an underlying graph (i.e., identifying bottleneck states; Schapiro et al., 2013; Solway et al., 2014), even after the passive observation of trajectories through the graph space. Intriguingly, the general tendency to use plans over model-free approaches appears to be correlated with the ability to infer latent graph-based structure from jumbled sequences of experiences (Rmus et al., 2022), potentially underscoring the utility of learning graph-structured representations in planning.

Mechanisms by Which Cognitive Graphs Could Facilitate Planning

States and observations or instances may not be mapped onto each other in a one-to-one fashion. This phenomenon, referred to as perceptual aliasing, could potentially destabilize control in reinforcement learning (Whitehead & Ballard, 1991). To overcome this, the agents must employ an accurate and parsimonious representation of experience that is able to split identical observations into different underlying states or merge seemingly different observations into a single state for generalization, depending on the context (Niv, 2019). In other words, correctly identifying the underlying latent state associated with an observation is crucial. Latent state inference thus plays an important role in constructing cognitive graphs—especially during early learning, when the small amount of experience can lead to highly uncertain estimates of the state structure, which adaptive decision-makers must account for (Harhen & Bornstein, 2023; Jiang et al., 2015).

When observations are aliased with respect to the underlying latent states, inferring the generative structure requires interpreting each observation relative to the others; formally, conditioning inference on some subset of the history of observations, rather than just the one presently available sensory input (Whittington et al., 2022). Hidden Markov models (HMM) provide a computational solution to latent state inference by decoupling the transition structure of the latent states (transition matrix) and the probability that a given observation maps onto latent states (emission matrix; George et al., 2021; Mark et al., 2020); this dichotomy is also dubbed as "stimulus–stimulus" associations and "stimulus–context latent" associations of content representations, respectively (see S. Wang et al., 2022).

A clone-structured cognitive graph is a version of the HMM that conditions the transition of latent spaces on actions (George et al., 2021). To elaborate, a given observation is explained in terms of two components: a transition tensor which accounts for the action-conditioned transitions between latent states and an emission matrix that assigns probabilities to the latent states given an observation. Within the transition tensor, each latent state in a sequence is identified in relation to its previous latent state and action, and whenever a new context—or a new combination of previous latent state and action—is encountered, a new clone is created. Clone-structured cognitive graphs have been able to capture phenomena thought to be important to structure learning in both spatial (George et al., 2021) and nonspatial (Swaminathan et al., 2023) domains: *splitting*, the ability to recover the ground-truth space from aliased observations, as well as *merging*, the ability to stitch overlapping latent states together from two disjoint observations. Thus, the clone-structured cognitive graph is an exciting proposal for how an agent can simultaneously learn both the structure (i.e., nodes) of the environment as well as its transition dynamics (i.e., edges). Within the framework we discuss here, the resulting representation is considered *uncompressed*, as it is attempting to capture the full, flexible environment model. Backed-up representations can be built by querying the resulting graph, as it stabilizes with sufficient experience (Wittkuhn et al., 2022).

Another variant of the HMM-based cognitive graph explicitly assumes the idea of predefined schemas for identifying the transition structure. Here, it is postulated that the transition dynamics emerge from predefined structural forms such as hexagonal grids or community structures (Mark et al., 2020), which could be grounded in the wider notion of inherent basis sets (Kemp & Tenenbaum, 2008; Luettgau et al., 2023; Tenenbaum et al., 2011) or generative grammar of sequences (Dragoi, 2024). The idea that cognitive graphs are constructed using the prior knowledge of structures could be empirically supported by results that human transfer learning is best explained by these models (Luettgau et al., 2023; Mark et al., 2020). The hippocampal–entorhinal system has been proposed to underlie decoupling, or factorizing, structure, and sensory observations (Whittington et al., 2018). Here, the medial entorhinal cortex contains grid cells (Hafting et al., 2005) that provide a basis set along which the transition structure is defined, and the lateral entorhinal cortex supports sensory representations. The conjunctive code of the transition structure and "emission" is hypothesized to be reflected in the hippocampus (Whittington et al., 2018). These distinct representational forms each play a critical role in the use of hippocampal replay to infer compositional structure across environments, permitting the construction of more compressed representations that can support efficient planning in novel environments (Kurth-Nelson et al., 2023). An area for future research is whether endowing artificial agents with this representational decomposition and algorithmic approach to replaying and recombining structure elements can allow them to perform an efficient approximation to graph compression.

Below, we describe how cognitive graphs support both merging and splitting in sequences of observations, and what specific mechanisms an unfolded graph could provide to facilitate early-stage learning.

Merging: Fast Generalization by Extrapolating Trajectories

Associative memory could be seen as the building block of cognitive graphs. One such instantiation is transitive inference, which is an example of leveraging the relational information of instances for faster generalization, observed in humans and animals (Bryant & Trabasso, 1971; Davis, 1992; Gillan, 1981). When an agent experiences $A > B$ and $B > C$, the unobserved relationship between $A > C$ can be inferred without direct experience (Eichenbaum et al., 1999). This can be achieved through forming supraordinate representations, comparable to cognitive graphs, such that $A > B > C$, which has been found to be supported by the hippocampus (Dusek & Eichenbaum, 1997; Greene et al., 2006; Zalesak & Heckers, 2009). Similarly, disparate fragments of event trajectories can be fused together, creating graph-like formations by leveraging the intersections of these trajectories (Eichenbaum & Cohen, 2014; Rmus et al., 2022). From these graphs, inferences can be made between the instances that were not directly experienced together, supporting flexible recombination and fast generalization (Eichenbaum, 2004). After learning sequences of objects that are generated based on the graphs that are either hexagonal or community-structured, humans are able to infer unobserved links using the transition structure of the latent graphs (Mark et al., 2020). This study provides direct evidence that people are able to extract long-run transition structures from sequences of events and also are able to transfer it for generalization.

Implementing this associative memory-based cognitive graph leads to efficient planning algorithms. For example, an episodic reinforcement learning algorithm called Episodic Reinforcement Learning with Associative Memory (ERLAM) augmented with associative memory showed increased sample efficiency compared to benchmarks (G. Zhu et al., 2020). In ERLAM, experienced trajectories are reorganized into graphs, which speeds the propagation of value learned from one instance to other related instances, thereby enhancing sample efficiency. In addition, clone-structured cognitive graphs introduced earlier have been shown to be capable of performing transitive inferences (George et al., 2021). This capability was demonstrated in the spatial domain, in agents navigating a larger environment divided into discrete "rooms." Here, two separate rooms are stitched together to form an overlapping region. Agents navigate each room separately and are tested on whether they can travel from a nonoverlapping region of one room to a region exclusive to the other room. Results show that agents are able to construct a latent map by stitching sequential observations from two disjoint episodes; overlapping observations from different trajectories are correctly assigned to the same hidden state.

In addition to conjoining separate sequences, associative memory binds seemingly independent choice options together into a temporal context, so that learning the value of a chosen option also influences the value of unchosen options (Biderman & Shohamy, 2021). This is referred to as counterfactual reasoning, another example of associative memory accelerating learning since information about an instance can be propagated to related experiences. Counterfactual reasoning is observed in reinforcement learning: humans not only deploy "factual" information through direct trial-and-error but also incorporate counterfactual learning (Boorman et al., 2011; Fischer & Ullsperger, 2013). Interestingly, counterfactual learning engages cognitive graphs for *both* model-based and model-*free* learning (Moran et al., 2021). In this process, the model-free values of options are positively reinforced by direct rewards and negatively influenced by the value of counterfactual options. Associative memory strength between options in reinforcement learning is correlated with how much learning about one option influences other unchosen options suggesting that counterfactual learning operates on a cognitive graph where edge weights are defined by associative memory strength between items (Biderman et al., 2023). An open question is whether factual and counterfactual learning are performed on the same cognitive graph. Some evidence points to a single representation supporting both kinds of reasoning (Boorman et al., 2009; Fischer & Ullsperger, 2013), whereas other evidence supports these forms of learning update distinct representations (Kishida et al., 2016; Li & Daw, 2011; Lohrenz et al., 2007). A common finding is that individuals are generally biased toward reinforcing their own choices ("confirmation bias," or the tendency to collect information partially according to the preexisting belief or action; Nickerson, 1998), in a way that they incorporate more information when the chosen option is more rewarded (i.e., greater learning rate for positive prediction errors of factual options) and when the unchosen option turns out to be less rewarding (i.e., greater learning rate for negative prediction errors of counterfactual options; Palminteri et al., 2017). Asymmetric updating in the other direction (more negative than positive) has been observed in individuals diagnosed with psychiatric disorders (e.g., depression; Rouhani & Niv, 2019); though this pattern has itself been shown to arise from the individual differences in representational precision (Harhen & Bornstein, 2024).

ERLAM provides an example of leveraging counterfactual combinatorial trajectories to facilitate the learning of artificial agents (G. Zhu et al., 2020). In this algorithm, trajectories are reorganized into graphs by merging the common elements (nodes) of the two trajectories. This agent would have an advantage over an agent who uses pure episodic memory in cases such as right after experiencing two intersecting trajectories that each lead to reward (e.g., $A > B > C >$ reward) and no-reward (e.g., $D > B > E >$ no-reward); while the ERLAM agent will be able to leverage the graph to plan an unexperienced route (e.g., $D > B > C >$ reward), an agent that only relies on episodic reinforcement learning would associate D with reward only after the direct experience. Recently, *expected eligibility traces* have been introduced as a form of leveraging counterfactual trajectories to accelerate learning (van Hasselt et al., 2021). Eligibility trace is a mechanism in reinforcement learning that provides a hindsight credit assignment with regard to the current state by keeping a trace of past experiences weighted by their recency (Singh & Sutton, 1996; Sutton & Barto, 2018). *Expected* eligibility traces improve the limitation of eligibility traces—that only one directly experienced trace is updated each time—by considering multiple counterfactual sequences that could have preceded a current state. Mirroring the relationship between the full forward model and successor representations, the predecessor representation is the fully backed-up version of the state tree supporting expected eligibility traces (Bailey & Mattar, 2022).

Splitting: Recovering Latent Structure From Aliased Sequences

It is possible that two different states are "aliased" or mapped onto overlapping observations. In this situation, as opposed to the example above, agents should be able to create a graph that *merges* two sequences—$A > D > C$ and $B > D > E$—an agent should be able to *split* D into two different nodes according to their contexts. Clone-structured cognitive graphs are able to accurately reconstruct correct latent graphs from sequences of aliased sensory observations by making clones of observations (George et al., 2021). Impressively, clone-structured cognitive graphs are not only able to both split aliased observation into latent states but also able to merge the reconstructed graphs as in transitive inference.

Indeed, as implied above, these are exactly the sorts of environments in which clone-structured cognitive graphs have an advantage over backed-up representations. Specifically, when presenting a clone-structured cognitive graph agent with a sequence of aliased observations from a graph with community structure (e.g., Figure 1, left), the agent is able to recover the modular structure. However, an agent that used a successor representation was not able to recover this structure (George et al., 2021). This suggests that environments in which modular structure is important to the task at hand benefit from having available less-compressed representations of experience. This idea aligns with the finding that sequences of observations generated by a modular versus lattice graph—where the two graphs only differ in terms of their higher-order structure—lead to more robust latent representations (Kahn et al., 2023).

Discussion

Multistep planning is a critical ability for autonomous organisms. Extensive research has identified multiple kinds of planning, each with their own benefits and appropriate to specific situations. These distinct approaches rely on different representational substrates and have different algorithmic commitments. Which kind of planning an individual performs in a given setting can dramatically change the outcome of their decisions. Therefore, it can be valuable to judgment and decision-making research to understand how to characterize the commonalities, differences, and appropriate uses of each form of planning. Here, we suggest that these seemingly distinct representational forms that support planning can be described as varying types of *cognitive graphs*, where these various manifestations of graphs exist along a spectrum of compression.

At one end of the spectrum, fully uncompressed representations capture every element of an associative network in detail. This full-featured model of the environment allows for flexible trajectory sampling at the time of decision and supports plans that are robust to changes in contingency and reward structure. In addition, this kind of uncompressed representation is a necessary first step for building more compressed representations because the latent structure (edges) in compressed representations requires inferring across multiple experiences (Lynn & Bassett, 2020; Wittkuhn et al., 2022). Since observations are often aliased, uniquely characterizing their latent state (nodes) and structure (edges) requires them to be placed in a sequence (Whittington et al., 2022). Network model simulations show that uncompressed sequences of events are necessary for building latent graphs that enable complex functions such as rapid value propagation (G. Zhu et al., 2020) or extracting higher order structures (George et al., 2021).

At the other extreme, a fully compressed graph—such as the *successor* and *predecessor* representations—captures summary-level statistical structure. These graphs are formed by "bootstrapping"—a repeated sampling of the full model to identify the long-run relationships between each pair of nodes in the network. These representations allow for fast, cheap multistep planning as they cache previous trajectory samples into a compact matrix format. Their *factorized* form, separating transition (edge) information from reward values, allows for replanning in the face of changing reward outcomes. However, the kind of planning they support is not robust to changes in contingency structure—these must be relearned, slowing planning until stable estimates can be obtained again. This is because backed-up representations like the successor representation are conditional on the specific policy that generated the compressed graphs; in other words, if the goal changes, the optimal action in each state should be relearned, thereby not transferable (Lehnert et al., 2017). Linear reinforcement learning, which incentivizes learning a "default" policy distributed uniformly across possible successor states, is a framework that addresses this limitation and explains flexible replanning in humans (Piray & Daw, 2021).

These different kinds of representations are learned simultaneously, which allows the agent to arbitrate between the most reliable representations at a given moment (S. Wang et al., 2022). In situations of high uncertainty, such as during the early phase of reinforcement learning where there is not enough data to construct a reliable model (Lengyel & Dayan, 2007) or in volatile environments (Nicholas et al., 2022), consulting on a subset of episodic samples provides a more reliable approximation of the value of observations. Arbitration between different representations has been proposed to be reflected in discontinuous "jumps" of subjective evidence ("jump-diffusions") in evidence accumulation models, where these sudden jumps during the sequential evidence sampling could indicate alternations to other sets of representations (S. Wang et al., 2022). The constantly changing ensemble of representations that lead to these jumps is hypothesized to occur in a bottom-up manner, akin to product-of-experts in machine learning (S. Wang et al., 2022).

By which mechanism does compression happen, such that more experience gradually leads to more compression? One possible mechanism could be the diffusion of information between nodes through a replay of events. The transition from the uncompressed graph to the fully backed-up form occurs via repeatedly sampling and aggregating features from neighboring nodes, analogous to message-passing algorithms (George et al., 2021; Hamilton et al., 2017; Parr et al., 2019). At the beginning of the learning process, the cognitive graph resembles an undiffused graph where a node, or a given state, holds limited information about others, thus requiring the agent in a state to explicitly traverse edges to infer about other states. At the same time, uncompressed graphs provide full representations of the contingency structure between states and actions, which allow for flexibility at the cost of greater computation time and behavioral variability. With more experience, the cognitive graphs undergo a transformation into a bootstrapped representation where information about future states is aggregated into each adjacent state, making explicit edge-based inferences between states less important. Caching these distal outcomes subserves rapid planning, while still retaining sensitivity to changes in reward availability. However, without additional mechanisms, it also confers a relative insensitivity to contingency changes that may be undesirable in novel or volatile environments. Replay of events could be a biological instantiation of message passing, given that the construction of backed-up representations is mediated by on-task replay in humans (Wittkuhn et al., 2022). A possible future direction for research would be to investigate whether replay contributes to maintaining and arbitrating between multiple kinds of representations.

One interesting direction to expand this concept of representational spectrum would be to test whether different modes of control arise as a function of the degree of compression (Moskovitz et al., 2022). Recall that agents using compressed representations should be adept at open-loop control because they can in principle select action sequences into a single operation. This eliminates the need for intermittent replanning during action execution (Eysenbach et al., 2021). However, if agents plan by sequentially sampling the next actions using uncompressed cognitive graphs, taking small steps could be more efficient than open-loop control, since the model has not been compiled yet to provide reliable future trajectories from a state. This edge-based planning is conceptually more similar to closed-loop planning, where an agent stops at each transition to replan. This leads us to the overarching question of whether the utility of using closed-loop versus open-loop planning aligns with the degree of compression in cognitive graphs. This alignment would be similar to the evolution of an episodic control system to a model-based, and then finally to model-free systems (Lengyel & Dayan, 2007; Yoo & Bornstein, 2024). Based on an interpretation that the seemingly model-free behaviors could actually be action sequences (Dezfouli & Balleine, 2012, 2013), an interesting hypothesis is that the model-free system at the end of the spectrum could be in fact representing action sequences formed by an open-loop control, likely a result of using highly compressed models.

Conclusions

To conclude, we highlight the kind of representations that could be used to support instance-based planning at early stages of learning—uncompressed cognitive graphs—and suggest that they could be in a spectrum, rather than discrete concepts, with backed-up representations at the other end. Further research may investigate whether this spectrum of *representations* directly induces a continuum of planning *algorithms*, such as closed- versus open-loop control.

Citation Diversity Statement

Recent work in several fields of science has identified a bias in citation practices such that articles from women and other minority scholars are undercited relative to the number of such articles in the field (Bertolero et al., 2020; Caplar et al., 2017; Chatterjee & Werner, 2021; Dion et al., 2018; Dworkin et al., 2020; Fulvio et al., 2021; Maliniak et al., 2013; Mitchell et al., 2013; X. Wang et al., 2021). Here, we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference using the databases that store the probability of a first name being carried by a woman (Dworkin et al., 2020; Zhou et al., 2020). By this measure (and excluding self-citations to the first and last authors of our current article), our references contain 5.35% woman (first)/woman (last), 13.0% man/woman, 18.59% woman/man, and 63.05% man/man. This method is limited in that (a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and (b) it cannot account for intersex, nonbinary, or transgender people. Second, we obtained predicted racial/ethnic category of the first and last author of each reference by the databases that store the probability of a first and last name being carried by an author of color (Ambekar et al., 2009; Sood & Laohaprapanon, 2018). By this measure (and excluding self-citations), our references contain 7.12% author of color (first)/author of color(last), 11.21% White author/author of color, 20.16% author of color/White author, and 61.51% White author/White author. This method is limited in that (a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity and (b) it cannot account for Indigenous and mixed-race authors or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

References

Ambekar, A., Ward, C., Mohammed, J., Male, S., & Skiena, S. (2009). *Name-ethnicity classification from open sources* [Conference session]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Bagaria, A., Senthil, J. K., & Konidaris, G. (2021). Skill discovery for exploration and planning using deep skill graphs. In M. Meila & T. Zhang (Eds.), *International conference on machine learning* (pp. 521–531). Proceedings of Machine Learning Research.

Bailey, D., & Mattar, M. (2022). *Predecessor features*. arXiv preprint. https://doi.org/10.48550/arXiv.2206.00303

Barreto, A., Dabney, W., Munos, R., Hunt, J. J., Schaul, T., van Hasselt, H. P., & Silver, D. (2017). Successor features for transfer in reinforcement learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30, pp. 4055–4065). Neural Information Processing Systems Foundation, Inc.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233. https://doi.org/10.1002/bdm.443

Barron, H. C., Dolan, R. J., & Behrens, T. E. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience*, *16*(10), 1492–1498. https://doi.org/10.1038/nn.3515

Bertolero, M. A., Dworkin, J. D., David, S. U., Lloreda, C. L., Srivastava, P., Stiso, J., Zhou, D., Dzirasa, K., Fair, D. A., Kaczkurkin, A. N., Marlin, B. J., Shohamy, D., Uddin, L. Q., Zurn, P., & Bassett, D. S. (2020). *Racial and ethnic imbalance in neuroscience reference lists and intersections with gender*. bioRxiv. https://doi.org/10.1101/2020.10.12.336230

Bertsekas, D. (2012). *Dynamic programming and optimal control: Volume I* (Vol. 4). Athena Scientific.

Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review*, *21*(5), 1095–1111. https://doi.org/10.3758/s13423-014-0650-1

Bhui, R. (2018). Case-based decision neuroscience: Economic judgment by similarity. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-directed decision making: Computations and neural circuits* (pp. 67–103). Academic Press. https://doi.org/10.1016/B978-0-12-812098-9.00004-8

Biderman, N., Gershman, S. J., & Shohamy, D. (2023). The role of memory in counterfactual valuation. *Journal of Experimental Psychology: General*, *152*(6), 1754–1767. https://doi.org/10.1037/xge0001364

Biderman, N., & Shohamy, D. (2021). Memory and decision making interact to shape the value of unchosen options. *Nature Communications*, *12*(1), Article 4648. https://doi.org/10.1038/s41467-021-24907-x

Blundell, C., Uria, B., Pritzel, A., Li, Y., Ruderman, A., Leibo, J. Z., Rae, J., Wierstra, D., & Hassabis, D. (2016). *Model-free episodic control*. arXiv preprint. https://doi.org/10.48550/arXiv.1606.04460

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLOS Biology*, *9*(6), Article e1001093. https://doi.org/10.1371/journal.pbio.1001093

Boorman, E. D., Behrens, T. E., Woolrich, M. W., & Rushworth, M. F. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*(5), 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*(7), 1011–1023. https://doi.org/10.1111/j.1460-9568.2011.07920.x

Bornstein, A. M., & Daw, N. D. (2013). Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLOS Computational Biology*, *9*(12), Article e1003387. https://doi.org/10.1371/journal.pcbi.1003387

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*(1), Article 15958. https://doi.org/10.1038/ncomms15958

Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*(7), 997–1003. https://doi.org/10.1038/nn.4573

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, *232*(5311), 456–458. https://doi.org/10.1038/232456a0

Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science*, *325*(5939), 414–416. https://doi.org/10.1126/science.1171022

Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, *1*(6), Article 0141. https://doi.org/10.1038/s41550-017-0141

Chatterjee, P., & Werner, R. M. (2021). Gender disparity in citations in high-impact journal articles. *JAMA Network Open*, *4*(7), Article e2114509. https://doi.org/10.1001/jamanetworkopen.2021.14509

Chrastil, E. R., & Warren, W. H. (2014). From cognitive maps to cognitive graphs. *PLOS ONE*, *9*(11), Article e112544. https://doi.org/10.1371/journal.pone.0112544

Davis, H. (1992). Transitive inference in rats (Rattus norvegicus). *Journal of Comparative Psychology*, *106*(4), 342–349. https://doi.org/10.1037/0735-7036.106.4.342

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, *5*(4), 613–624. https://doi.org/10.1162/neco.1993.5.4.613

Dayan, P. (2008). The role of value systems in decision making. In C. Engel & W. Singer (Eds.), *Better than conscious? Decision making, the human mind, and implications for institutions* (pp. 51–70). MIT Press.

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *European Journal of Neuroscience*, *35*(7), 1036–1051. https://doi.org/10.1111/j.1460-9568.2012.08050.x

Dezfouli, A., & Balleine, B. W. (2013). Actions, action sequences and habits: Evidence that goal-directed and habitual action control are hierarchically organized. *PLOS Computational Biology*, *9*(12), Article e1003364. https://doi.org/10.1371/journal.pcbi.1003364

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, *26*(3), 312–327. https://doi.org/10.1017/pan.2018.12

Doya, K., Samejima, K., Katagiri, K.-i., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation*, *14*(6), 1347–1369. https://doi.org/10.1162/089976602753712972

Dragoi, G. (2024). The generative grammar of the brain: A critique of internally generated representations. *Nature Reviews Neuroscience*, *25*(1), 60–75. https://doi.org/10.1038/s41583-023-00763-0

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(13), 7109–7114. https://doi.org/10.1073/pnas.94.13.7109

Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). *The extent and drivers of gender imbalance in neuroscience reference lists*. bioRxiv. https://doi.org/10.1101/2020.01.03.894378

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. https://doi.org/10.1016/j.neuron.2004.08.028

Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, *83*(4), 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, *23*(2), 209–226. https://doi.org/10.1016/s0896-6273(00)80773-4

Eysenbach, B., Salakhutdinov, R. R., & Levine, S. (2021). Robust predictable control. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems* (Vol. 34, pp. 27813–27825). Neural Information Processing Systems Foundation, Inc.

Feld, G., Bernard, M., Rawson, A., & Spiers, H. (2022). Sleep targets highly connected global and local nodes to aid consolidation of learned graph networks. *Scientific Reports*, *12*(1), Article 15086. https://doi.org/10.1038/s41598-022-17747-2

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*(4), 659–676. https://doi.org/10.1037/0033-295X.107.4.659

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243–1255. https://doi.org/10.1016/j.neuron.2013.07.006

Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*(2), 159–161. https://doi.org/10.1017/S1930297500002370

Fox, R., Krishnan, S., Stoica, I., & Goldberg, K. (2017). *Multi-level discovery of deep options*. arXiv preprint. https://doi.org/10.48550/arXiv.1703.08294

Fulvio, J. M., Akinnola, I., & Postle, B. R. (2021). Gender (im)balance in citation practices in cognitive neuroscience. *Journal of Cognitive Neuroscience*, *33*(1), 3–7. https://doi.org/10.1162/jocn_a_01643

Geerts, J. P., Gershman, S. J., Burgess, N., & Stachenfeld, K. L. (2022). *A probabilistic successor representation for context-dependent prediction*. bioRxiv. https://doi.org/10.1101/2022.06.03.494671

George, D., Rikhye, R. V., Gothoskar, N., Guntupalli, J. S., Dedieu, A., & Lázaro-Gredilla, M. (2021). Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nature Communications*, *12*(1), Article 2392. https://doi.org/10.1038/s41467-021-22559-5

Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience*, *38*(33), 7193–7200. https://doi.org/10.1523/jneurosci.0151-18.2018

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, *68*, 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The successor representation and temporal context. *Neural Computation*, *24*(6), 1553–1568. https://doi.org/10.1162/neco_a_00282

Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*, *110*(3), 605–639. https://doi.org/10.2307/2946694

Gillan, D. J. (1981). Reasoning in the chimpanzee: II. Transitive inference. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*(2), 150–164. https://doi.org/10.1037/0097-7403.7.2.150

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*(1), 3–32. https://doi.org/10.1037/0033-295X.111.1.3

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, *8*(8), 371–377. https://doi.org/10.1016/j.tics.2004.06.005

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2006). An FMRI analysis of the human hippocampus: Inference, context, and task awareness. *Journal of Cognitive Neuroscience*, *18*(7), 1156–1173. https://doi.org/10.1162/jocn.2006.18.7.1156

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. https://doi.org/10.1038/nature03721

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30, pp. 1024–1034). Neural Information Processing Systems Foundation, Inc.

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(13), Article e2216524120. https://doi.org/10.1073/pnas.2216524120

Harhen, N. C., & Bornstein, A. M. (2024). Interval timing as a computational pathway from early life adversity to affective disorders. *Topics in Cognitive Science*, *16*(1), 92–112. https://doi.org/10.1111/tops.12701

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009

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*, *606*(7912), 129–136. https://doi.org/10.1038/s41586-022-04743-9

Hotaling, J. M. (2020). Decision field theory-planning: A cognitive model of planning on the fly in multistage decision making. *Decision*, *7*(1), 20–42. https://doi.org/10.1037/dec0000113

Hotaling, J. M., & Kellen, D. (2022). Dynamic decision making: Empirical and theoretical directions. In K. D. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 76, pp. 207–238). Elsevier.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hunt, L., Daw, N., Kaanders, P., MacIver, M., Mugan, U., Procyk, E., Redish, A., Russo, E., Scholl, J., Stachenfeld, K., Wilson, C. R. E., & Kolling, N. (2021). Formalizing planning and information search in naturalistic decision-making. *Nature Neuroscience*, *24*(8), 1051–1064. https://doi.org/10.1038/s41593-021-00866-w

Jeong, H., Taylor, A., Floeder, J. R., Lohmann, M., Mihalas, S., Wu, B., Zhou, M., Burke, D. A., & Namboodiri, V. M. K. (2022). Mesolimbic dopamine release conveys causal associations. *Science*, *378*(6626), Article eabq6740. https://doi.org/10.1126/science.abq6740

Jiang, N., Kulesza, A., Singh, S., & Lewis, R. (2015). *The dependence of effective planning horizon on model accuracy* [Conference session]. Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems.

Kahn, A. E., Karuza, E. A., Vettel, J. M., & Bassett, D. S. (2018). Network constraints on learnability of probabilistic motor sequences. *Nature Human Behaviour*, *2*(12), 936–947. https://doi.org/10.1038/s41562-018-0463-8

Kahn, A. E., Szymula, K., Bassett, D. S., Nyema, N., Haggerty, E. B., & Aguirre, G. K. (2023). *Network structure influences the strength of learned neural representations*. bioRxiv. https://doi.org/10.1101/2023.01.23.525254

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10687–10692. https://doi.org/10.1073/pnas.0802631105

Kinney, D., & Lombrozo, T. (2023a). *Building compressed causal models of the world*. PsyArXiv. https://doi.org/10.31234/osf.io/2f7x6

Kinney, D., & Lombrozo, T. (2023b). *Lossy compression and the granularity of causal representation* [Poster presentation]. NeurIPS 2023 Workshop: Information-Theoretic Principles in Cognitive Systems.

Kishida, K. T., Saez, I., Lohrenz, T., Witcher, M. R., Laxton, A. W., Tatter, S. B., White, J. P., Ellis, T. L., Phillips, P. E., & Montague, P. R. (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 200–205. https://doi.org/10.1073/pnas.1513619112

Korf, R. E. (1987). Planning as search: A quantitative approach. *Artificial Intelligence*, *33*(1), 65–88. https://doi.org/10.1016/0004-3702(87)90051-8

Kurth-Nelson, Z., Behrens, T., Wayne, G., Miller, K., Luettgau, L., Dolan, R., Liu, Y., & Schwartenbeck, P. (2023). Replay and compositional computation. *Neuron*, *111*(4), 454–469. https://doi.org/10.1016/j.neuron.2022.12.028

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. In K. D. Federmeier (Ed.), *Psychology of learning and motivation* (pp. 195–232). Elsevier.

Lai, L., Huang, A. Z., & Gershman, S. J. (2022). *Action chunking as policy compression*. PsyArXiv. https://doi.org/10.31234/osf.io/z8yrv

Lehnert, L., Tellex, S., & Littman, M. L. (2017). *Advantages and limitations of using successor features for transfer in reinforcement learning*. arXiv preprint. https://doi.org/10.48550/arXiv.1708.00102

Lengyel, M., & Dayan, P. (2007). Hippocampal contributions to control: The third way. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (Vol. 20, pp. 889–896). Neural Information Processing Systems Foundation, Inc.

Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience*, *31*(14), 5504–5511. https://doi.org/10.1523/jneurosci.6316-10.2011

Liu, K., Kurutach, T., Tung, C., Abbeel, P., & Tamar, A. (2020). Hallucinative topological memory for zero-shot visual planning. In H. Daumé III & A. Singh (Eds.), *International conference on machine learning* (pp. 6259–6270). Proceedings of Machine Learning Research.

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(22), 9493–9498. https://doi.org/10.1073/pnas.0608842104

Luettgau, L., Erdmann, T., Veselic, S., Stachenfeld, K., Moran, R., Kurth-Nelson, Z., & Dolan, R. (2023). *Decomposing dynamical subprocesses for compositional generalization*. PsyArXiv. https://doi.org/10.31234/osf.io/sxn4a

Lynn, C. W., & Bassett, D. S. (2020). How humans learn and represent networks. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(47), 29407–29415. https://doi.org/10.1073/pnas.1912328117

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), Article 2313. https://doi.org/10.1038/s41467-020-15146-7

Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, *67*(4), 889–922. https://doi.org/10.1017/S0020818313000209

Mark, S., Moran, R., Parr, T., Kennerley, S. W., & Behrens, T. E. (2020). Transferring structural knowledge across cognitive maps in humans and models. *Nature Communications*, *11*(1), Article 4783. https://doi.org/10.1038/s41467-020-18254-6

Mitchell, S. M., Lange, S., & Brus, H. (2013). Gendered citation patterns in international relations journals. *International Studies Perspectives*, *14*(4), 485–492. https://doi.org/10.1111/insp.12026

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692. https://doi.org/10.1038/s41562-017-0180-8

Moran, R., Dayan, P., & Dolan, R. J. (2021). Human subjects exploit a cognitive map for credit assignment. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(4), Article e2016884118. https://doi.org/10.1073/pnas.2016884118

Moskovitz, T., Hromadka, S., Touati, A., Borsa, D., & Sahani, M. (2023). *A state representation for diminishing rewards*. arXiv preprint. https://doi.org/10.48550/arXiv.2309.03710

Moskovitz, T., Miller, K., Sahani, M., & Botvinick, M. M. (2022). *A unified theory of dual-process control*. arXiv preprint. https://doi.org/10.48550/arXiv.2211.07036

Moskovitz, T., Wilson, S. R., & Sahani, M. (2021). *A first-occupancy representation for reinforcement learning*. arXiv preprint. https://doi.org/10.48550/arXiv.2109.13863

Moss, L. S. (1984). The laws of human relations and the rules of human action derived therefrom. By Hermann Heinrich Gossen. Translated by Rudolph C. Blitz with an introductory essay by Nicholas Georgescu-Roegen. Cambridge: MIT Press, 1983. pp. 460. $47.50. *The Journal of Economic History*, *44*(4), 1130–1132. https://doi.org/10.1017/S0022050700033350

Muller, R. U., Stead, M., & Pach, J. (1996). The hippocampus as a cognitive graph. *The Journal of General Physiology*, *107*(6), 663–694. https://doi.org/10.1085/jgp.107.6.663

Natarajan, M., & Kolobov, A. (2022). *Planning with Markov decision processes: An AI perspective*. Springer Nature.

Nicholas, J., Daw, N. D., & Shohamy, D. (2022). Uncertainty alters the balance between incremental learning and episodic memory. *Elife*, *11*, Article e81679. https://doi.org/10.7554/elife.81679

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Noh, S. M., Singla, U. K., Bennett, I. J., & Bornstein, A. (2023). Memory precision and age differentially predict the use of decision-making strategies across the lifespan. *Scientific Reports*, *13*(1), Article 17014. https://doi.org/10.1038/s41598-023-44107-5

Otto, A. R., Devine, S., Schulz, E., Bornstein, A. M., & Louie, K. (2022). Context-dependent choice and evaluation in real-world consumer behavior. *Scientific Reports*, *12*(1), Article 17744. https://doi.org/10.1038/s41598-022-22416-5

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), Article 8096. https://doi.org/10.1038/ncomms9096

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S.-J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLOS Computational Biology*, *13*(8), Article e1005684. https://doi.org/10.1371/journal.pcbi.1005684

Parr, T., Markovic, D., Kiebel, S. J., & Friston, K. J. (2019). Neuronal message passing using mean-field, Bethe, and marginal approximations. *Scientific Reports*, *9*(1), Article 1889. https://doi.org/10.1038/s41598-018-38246-3

Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends in Cognitive Sciences*, *25*(1), 37–54. https://doi.org/10.1016/j.tics.2020.10.004

Piray, P., & Daw, N. D. (2021). Linear reinforcement learning in planning, grid fields, and cognitive control. *Nature Communications*, *12*(1), Article 4942. https://doi.org/10.1038/s41467-021-25123-3

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*(4), 621–647. https://doi.org/10.1037/a0039413

Pritzel, A., Uria, B., Srinivasan, S., Badia, A. P., Vinyals, O., Hassabis, D., Wierstra, D., & Blundell, C. (2017). Neural episodic control. In D. Precup & Y. W. Teh (Eds.), *International conference on machine learning* (pp. 2827–2836). Proceedings of Machine Learning Research.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Rhodes, T., & Turvey, M. T. (2007). Human memory retrieval as Lévy foraging. *Physica A: Statistical Mechanics and its Applications*, *385*(1), 255–260. https://doi.org/10.1016/j.physa.2007.07.001

Rmus, M., Ritz, H., Hunter, L. E., Bornstein, A. M., & Shenhav, A. (2022). Humans can navigate complex graph structures acquired during latent learning. *Cognition*, *225*, Article 105103. https://doi.org/10.1016/j.cognition.2022.105103

Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology*, *236*(8), 2425–2435. https://doi.org/10.1007/s00213-019-05322-z

Savinov, N., Dosovitskiy, A., & Koltun, V. (2018). *Semi-parametric topological memory for navigation*. arXiv preprint. https://doi.org/10.48550/arXiv.1803.00653

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. https://doi.org/10.1038/nn.3331

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T., & Silver, D. (2020). Mastering Atari, go, chess and shogi by planning with a learned model. *Nature*, *588*(7839), 604–609. https://doi.org/10.1038/s41586-020-03051-4

Silver, D., & Veness, J. (2010). Monte-Carlo planning in large POMDPS. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems* (Vol. 23, pp. 2164–2172). Neural Information Processing Systems Foundation, Inc.

Singh, S. P., & Sutton, R. S. (1996). Reinforcement learning with replacing eligibility traces. *Machine Learning*, *22*(1–3), 123–158. https://doi.org/10.1007/BF00114726

Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, *79*(2), 361–374. https://doi.org/10.1016/j.neuron.2013.05.038

Solomon, E. A., Lega, B. C., Sperling, M. R., & Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the*

National Academy of Sciences of the United States of America, *116*(48), 24343–24352. https://doi.org/10.1073/pnas.1906729116

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., & Botvinick, M. M. (2014). Optimal behavioral hierarchy. *PLOS Computational Biology*, *10*(8), Article e1003779. https://doi.org/10.1371/journal.pcbi.1003779

Sommerville, J. A., & Woodward, A. L. (2005a). Infants' sensitivity to the causal features of means-end support sequences in action and perception. *Infancy*, *8*(2), 119–145. https://doi.org/10.1207/s15327078in0802_2

Sommerville, J. A., & Woodward, A. L. (2005b). Pulling out the intentional structure of action: The relation between action processing and action production in infancy. *Cognition*, *95*(1), 1–30. https://doi.org/10.1016/j.cognition.2003.12.004

Sood, G., & Laohaprapanon, S. (2018). *Predicting race and ethnicity from the sequence of characters in a name*. arXiv preprint. https://doi.org/10.48550/arXiv.1805.02109

Sriram, S., Vijayarangan, R., Raghuraman, S., & Yuan, X. (2009). *Implementing a no-loss state in the game of tic-tac-toe using a customized decision tree algorithm* [Conference session]. 2009 International Conference on Information and Automation. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5204878

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653. https://doi.org/10.1038/nn.4650

Stiso, J., Lynn, C. W., Kahn, A. E., Rangarajan, V., Szymula, K. P., Archer, R., Revell, A., Stein, J. M., Litt, B., Davis, K. A., Lucas, T. H., & Bassett, D. S. (2022). Neurophysiological evidence for cognitive map formation during sequence learning. *eNeuro*, *9*(2), 1–24. https://doi.org/10.1523/eneuro.0361-21.2022

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

Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*(1–2), 181–211. https://doi.org/10.1016/S0004-3702(99)00052-1

Swaminathan, S., Dedieu, A., Raju, R. V., Shanahan, M., Lazaro-Gredilla, M., & George, D. (2023). *Schema-learning and rebinding as mechanisms of in-context learning and emergence*. arXiv preprint. https://doi.org/10.48550/arXiv.2307.01201

Talvitie, E. (2017). *Self-correcting models for model-based reinforcement learning* [Conference session]. Thirty-First AAAI Conference on Artificial Intelligence.

Tambini, A., Miller, J., Ehlert, L., Kiyonaga, A., & D'Esposito, M. (2023). *Structured memory representations develop at multiple time scales in hippocampal-cortical networks*. bioRxiv. https://doi.org/10.1101/2023.04.06.535935

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tomov, M. S., Yagati, S., Kumar, A., Yang, W., & Gershman, S. J. (2020). Discovery of hierarchical representations for efficient planning. *PLOS Computational Biology*, *16*(4), Article e1007594. https://doi.org/10.1371/journal.pcbi.1007594

van Hasselt, H., Madjiheurem, S., Hessel, M., Silver, D., Barreto, A., & Borsa, D. (2021). Expected eligibility traces. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(11), 9997–10005. https://doi.org/10.1609/aaai.v35i11.17200

Wang, S., Feng, S. F., & Bornstein, A. M. (2022). Mixing memory and desire: How memory reactivation supports deliberative decision-making. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(2), Article e1581. https://doi.org/10.1002/wcs.1581

Wang, X., Dworkin, J. D., Zhou, D., Stiso, J., Falk, E. B., Bassett, D. S., Zurn, P., & Lydon-Staley, D. M. (2021). Gendered citation practices in the field of communication. *Annals of the International Communication Association*, *45*(2), 134–153. https://doi.org/10.1080/23808985.2021.1960180

Warren, W. H. (2019). Non-Euclidean navigation. *Journal of Experimental Biology*, *222*(Suppl. 1), Article jeb187971. https://doi.org/10.1242/jeb.187971

Weber, E. U., & Johnson, E. J. (2006). Constructing preferences from memory. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (pp. 397–410). Cambridge University Press.

Whitehead, S. D., & Ballard, D. H. (1991). Learning to perceive and act by trial and error. *Machine Learning*, *7*, 45–83. https://doi.org/10.1023/A:1022619109594

Whittington, J. C., McCaffary, D., Bakermans, J. J., & Behrens, T. E. (2022). *How to build a cognitive map: Insights from models of the hippocampal formation*. arXiv preprint. https://doi.org/10.48550/arXiv.2202.01682

Whittington, J. C., Muller, T., Mark, S., Barry, C., & Behrens, T. (2018). Generalisation of structural knowledge in the hippocampal–entorhinal system. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31, pp. 8484–8495). Neural Information Processing Systems Foundation, Inc.

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. (2020). The tolman-eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, *183*(5), 1249–1263. https://doi.org/10.1016/j.cell.2020.10.024

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252

Wittkuhn, L., Krippner, L. M., & Schuck, N. W. (2022). *Statistical learning of successor representations is related to on-task replay*. bioRxiv. https://doi.org/10.1101/2022.02.02.478787

Wu, C. M., Schulz, E., & Gershman, S. J. (2019). *Generalization as diffusion: Human function learning on graphs*. bioRxiv. https://doi.org/10.1101/538934

Wu, C. M., Schulz, E., & Gershman, S. J. (2021). Inference and search on graph-structured spaces. *Computational Brain & Behavior*, *4*(2), 125–147. https://doi.org/10.1007/s42113-020-00091-x

Yoo, J., & Bornstein, A. M. (2024). *Temporal dynamics of model-based control reveal arbitration between multiple task representations*. https://doi.org/10.31234/osf.io/sgcy5

Zalesak, M., & Heckers, S. (2009). The role of the hippocampus in transitive inference. *Psychiatry Research: Neuroimaging*, *172*(1), 24–30. https://doi.org/10.1016/j.pscychresns.2008.09.008

Zhang, L., Yang, G., & Stadie, B. C. (2021). *World model as a graph: Learning latent landmarks for planning* [Conference session]. International Conference on Machine Learning.

Zhao, W. J., Richie, R., & Bhatia, S. (2022). Process and content in decisions from memory. *Psychological Review*, *129*(1), 73–106. https://doi.org/10.1037/rev0000318

Zhou, D., Cornblath, E. J., Stiso, J., Teich, E. G., Dworkin, J. D., Blevins, A. S., & Bassett, D. S. (2020). *Gender diversity statement and code notebook v1.0* (Version v1.0). Zenodo. https://doi.org/10.5281/zenodo.3672110

Zhu, G., Lin, Z., Yang, G., & Zhang, C. (2020). *Episodic reinforcement learning with associative memory* [Conference session]. International Conference on Learning Representations (ICLR) 2020.

Zhu, J., Sanborn, A., & Chater, N. (2018). Mental sampling in multimodal representations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31, pp. 5748–5759). Neural Information Processing Systems Foundation, Inc.

Received July 21, 2023
Revision received June 14, 2024
Accepted June 21, 2024 ■