Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Behavioral Sciences

Multi-step planning in the brain

Kevin J Miller[1,2] and Sarah Jo C Venditto[3]

Decisions in the natural world are rarely made in isolation. Each action that an organism selects will affect the future situations in which it finds itself, and those situations will in turn affect the future actions that are available. Achieving real-world goals often requires successfully navigating a sequence of many actions. An efficient and flexible way to achieve such goals is to construct an internal model of the environment, and use it to plan behavior multiple steps into the future. This process is known as multi-step planning, and its neural mechanisms are only beginning to be understood. Here, we review recent advances in our understanding of these mechanisms, many of which take advantage of multi-step decision tasks for humans and animals.

Addresses
[1] DeepMind, London, UK
[2] UCL Institute of Ophthalmology, University College London, London UK
[3] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA

Corresponding author: Miller, Kevin J (kevin.miller@ucl.ac.uk)

Current Opinion in Behavioral Sciences 2020, **38**:29–39

This review comes from a themed issue on **Computational cognitive neuroscience**

Edited by **Geoff Schoenbaum** and **Angela J Langdon**

https://doi.org/10.1016/j.cobeha.2020.07.003

2352-1546/© 2020 Published by Elsevier Ltd.

Introduction: multi-step planning in the brain

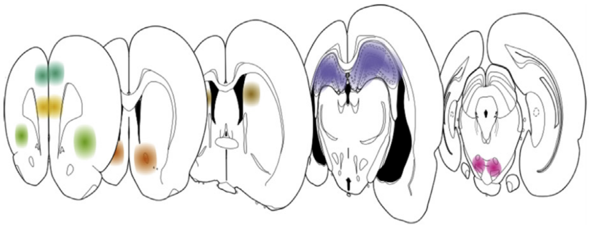
Humans and animals construct internal models of their environments, and use these models to inform behavior [1]. This capacity is known as *planning*, and it allows an organism to direct its behavior flexibly towards different possible goals. Planning is especially useful when achieving a goal requires a sequence of many actions, a situation we refer to as 'multi-step planning'. Recent years have seen a surge of progress towards understanding the neural mechanisms of multi-step planning. This progress was enabled, in large part, by the development about a decade ago of a variety of multi-step decision tasks for human subjects [2–6, reviewed by Ref. 7], expanding on older work that largely relied upon variants of a single task [8,9]. This behavioral toolkit has since continued to adapt and expand [10–12,13••,14,15], allowing increasingly detailed behavioral and neural study of planning. One of these multi-step decision tasks has been modified for animals, including both rodents [16••,17••,18•,19••,20•] and non-human primates [21,22•], greatly expanding the experimental toolkit that can be brought to bear on the study of multi-step planning in the brain. Here, we review some of what has been learned by these studies about the neural mechanisms of planning in multi-step tasks.

This work has built upon, and increasingly intertwined with, several other lines of neuroscience research. A first line investigates how animals learn the structure of their environments, proposing that this knowledge consists of a network of associations between various stimuli, actions, and outcomes. Multi-step planning can be thought of as a process that combines chains of such associations in order to guide behavior. A second line of work investigates the neural mechanisms of foraging for food. Foraging tasks frequently involve decisions which affect not just the immediate expected outcome, but also the future state of the world, and it has increasingly been recognized that animals may deploy multi-step planning in order to solve them. A final line of work investigates the role of the hippocampus in spatial navigation. This work is guided by the idea that the hippocampus represents a 'cognitive map' representing the spatial layout of the external world. Multi-step planning can be thought of as a process that uses this map to guide an extended sequence of behavior towards a distant goal. Recent work using multi-step decision tasks suggests that many neural structures which contribute to associative learning, to food foraging, or to spatial navigation may have similar roles to play in multi-step planning as well (Figure 1).

Multi-step planning shares neural mechanisms with associative learning

Work on the neural mechanisms of planning builds upon a rich body of work investigating cognitive capacities upon which planning may depend. One of these is the ability to associate stimuli or actions with specific expected outcomes (e.g. walk south → arrive at lab; walk north → arrive at coffee shop). Behavior guided by such outcome-specific associations can be thought of as exercising a simple one-step form of planning. Researchers have developed several selective assays of this capacity and have used them extensively to identify and characterize the neural structures that support outcome-specific associations [23•]. A second cognitive capacity necessary for multi-step planning is 'inference': the ability to combine separately learned associations in order to form new associations between items that may never have been encountered together before (e.g. walk north → arrive at

Figure 1



	Related Functions	Multi-Step Tasks
Anterior Cingulate	• Encodes foraging-related task variables [69-71, 73]	• Rodents: Encodes several planning-related variables; silencing impairs behavior [19]. • Humans: Encodes planning-related variables [13,72,73]
Prelimbic Cortex	• Silencing impairs learning of action-outcome associations [39-42]	• Possible homolog dlPFC: Disruption impairs planning [44]; activity correlates with planning difficulty [45, 46], state prediction error [6,47]. • Possible homolog vmPFC: Activity signals planning-informed value [2,6,47,49].
Orbitofrontal Cortex	• Encodes expected outcomes [33] • Silencing impairs inference [27] and use of outcome-specific associations [28-30]	• Rodents: Silencing impairs planning [16]. Dopamine content correlates with planning [17]. May specifically support learning [37].
Ventral Striatum	• Activity correlates with stimulus-outcome information [56] • Lesions impair stimulus-outcome behaviors [57, 58]	• Rodents: Dopamine content correlates with planning performance [17]. • Humans: Dopamine content correlates with planning [59]. Activity correlates with planning-informed prediction errors [2,3].
Dorsomedial Striatum	• Silencing impairs use of action-outcome associations [52]	• Humans: Activity correlates with planning difficulty [45], planning-informed value [5]; size correlates with planning performance [55].
Hippocampus	• Lesions do not impair classic assays of outcome-specific associations [91]. • Supports navigation and episodic memory [88-90] • Exhibits theta sequences during active behavior [94]. • Exhibits sharp-wave ripple sequences during rest [93].	• Rodents: Silencing impairs planning [16]. • Humans: Lesions impair planning [92] • Humans: Planning associated with increased theta power [100] • Humans: Offline sequences decoded with MEG associated with increased ripple power [107,115]
Midbrain Dopamine	• Contributes to learning of outcome-specific associations [24,61,62]	• Rodents and humans: Individual differences correlate with planning performance [17,59,65]. Precursor drugs enhance planning [66,67].

Current Opinion in Behavioral Sciences

Brain regions identified as playing a role in multi-step planning. Summary of roles in related cognitive processes, and of recent data from multi-step planning tasks. Coronal sections modified from Ref. [74].

coffee shop; place order → receive coffee; ∴ to obtain coffee, walk north). Researchers have developed assays of this capacity as well, and have used them to identify neural structures necessary for combining separately learned associations [e.g. Refs. 24–26]. Mechanisms for outcome-specific associations and for inference are universally present in theoretical accounts of multi-step planning (see Box 1). Recent work has borne out the idea that the neural structures which support outcome-specific associations and inference in associative learning tasks likely support planning in multi-step decision tasks as well.

Orbitofrontal cortex and model-based prediction

Regions of frontal cortex belonging to the orbital network (especially areas LO and AIv in rodents and areas 13 and 11 l in primates, often referred to as 'OFC' or 'lateral OFC') play an important role both in outcome-specific associations and in inference. Lesions or inactivations of OFC have been shown to impair performance on behavioral assays of these capacities in rodents [27–29] and to impair use of outcome-specific associations in nonhuman primates [30]. Recent data have extended these findings to humans as well, showing that disruption of OFC activity impairs performance on assays of outcome-specific associations and of inference [31,32]. Correspondingly, neural activity in the OFC has been shown to encode expected outcomes across many tasks and species [for review: Ref. 33], an observation which has recently been extended to encompass not only rewarding outcomes but also neutral stimuli [25,34]. Together, these studies provide strong reason to believe that OFC supports cognitive capacities that are necessary for planning, likely mediated by its representations of specific expected outcomes. They raise the possibility that it may play a similar role in support of multi-step planning.

Recent studies using multi-step decision tasks in rodents have supported this idea, demonstrating that the dopamine content of the OFC correlates with planning performance [17••], and that silencing neural activity in the OFC specifically impairs planning [16••]. If the OFC

Box 1 Computational theories of multi-step planning in the brain.

Modern computational neuroscience theories of planning typically use the theoretical framework of reinforcement learning (RL) [75], which considers the problem of an agent interacting with an environment in order to maximize rewards. An RL problem is defined by four components: a set of environmental states *s*; a set of available actions *a*; a reward function *R(s)* specifying the immediate reward available in state *s*; and a transition function *T(s'|s,a)* giving the probability that taking action α in state *s* will lead to state *s*'. A solution to an RL problem is defined by a policy function $\pi(a|s)$, giving the probability that the agent will take action α when in state *s*. For the purposes of RL, any method which has access to ***T*** and ***R*** (or approximations of them), and which uses them to help construct or improve a policy, is referred to as a 'planning' method [75]. A variety of planning methods have been adapted from machine learning to serve as theories of planning in the brain.

1) **Tree search.** A popular proposal is that humans and animals plan by constructing and evaluating a tree of possible future trajectories beginning from the current state [76]. Expanding the tree involves selecting a state within the tree and an action, using ***T*** and ***R*** to determine the likely subsequent states and their rewards, and adding these as new entries to the tree. Information from throughout the tree is used to estimate the long-term expected reward of each currently available action. Practical tree search methods typically do not search from scratch each time a decision must be made, but rather cache estimates of the values of different states, and use these estimates in future searches: either to evaluate states at the periphery of the tree or to guide how the tree should be expanded [4,15]. A variant of tree search proposes building not just a forward tree from the current state, but also a backward tree from a particular goal state [77].
2) **Learning from simulated experience.** Another proposal is that humans and animals use internal models to generate simulated behavioral trajectories, comprising sequences of states, actions, and rewards, and that these in turn provide training input to a separate model-free learning system [78]. The model-free system is responsible for maintaining the behavioral policy, and it improves this policy using the simulated experience as it would using genuine experience coming from the external world. Learning from simulated experience is most often proposed as a strategy for 'offline planning', which is directed not at guiding behavior in a particular moment, but in consolidating model-based knowledge into a behavioral policy that will be generally useful in the future.
3) **Dynamic programming.** One way to behave optimally on an RL problem is to compute the 'optimal value function', giving the best possible long-term expected reward that can be achieved from each state, and to always select the action that leads to the best-scoring state according to this metric. The optimal value function ***V**** for a particular RL problem is given by the Bellman optimality equation:

$$V^*(s) = R(s) + \gamma_\alpha^{\max} \sum_{s\prime} T(s\prime|s,\alpha)V^*(s\prime) \quad (1)$$

where γ is a 'discount factor' specifying the relative importance of immediate versus delayed rewards. Importantly, ***V**** is defined recursively — the value of each state depends on the values of other states, which may in turn depend on the value of the first. One strategy for computing it is to initialize some estimate ***V***, and then to repeatedly improve it by applying an update like:

$$V(s) \leftarrow R(s) + \gamma \max_\alpha \sum_{s\prime} T(s\prime|s,\alpha)V(s\prime) \quad (2)$$

termed a 'Bellman backup', which updates the estimated value for a particular state *s* based on the current estimated values for the others. If a sufficient number of updates are applied to all states, ***V*** is guaranteed to converge to ***V**** [75]. The number of iterations required for convergence is often very large, so practical planning methods seek to achieve a good-enough approximation using only a limited budget of updates [79••]. Planning by dynamic programming can be used at decision time, focusing on the current state and its likely successors, or for offline planning, focusing on states that are likely to be encountered in the future.

4) **Efficient re-planning.** Many real-world planning problems involve achieving new goals in well-learned environments. This observation has given rise to accounts of planning which involve a lengthy or costly computation that depends only on the transition function *T(s'|s,a)*, defining the environment's dynamics, followed by an inexpensive computation that depends on the reward function *R(s)*, defining the agent's current goals. The results of the costly computation can be cached, resulting in an agent that is able to efficiently plan towards new goals in familiar environments. One such strategy is the successor representation [80,81••,82], which caches a 'successor matrix' $S^\pi(s,s')$ giving the long-term expected future occupancy of state *s*', for an agent currently occupying state *s* and following behavioral policy π. An approximation to the optimal value function, ***V****, for a new goal can be obtained by multiplying this matrix by the reward function *R(s)* defining that goal:

$$V^{SR} = S^\pi R \quad (3)$$

VSR can be an effective approximation of ***V**** in some situations. It is limited, however, by the fact that the successor matrix ***Sπ*** depends strongly on the behavioral policy that was used to generate it (i.e. which states are likely to follow after others depends strongly on which actions the agent chooses to take). This can render ***VSR*** a very poor approximation in situations where the policy used to compute it is very different from the optimal policy [81••].

An alternative approach adopts methods from control theory [83], which substitute the original RL problem with a surrogate problem that can be solved without recursion. The Bellman equation for the surrogate problem reduces to

$$V^{linear} = \log(MPe^R) \quad (4)$$

where **M** and **P** can be computed from the transition function alone. If the brain were to cache these quantities for a familiar environment, it could combine them with a new reward function in a single linear computation and compute $\boldsymbol{V^{linear}}$ suitable for achieving new goals in that environment [84••]. Behavior based on $\boldsymbol{V^{linear}}$ has been shown to closely approximate optimal behavior on a variety of challenging planning problems [83,84••].

A final proposal [85••] uses the fact that certain functions of a matrix can be linearly computed if given the eigenvectors and eigenvalues of that matrix. If the brain were to cache these quantities for an adjacency matrix defining the structure of an environment, it could use them to efficiently compute an approximation of the distance between any two points in that environment and use this to plan shortest-path routes [85••].

5) **Neural networks.** The planning methods described above typically consider problems involving discrete states and fully known environmental dynamics (*T* and *R*). Many real-world planning problems feature environments that are continuous and/or partially unknown to the agent. The development of effective planning methods suitable for these problems remains an active research goal. Recent work in machine learning has resulted in a variety of novel methods, many of which integrate planning strategies like those described above with deep neural networks [e.g. Ref. 86]. The implications of this work for psychology and neuroscience present an important opportunity for future research [87•].

supports multi-step planning using its representations of expected outcomes, at least two computational roles are possible. The first is a role in choice: expected outcomes of available actions are compared, and the action with the best outcome is selected. The second is a role in learning: expected outcomes are compared to outcomes actually received, and future expectations are updated accordingly. Lesion studies using simpler tasks in primates had suggested that different subregions of frontal cortex may be involved in choosing and in learning [35,36], but it was unclear whether either of these was supported by OFC proper [37] and whether these functions generalized to multi-step tasks. A recent study addresses the role of the OFC specifically in multi-step planning using a rodent task that separates task items into some that are learned about, but not chosen between, and others that are chosen between, but not learned about [38••]. Activity in OFC correlates with the expected outcomes of items that are learned about, but not with those of items that are chosen between. Consistent with this, silencing OFC activity on a particular trial selectively impairs the influence of this expected outcome signal on learning but does not impair choosing. These results suggest that OFC in the rodent participates in planning not by driving choice directly, but by supporting the learning of associations upon which planning depends.

Prelimbic cortex and acquisition of action-outcome associations

Another region of the rodent frontal cortex that plays a role in associative learning is the prelimbic cortex (PL). Early results showed that lesions of PL impaired the ability of animals to use associations between actions and specific outcomes to guide behavior [39]. Later results revealed that this effect is specific to learning: once an action-outcome association has been formed, activity in PL is not necessary for the later use of that association to guide behavior [40–42]. Action-outcome associations play a prominent role in computational models of planning (e.g. the 'transition function', see Box 1), raising the possibility that PL plays a role in planning on multi-step decision tasks as well. While no data that we are aware of have been published from this region in a rodent multi-step task, studies with humans have implicated two other frontal regions that may be its functional homologs. The first is the dorsolateral prefrontal cortex (dlPFC) [43]. Disruption of neural activity in this region impairs planning performance [44], while neuroimaging experiments reveal greater activity in difficult versus easy planning problems [45,46]. Studies using tasks with probabilistic action-outcome relationships report that dlPFC activity correlates with a 'state prediction error', indicating the extent to which an outcome is surprising [6,47]. This signal is consistent with a role either in learning about action-outcome relationships or in forming a new plan when a previous one is interrupted. A second candidate functional homolog of PL is ventromedial prefrontal cortex (vmPFC, sometimes called 'medial OFC') [48]. Neural activity in this region is widely reported to represent expected reward probability; studies using multi-step planning tasks find that this representation occurs even when computing it requires planning [2,6,47,49], and that it codes related variables like the current distance to a goal [11,50]. Rodent studies in PL also have identified correlates of expected reward [51], though the necessary experiments to determine whether this signal can be informed by planning have not yet been performed. Taken together, these results are broadly consistent with the idea that a role of PL in associative learning in rodents — that of learning action-outcome relationships — may be performed by related regions in human PFC in support of multi-step planning. However, prefrontal cortex in primates is considerably larger and more differentiated than in rodents. Different planning-related signals have been found in many of its subregions [11,13••,46,47,49], which may play a variety of roles in planning. Future work is needed to clarify the nature of these roles as well as to understand the extent to which primate and rodent planning rely on similar mechanisms.

Dorsomedial striatum and action-outcome associations

The role of PL in learning action-outcome associations has recently been shown to depend specifically on its projections to the dorsomedial striatum (DMS) [42]. Silencing neural activity in the DMS itself impairs behaviors that depend on action-outcome associations [52], giving rise to the view that this region is the site where these associations themselves are stored [23•,48]. Recent work has begun to reveal in detail the circuit mechanisms by which action-outcome associations in the DMS are formed and modified [53,54]. Many theoretical accounts of planning rely on chaining together multiple action-outcome associations (see Box 1), raising the possibility that the associations stored in DMS may support planning. In multi-step decision tasks in humans, activity in the DMS is greater during difficult versus easy planning problems [45], and individual differences in planning performance correlate with differences in the physical size of the DMS [55]. Activity in human DMS has also been found to correlate with expected future reward, but only in situations where computing this requires forward planning [5]. Together, results from human multi-step tasks provide strong reason to believe that DMS plays some role in planning, though they have only begun to characterize the nature of this role. A strong hypothesis based on the role of DMS in associative learning is that it stores action-outcome associations that can be chained together to form multi-step plans.

Ventral striatum: stimulus-outcome associations and goal selection

In contrast with the dorsal striatum, activity in ventral striatum has been shown to correlate not with action-outcome, but with stimulus-outcome contingencies [56]. Correspondingly, lesions to ventral striatum impair stimulus-based, but not action-based, behavior [57] as well as an effect known as 'Pavlovian-instrumental transfer', in which the presentation of a stimulus that is associated with a particular outcome increases performance of an action that is associated with that same outcome [58]. These findings have given rise to the view that the ventral striatum both stores stimulus-outcome associations and also contributes to a process by which particular outcomes are selected as goals [23•,48]. In human neuroimaging experiments, activity in the ventral striatum is widely found to correlate with a 'reward prediction error' signal. In multi-step decision tasks, this signal has been shown to be sensitive to model-based information about the structure of the world [2,3]. Further evidence that ventral striatum plays a role in multi-step planning comes from a pair of recent studies that found that individual differences in dopamine levels in ventral, but not in dorsal, striatum correlate with individual differences in planning performance [17••,59]. These results provide strong reason to believe that ventral striatum plays a role both in associative learning and in multi-step planning, but further research is needed to clarify the relationship of these roles.

Dopamine and model-based learning

The neurotransmitter dopamine is thought to participate in learning, with a popular computational theory proposing that it signals a 'reward prediction error' that guides the learning of associations between stimuli or actions and general affective value [60]. This 'model-free' form of learning would not result in the outcome-specific associations needed to support multi-step planning. Recent data, however, have provided strong evidence that dopamine's role is not limited to model-free learning. Artificially inhibiting dopamine neurons, for example, prevents an animal from learning predictive relationships between sensory stimuli, while artificially activating them is sufficient to cause these associations to form Ref. [24]. Recordings of dopamine neurons in rodents and measurements of activity from dopaminergic regions in humans reveal an 'identity prediction error' signal that occurs when an expected reinforcer is substituted with an unexpected one of equal value [61,62]. These findings suggest that dopamine plays a role in learning associations between actions or stimuli and specific outcomes [63•,64]. This raises the possibility that it might support planning in multi-step tasks, which depends upon associations of this kind.

Experiments using multi-step tasks have provided significant evidence that this is the case. One study examined individual differences in dopamine-related genes in humans and found that these relate to planning, but not to model-free learning [65]. Others examined individual differences in endogenous dopamine levels and found that greater dopamine was related to enhanced planning, both in humans [59] and in rats [17••]. Other studies have found that artificially increasing dopamine content in humans using precursor drugs is sufficient to enhance planning [66,67]. Together, these results suggest that dopamine supports planning in multi-step tasks. A hypothesis based on the associative learning literature is that it does so indirectly, by driving learning of the outcome-specific representations upon which planning depends.

Multi-step planning shares neural mechanisms with foraging

Animals seek many goals, but two of the most common are evading predators and seeking food. The cognitive challenges of these goals may have driven the evolution of multi-step planning [68]. The neural mechanisms of naturalistic food-seeking have been studied using patch foraging tasks, in which an animal makes sequential decisions about whether to continue harvesting food from a depleting patch versus search for a new patch. These studies have identified a role for the dorsal anterior cingulate cortex (dACC) in encoding foraging-related variables [69,70]. The role of dACC has recently been reproduced in a recent study using a seemingly very different food-seeking task, in which monkeys use a joystick to control a digital predator pursuing digital prey [71•]. By their nature, these tasks involve sequences of actions whose consequences impact the state of the world in meaningful ways, and there is growing evidence that subjects may deploy planning to solve them [72•]. This is especially true in recent studies using more highly structured foraging-type tasks [13••,73], which have continued to find task-related variables in dACC.

A recent study addresses the role of the ACC in multi-step planning directly, using a task with dynamically variable transitions from first step action to second step state, in addition to the usual dynamically variable transitions from second step state to reward [19••]. The authors find that neural activity in ACC encodes a wide variety of task-related variables, including the current action-state relationship, but not the current state-reward relationship. Correspondingly, silencing of the ACC on a particular trial disrupts the influence of that trial's transition, but not that trial's reward, on future choice. These results suggest that ACC may play a role in representing the current structure of a dynamic environment and in deploying this knowledge to guide choice.

Together, these results suggest that common mechanisms may be at play in both sequential decisions involving foraging and in sequential decisions involving multi-step planning. Whether this indicates a role for foraging-specialized circuits in multi-step planning tasks (many of which can be viewed as structured food- or water-seeking tasks), a role for general-purpose planning circuits in foraging, or a common cognitive function underpinning tasks of both types, remains a topic for future research.

Multi-step planning shares neural mechanisms with navigation

Multi-step planning requires the ability to chain together sequences of action-outcome associations in order to guide multiple steps of behavior towards a potentially distant goal. Modern computational accounts of planning formalize this idea by using an internal model (the 'transition function' and 'reward function', see Box 1). Older psychological accounts refer to an analogous construct as the 'cognitive map' [1], which is thought to support episodic, spatial, and relational learning and memory as well as multi-step planning. The discovery of place cells highlighted the hippocampus, already known to support episodic learning and memory, as a region critical for the cognitive map, especially of physical space [88]. Decades of research that followed have fleshed out the richness of representations contained within the hippocampus and associated regions to support a complex and high-dimensional cognitive map underlying flexible behavior [89,90•].

While activity in the hippocampus does not seem required for the use of action-outcome associations in 'one-step' associative learning assays [91], it has recently been shown that it does contribute to planning in multi-step decision tasks [16••,92••]. This suggests that the hippocampus may be involved in chaining together multiple action-outcome associations, and more broadly that multi-step planning may share a common neural substrate, the hippocampal cognitive map, with spatial navigation.

Hippocampal sequences and multi-step planning

Many computational methods for multi-step planning — including tree search, learning from simulated experience, and dynamic programming (see Box 1) — involve considering a set of adjacent states in sequence. In the hippocampus, place cells typically report the current location of the animal, but they also have been shown to transiently 'sweep out' nonlocal paths through space in fast sequences of neural activity [93,94]. Two major types of sequences exist, theta sequences and sharp-wave ripple (SWR) sequences, both of which have been proposed to play a role in multi-step planning [79••,95•].

Theta sequences typically occur when an animal is moving, and represent trajectories beginning at or immediately behind the animal's current location and extending forward to expected future locations [94,96]. They have been shown both to reflect information about current goals [97] and to consider multiple possible future trajectories during moments when an animal pauses at a choice point [98]. Theta sequences have been proposed to reflect an online planning process, perhaps similar to tree search or to online dynamic programming [95•] (see Box 1). Recently, they have been shown to cycle quickly between different possible future trajectories even in the absence of overtly deliberate behavior [99]. This cycling involved not just cells that code for location, but also those that code for heading direction, suggesting that theta sequences are not limited to evaluating possible future locations, but rather possible future states more generally. Theta sequences have yet to be directly studied during a multi-step planning task, therefore it is currently unknown how these mechanisms extend to planning over multiple steps. However, a recent study [100••] demonstrated increased theta power in humans during multi-step spatial planning, consistent with the idea that theta sequences might be involved in this process.

Sharp-wave ripple sequences [93] typically occur when an animal is stationary and have been characterized both during rest and during brief pauses in active behavior [101]. They have been shown both to 'replay' trajectories that the animal has taken in the past, as well as to follow novel trajectories that the animal has never before taken [101–103]. Sharp-wave ripple sequences at rest have been proposed to reflect an offline planning process, perhaps similar to simulated experience or to offline dynamic programming (see Box 1) [79••,95•]. These computations are designed not to support immediate choice, but rather to consolidate model-based knowledge for efficient later use. Recent studies examining the link between sharp-wave ripple content and current goals have painted a complicated picture, with reports that in different situations, sharp-wave ripple sequences can be biased towards a current goal [101,103], away from a current goal [104], or even entirely random [105]. As with theta sequences, the role of sharp-wave ripple sequences in multi-step planning remains unclear, as they have yet to be directly studied during a multi-step planning task. Recent studies with human subjects have provided hints of a way forward. One such study reports signatures of offline replay, measured using fMRI, in human hippocampus [106], while another reports a specific link between whole-brain replay, measured using MEG, and ripple-band power in the hippocampus [107]. Further study, both using these tools in human subjects and using multi-step tasks in animal models, is needed to clarify the role of hippocampal sequences in multi-step planning.

Sequences beyond the hippocampus

For effective planning, it is not sufficient to simply represent sequences of possible future states. These sequences must also be evaluated and a behavioral policy updated. Internal decisions must also be made regarding which sequences of states should be considered. Theoretical accounts propose that the hippocampus does not perform all of these functions on its own, but rather that hippocampus interacts with other brain regions, including prefrontal cortex, ventral striatum, and OFC [108–110]. Recent data have added to the evidence for this view, with one study showing specific behavioral correlates for replay events that were coordinated between hippocampus and prefrontal cortex [111], and another study showing that silencing prefrontal cortex caused disruptions in theta sequences [112].

While these interactions have not been investigated using multi-step tasks in rodents, data from human subjects suggests that planning is accompanied by sequences beyond the hippocampus. One recent study reports that planning difficulty and also increased planning success are associated with increased coupling between the hippocampus and prefrontal cortex [50]. Another set of studies identifies reactivation of task state representations throughout the brain using fMRI during both online and offline planning [113,114••]. However, the methods used in these studies do not have the temporal resolution necessary to identify sequences of activity. A further set of studies instead uses MEG, which is able to track the temporal evolution of these representations, and they report that they are organized into sequences of task states [115••]. One recent study has used this technique to differentiate neural sequences that happened during task performance, found to correlate with flexible replanning of future choices, from those that happened during rest periods, found to correlate with consolidation of past choices and inflexible future decisions [116••].

While human studies do not provide the same spatial and temporal resolution as sequential activity studied in rodents, they provide intriguing links to rodent literature as well as hypotheses for possible patterns of sequences in rodents engaging in multi-step planning.

Predictive representations

Computational models that use sequences to plan (including tree search, simulated experience, and dynamic programming; see Box 1) are typically computationally expensive, requiring a large number of sequences to be considered in order to compute optimal behavior. Several recent computational accounts seek to address this issue by proposing mechanisms for computationally efficient approximate planning (see Box 1). Each of these relies upon a particular structured representation of the environment, and each points to evidence that this structured representation may be present within the hippocampus or the entorhinal cortex.

The first is the successor representation (SR) [81••,82], which proposes that the representation of each state should contain information about future states that are likely to follow it. This has the consequence that different states should be represented similarly if those states predict similar likely futures. A recent theoretical account proposes that hippocampal place cells carry this type of representation [117••], consistent both with long-standing observations from rodent electrophysiology as well as recent observations from human fMRI [118,119].

While methods that use the SR for efficient planning can be very flexible with respect to some types of changes in the environment, they can be extremely inflexible with respect to others [81••]. A pair of recent accounts [84••,85••] (see Box 1) proposes alternative methods for efficient approximate planning which may be more generally flexible. Both of these methods posit special representations of the structure of the environment, and both propose that these special representations may be present in the entorhinal cortex. The entorhinal cortex has been proposed to support representation and generalization of structured behavior in a variety of contexts [90•,120], but the idea that it is involved specifically in efficient approximate planning remains to be tested.

Conclusions and future directions

Over the last decade, work using multi-step decision tasks has begun to provide insight into neural mechanisms of multi-step planning. This work has revealed hints into the functions of a number of brain regions (Figure 1), but an integrated account of how the brain forms plans is still lacking. One particular challenge for future research is that planning likely involves interactions between a number of different brain regions, and understanding its neural mechanisms will likely require experiments investigating these interactions specifically. Another challenge is that there are a variety of different ways to plan (e.g. online/offline, sequence-based/efficient-approximate), and the brain may implement several of them to be deployed in different situations. If this is the case, future research will benefit greatly from new behavioral tools that can isolate planning in its various forms. In general, producing an integrated account of planning in the brain is likely to require the combination of both a sophisticated behavioral tools to characterize planning strategies and quantify changes caused by external perturbations, as well as a sophisticated neuroscience tools for measuring neural correlates of planning and for perturbing the brain in selective ways. Recent work has taken important steps in this direction: researchers working with human subjects have developed behavioral tasks that characterize and assay planning in increasingly detailed ways, while researchers working with animals have adapted a subset of these tasks, allowing the studies which measure and perturb the brain in increasingly precise ways.

Conflict of interest statement

Nothing declared.

Acknowledgements

We would like to thank Chris Summerfield, Kim Stachenfeld and Jan Balaguer for helpful comments on the manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1656466 to SJV. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. KJM was supported by DeepMind.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Tolman EC: **Cognitive maps in rats and men**. *Psychol Rev* 1948, **55**:189-208.

2. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model-based influences on humans' choices and striatal prediction errors**. *Neuron* 2011, **69**:1204-1215.

3. Simon DA, Daw ND: **Neural correlates of forward planning in a spatial decision task in humans**. *J Neurosci* 2011, **31**:5526-5539.

4. Huys QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP: **Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees**. *PLoS Comput Biol* 2012, **8**:e1002410.

5. Wunderlich K, Dayan P, Dolan RJ: **Mapping value based planning and extensively trained choice in the human brain**. *Nat Neurosci* 2012, **15**:786-791.

6. Gläscher J, Daw N, Dayan P, O'Doherty JP: **States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning**. *Neuron* 2010, **66**:585-595.

7. Dolan RJ, Dayan P: **Goals and habits in the brain**. *Neuron* 2013, **80**:312-325.

8. Shallice T: **Specific impairments of planning**. *Philos Trans R Soc Lond B Biol Sci* 1982, **298**:199-209.

9. Unterrainer JM, Owen AM: **Planning and problem solving: from neuropsychology to functional neuroimaging**. *J Physiol Paris* 2006, **99**:308-317.

10. Snider J, Lee D, Poizner H, Gepshtein S: **Prospective optimization with limited resources**. *PLoS Comput Biol* 2015, **11**:e1004501.

11. Balaguer J, Spiers H, Hassabis D, Summerfield C: **Neural mechanisms of hierarchical planning in a virtual subway network**. *Neuron* 2016, **90**:893-903.

12. Callaway F, Lieder F, Das P, Gul S, Krueger PM, Griffiths T: **A resource-rational analysis of human planning**. *Cogn Sci* 2018 . Available: *http://173.236.226.255/papers/Callaway_CogSci_2018.pdf*.

13. Kolling N, Scholl J, Chekroud A, Trier HA, Rushworth MFS: •• **Prospection, perseverance, and insight in sequential behavior**. *Neuron* 2018, **99**:1069-1082.e7.
Introduces a novel multi-step task for human subjects,which bridges a gap between previous multi-step decision task used to study planning and sequential behaviors used to study foraging. Identifies behavioral evidence of planning and neural evidence for the roles of several prefrontal regions.

14. van Opheusden B, Galbiati G, Bnaya Z, Li Y, Ma WJ: **Modeling decision tree search in a two-player game**. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* 2017:1254-1259.

15. Keramati M, Smittenaar P, Dolan RJ, Dayan P: **Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum**. *Proc Natl Acad Sci USA* 2016, **113(45)**:12868-12873.

16. Miller KJ, Botvinick MM, Brody CD: **Dorsal hippocampus** •• **contributes to model-based planning**. *Nat Neurosci* 2017, **20**:1269-1276.
Presents a multi-step decision task for rodents, and demonstrates that their behavior is consistent with planning. Demonstrates that silencing either OFC or dorsal hippocampus selectively impairs planning.

17. Groman SM, Massi B, Mathias SR, Curry DW, Lee D, Taylor JR: •• **Neurochemical and behavioral dissections of decision-making in a rodent multistage task**. *J Neurosci* 2019, **39**:295-306.
Presents a multi-step decision task for rats. Quantifies differences in the extent to which indivudals use a planning strategy vs. a model-free strategy, and relates these to individual differences in dopamine concentration in different brain regions.Finds that planning correlates with dopamine concentration in the OFC and ventral striatum, but not in the dorsal striatum.

18. Dezfouli A, Balleine BW: **Learning the structure of the world: the** • **adaptive nature of state-space and action representations in multi-stage decision-making**. *PLoS Comput Biol* 2019, **15**: e1007334.
Presents a rodent adaptation of a multi-step decision task. Describes the timecourse of task learning, as animals increasingly discover and then exploit the structure of the task

19. Akam T, Rodrigues-Vaz I, Marcelo I, Zhang X, Pereira M, Oliveira R •• *et al.*: **Anterior cingulate cortex represents action-state predictions and causally 1mediates model-based reinforcement learning in a two-step decision task**. *bioRxiv* 2020 http://dx.doi.org/10.1101/126292.
Presents a multi-step decision task for mice, using a design in which both rewrad probabilities and also state transition probabilities change unpredictably within individual sessions. Investigates the role of ACC, and finds that it encodes a variety of task-related variables, including an estimate of the current stae transition probabilities. Finds the silencing ACC specifically imparis the influence of state transitions on future behavior.

20. Hasz BM, David Redish A: **Deliberation and procedural** • **automation on a two-step task for rats**. *Front Integr Neurosci* 2018, **12**:30.
Presents a rodent adaptation of a multi-step decision task, involving navigation through a large-scale maze. Finds that behavior reflects a mixture of planning and model-free strategies, and relates these to "vicarious trial-and-error" behaviors, which have separately been shown to relate to hippocampal theta sequences.

21. Ford C, Wallis J: **Dissociating model-based and model-free reinforcement learning in a non-human primate model**. *Reinf Learn Decis Mak* 2019. Montreal, Canada.

22. Miranda B, Malalasekera N, Behrens T, Dayan P, Kennerley S: • **Combined model-free and model-sensitive reinforcement learning in non-human primates**. *PLOS Comp Bio* 2020, **16**: e1007944.
Presents a multi-step decision task adapted for nonhuman primates, along with a careful characterization of the behavioral strategies that they adopt. Finds that behavior is strongly influenced by a planning strategy, with weaker influences of other strategies including model-free learning.

23. Balleine BW: **The meaning of behavior: discriminating reflex** • **and volition in the brain**. *Neuron* 2019, **104**:47-62.
Reviews and synthesizes the large number of studies which use associative learning assays to probe the neural mechanisms that support action-outcome associations.

24. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL *et al.*: **Dopamine transients are sufficient and necessary for acquisition of model-based associations**. *Nat Neurosci* 2017, **20**:735-742.

25. Sadacca BF, Wied HM, Lopatina N, Saini GK, Nemirovsky D, Schoenbaum G: **Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task**. *eLife* 2018, **7**.

26. Wang F, Schoenbaum G, Kahnt T: **Interactions between human orbitofrontal cortex and hippocampus support model-based inference**. *PLoS Biol* 2020, **18**:e3000578.

27. Jones JL, Esber GR, McDannald MA, Gruber AJ, Hernandez A, Mirenzi A *et al.*: **Orbitofrontal cortex supports behavior and learning using inferred but not cached values**. *Science* 2012, **338**:953-956.

28. Gallagher M, McMahan RW, Schoenbaum G: **Orbitofrontal cortex and representation of incentive value in associative learning**. *J Neurosci* 1999, **19**:6610-6614.

29. Gremel CM, Costa RM: **Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions**. *Nat Commun* 2013, **4**:2264.

30. Rudebeck PH, Saunders RC, Prescott AT, Chau LS, Murray EA: **Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating**. *Nat Neurosci* 2013, **16**:1140-1145.

31. Howard JD, Reynolds R, Smith DE, Voss JL, Schoenbaum G, Kahnt T: **Targeted stimulation of human orbitofrontal networks disrupts outcome-guided behavior**. *Curr Biol* 2020, **30**:490-498.e4.

32. Wang F, Howard JD, Voss JL, Schoenbaum G, Kahnt T: **Targeted stimulation of an orbitofrontal network disrupts decisions based on inferred, not experienced, outcomes**. *bioRxiv* 2020 http://dx.doi.org/10.1101/2020.04.24.059808.

33. Rudebeck PH, Murray EA: **The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes**. *Neuron* 2014, **84**:1143-1156.

34. Pauli WM, Gentile G, Collette S, Tyszka JM, O'Doherty JP: **Evidence for model-based encoding of Pavlovian contingencies in the human brain**. *Nat Commun* 2019, **10**:1-11.

35. Noonan MP, Walton ME, Behrens TEJ, Sallet J, Buckley MJ, Rushworth MFS: **Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex**. *Proc Natl Acad Sci USA* 2010, **107**:20547-20552.

36. Noonan MP, Chau BKH, Rushworth MFS, Fellows LK: **Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans**. *J Neurosci* 2017, **37**:7023-7035.

37. Rudebeck PH, Saunders RC, Lundgren DA, Murray EA: **Specialized representations of value in the orbital and ventrolateral prefrontal cortex: desirability versus availability of outcomes**. *Neuron* 2017, **95**:1208-1220.e5.

38. •• Miller KJ, Botvinick MM, Brody CD: **Value representations in orbitofrontal cortex drive learning, not choice**. *bioRxiv* 2020. Available: https://www.biorxiv.org/content/10.1101/245720v4.abstract.

Uses a multi-step decision task for rats to investigate the role of the OFC. Finds that neural activity in the OFC encodes expected outcomes in a way suitable to guide learning of the environmental structure, but less suitable for driving immediate choice. Also finds that silencing the OFC has behavioral effects consistent with impaired learning, but not with impaired decision-making.

39. Balleine BW, Dickinson A: **Goal-directed instrumental action: contingency and incentive learning and their cortical substrates**. *Neuropharmacology* 1998, **37**:407-419.

40. Ostlund SB, Balleine BW: **Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning**. *J Neurosci* 2005, **25**:7763-7770.

41. Hart G, Bradfield LA, Balleine BW: **Prefrontal corticostriatal disconnection blocks the acquisition of goal-directed action**. *J Neurosci* 2018, **38**:1311-1322.

42. Hart G, Bradfield LA, Fok SY, Chieng B, Balleine BW: **The bilateral prefronto-striatal pathway is necessary for learning new goal-directed actions**. *Curr Biol* 2018:2218-2229.e7.

43. Uylings HBM, Groenewegen HJ, Kolb B: **Do rats have a prefrontal cortex?** *Behav Brain Res* 2003, **146**:3-17.

44. Smittenaar P, FitzGerald THB, Romei V, Wright ND, Dolan RJ: **Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans**. *Neuron* 2013, **80**:914-919.

45. Dagher A, Owen AM, Boecker H, Brooks DJ: **Mapping the network for planning: a correlational PET activation study with the Tower of London task**. *Brain* 1999, **122**:1973-1987.

46. Wagner G, Koch K, Reichenbach JR, Sauer H, Schlösser RGM: **The special involvement of the rostrolateral prefrontal cortex in planning abilities: an event-related fMRI study with the Tower of London paradigm**. *Neuropsychologia* 2006, **44**:2337-2347.

47. Lee SW, Shimojo S, O'Doherty JP: **Neural computations underlying arbitration between model-based and model-free learning**. *Neuron* 2014, **81**:687-699.

48. Balleine BW, O'Doherty JP: **Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action**. *Neuropsychopharmacology* 2010, **35**:48-69.

49. Kim D, Park GY, O′Doherty JP, Lee SW: **Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning**. *Nat Commun* 2019, **10**:1-14.

50. Kaplan R, King J, Koster R, Penny WD, Burgess N, Friston KJ: **The neural representation of prospective choice during spatial planning and decisions**. *PLoS Biol* 2017, **15**:e1002588.

51. Sul JH, Kim H, Huh N, Lee D, Jung MW: **Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making**. *Neuron* 2010, **66**:449-460.

52. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW: **The role of the dorsomedial striatum in instrumental conditioning**. *Eur J Neurosci* 2005, **22**:513-523.

53. Peak J, Chieng B, Hart G, Balleine B: **Striatal direct and indirect pathway neurons differentially control the encoding and updating of goal-directed learning**. *bioRxiv* 2020. Available: http://biorxiv.org/content/10.1101/2020.02.18.955385v1.full.

54. Matamales M, McGovern AE, Mi JD, Mazzone SB, Balleine BW, Bertran-Gonzalez J: **Local D2- to D1-neuron transmodulation updates goal-directed learning in the striatum**. *Science* 2020, **367**:549-555.

55. Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J *et al.*: **Disorders of compulsivity: a common bias towards learning habits**. *Mol Psychiatry* 2015, **20**:345-352.

56. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ: **Dissociable roles of ventral and dorsal striatum in instrumental conditioning**. *Science* 2004, **304**:452-454.

57. Rothenhoefer KM, Costa VD, Bartolo R, Vicario-Feliciano R, Murray EA, Averbeck BB: **Effects of ventral striatum lesions on stimulus-based versus action-based reinforcement learning**. *J Neurosci* 2017, **37**:6902-6914.

58. Corbit LH, Balleine BW: **The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell**. *J Neurosci* 2011, **31**:11786-11794.

59. Deserno L, Huys QJM, Boehme R, Buchert R, Heinze H-J, Grace AA *et al.*: **Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making**. *Proc Natl Acad Sci U S A* 2015, **112**:1595-1600.

60. Schultz W, Dayan P, Montague PR: **A neural substrate of prediction and reward**. *Science* 1997, **275**:1593-1599.

61. Takahashi YK, Batchelor HM, Liu B, Khanna A, Morales M, Schoenbaum G: **Dopamine neurons respond to errors in the**

prediction of sensory features of expected rewards. *Neuron* 2017, **95**:1395-1405.e3.

62. Howard JD, Kahnt T: **Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex**. *Nat Commun* 2018, **9**:1611.

63. • Langdon AJ, Sharpe MJ, Schoenbaum G, Niv Y: **Model-based predictions for dopamine**. *Curr Opin Neurobiol* 2018, **49**:1-7.
Reviews growing evidence that dopamine plays a role in learning outcome-specific associations, in the context of associative learning tasks. It may play a similar role in multi-step planning as well.

64. Gardner MPH, Schoenbaum G, Gershman SJ: **Rethinking dopamine as generalized prediction error**. *Proc R Soc B* 2018, **285**:20181645.

65. Doll BB, Bath KG, Daw ND, Frank MJ: **Variability in dopamine genes dissociates model-based and model-free reinforcement learning**. *J Neurosci* 2016, **36**:1211-1222.

66. Wunderlich K, Smittenaar P, Dolan RJ: **Dopamine enhances model-based over model-free choice behavior**. *Neuron* 2012, **75**:418-424.

67. Sharp ME, Foerde K, Daw ND, Shohamy D: **Dopamine selectively remediates “model-based” reward learning: a computational approach**. *Brain* 2016, **139**:335-364.

68. Mugan U, MacIver MA: **Spatial planning with long visual range benefits escape from visual predators in complex naturalistic environments**. *Nat Comm* 2020, **11**:3057.

69. Hayden BY, Pearson JM, Platt ML: **Neuronal basis of sequential foraging decisions in a patchy environment**. *Nat Neurosci* 2011, **14**:933-939.

70. Kolling N, Behrens TEJ, Mars RB, Rushworth MFS: **Neural mechanisms of foraging**. *Science* 2012, **336**:95-98.

71. • Yoo SBM, Tu JC, Piantadosi ST, Hayden BY: **The neural basis of predictive pursuit**. *Nat Neurosci* 2020, **23**:252-259.
Introduces a novel task for primates that involves using a joystick to control a virtual predator pursuing virtual prey, which requires many actions to be made in sequence before a goal is achieved. Includes a detailed analysis of behavior and of neural signals in the ACC, which encode a variety of task-related information.

72. • Kolling N, Akam T: **(Reinforcement?) Learning to forage optimally**. *Curr Opin Neurobiol* 2017, **46**:162-169.
Review that explicitly draws a connection between planning, as studied using multi-step decision tasks, and sequential decision-making as studied in foraging.

73. Wittmann MK, Kolling N, Akaishi R, Chau BKH, Brown JW, Nelissen N *et al.*: **Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex**. *Nat Commun* 2016, **7**:12327.

74. Paxinos G, Watson C: *The Rat Brain in Stereotaxic Coordinates: Hard Cover Edition*. Elsevier; 2006.

75. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction*. edn 2. Cambridge, MA, USA: MIT Press; 2017.

76. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *Nat Neurosci* 2005, **8**:1704-1711.

77. Afsardeir A, Keramati M: **Behavioural signatures of backward planning in animals**. *Eur J Neurosci* 2018, **47**:479-487.

78. Gershman SJ, Markman AB, Otto AR: **Retrospective revaluation in sequential decision making: a tale of two systems**. *J Exp Psychol Gen* 2014, **143**:182-194.

79. •• Mattar MG, Daw ND: **Prioritized memory access explains planning and hippocampal replay**. *Nat Neurosci* 2018, **21**:1609-1617.
Considers the problem of efficiently allocating limited computational resources to solve planning problems. Observes that these solutions often involve evaluating sequences of adjacent states, and relates them to a variety of experimental data on neural sequences in the hippocampus.

80. Dayan P: **Improving generalization for temporal difference learning: the successor representation**. *Neural Comput* 1993, **5**:613-624.

81. •• Russek EM, Momennejad I, Botvinick MM, Gershman SJ, Daw ND: **Predictive representations can link model-based reinforcement learning to model-free mechanisms**. *PLoS Comput Biol* 2017, **13**:e1005768.
Considers the successor representation as a mechanisms for computationally-efficient approximate planning. Extends this idea in several ways, and ildentifies situations in which it does and does not provide useful flexibility at low computational cost.

82. Gershman SJ: **The successor representation: its computational logic and neural substrates**. *J Neurosci* 2018, **38**:7193-7200.

83. Todorov E: **Efficient computation of optimal actions**. *Proc Natl Acad Sci U S A* 2009, **106**:11478-11483.

84. •• Piray P, Daw ND: **Linear reinforcement learning: flexible reuse of computation in planning, grid fields, and cognitive control**. *bioRxiv* 2020 http://dx.doi.org/10.1101/856849.
Adapts methods from control theory, designed explicitly for computatoinally-efficient approximate planning, as a theory of planning in the brain. Points out several ways in which they overcome limitations of the successor representation strategy. Draws connections between the types of representations these strategies use and those found in entorhinal cortex, as well as between the overall architecture of the strategy and the psychology of cognitive control.

85. •• Baram AB, Muller TH, Whittington JCR, Behrens TEJ: **Intuitive planning: global navigation through cognitive maps based on grid-like codes**. *bioRxiv* 2018. Available: https://www.biorxiv.org/content/10.1101/421461v1.abstract.
Proposes a version of computationally efficient approximate planning that efficiently estimates the distance between any two locations using representations similar to those found in entorhinal cortex.

86. Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S *et al.*: **Mastering atari, go, chess and shogi by planning with a learned model**. *arXiv [cs.LG]* 2019 . Available: In: *http://arxiv.org/abs/1911.08265*.

87. • Hamrick JB: **Analogues of mental simulation and imagination in deep learning**. *Curr Opin Behav Sci* 2019, **29**:8-16.
Reviews recent advances in artificial intelligence techniques using deep neural networks for multi-step planning, and their implications for cognitive science research.

88. O’Keefe J, Nadel L: *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press; 1978.

89. Epstein RA, Patai EZ, Julian JB, Spiers HJ: **The cognitive map in humans: spatial navigation and beyond**. *Nat Neurosci* 2017, **20**:1504-1513.

90. • Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL *et al.*: **What is a cognitive map? Organizing knowledge for flexible behavior**. *Neuron* 2018, **100**:490-509.
Reviews literature on conitive maps, both in spatial navigation and in nonspatial tasks. Makes the case that both type sof maps can be understood in a common framework.

91. Corbit LH, Ostlund SB, Balleine BW: **Sensitivity to instrumental contingency degradation is mediated by the entorhinal cortex and its efferents via the dorsal hippocampus**. *J Neurosci* 2002, **22**:10976-10984.

92. •• Vikbladh OM, Meager MR, King J, Blackmon K, Devinsky O, Shohamy D *et al.*: **Hippocampal contributions to model-based planning and spatial memory**. *Neuron* 2019, **102**:683-693.e4.
Reports that hippocampal lesions impair multi-step planning in human subjects, and describes evidence that spatial navigation and multi-step planning share a common neural substrate.

93. Foster DJ: **Replay comes of age**. *Annu Rev Neurosci* 2017, **40**:581-602.

94. Redish AD: **Vicarious trial and error**. *Nat Rev Neurosci* 2016, **17**:147-159.

95. • Pezzulo G, Donnarumma F, Maisto D, Stoianov I: **Planning at decision time and in the background during spatial navigation**. *Curr Opin Behav Sci* 2019, **29**:69-76.
Reviews the current evidence that hippocampal sequences have properties suitable for a role in multi-step planning. Makes the case that different types of sequences may support planning at decision time and offline planning.

96. Skaggs WE, McNaughton BL, Wilson MA, Barnes CA: **Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences**. *Hippocampus* 1996, **6**:149-172.

97. Wikenheiser AM, Redish AD: **Hippocampal theta sequences reflect current goals**. *Nat Neurosci* 2015, **18**:289-294.

98. Johnson A, Redish AD: **Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point**. *J Neurosci* 2007, **27**:12176-12189.

99. Kay K, Chung JE, Sosa M, Schor JS, Karlsson MP, Larkin MC *et al.*: **Constant sub-second cycling between representations of possible futures in the hippocampus**. *Cell* 2020, **180**:552-567.e25.

100. •• Kaplan R, Tauste Campo A, Bush D, King J, Principe A, Koster R *et al.*: **Human hippocampal theta oscillations reflect sequential dependencies during spatial planning**. *Cogn Neurosci* 2019:1-10.

Uses MEG in human subjects during a multi-step planning task, and find a correlation between planning and hippocampal theta power. This provides evidence for the idea that theta sequences support prospective planning.

101. Pfeiffer BE, Foster DJ: **Hippocampal place-cell sequences depict future paths to remembered goals**. *Nature* 2013, **497**:74-79.

102. Gupta AS, van der Meer MAA, Touretzky DS, Redish AD: **Hippocampal replay is not a simple function of experience**. *Neuron* 2010, **65**:695-705.

103. Ólafsdóttir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ: **Hippocampal place cells construct reward related sequences through unexplored space**. *eLife* 2015, **4**:e06063.

104. Carey AA, Tanaka Y, van der Meer MAA: **Reward revaluation biases hippocampal replay content away from the preferred outcome**. *Nat Neurosci* 2019, **22**:1450-1459.

105. Stella F, Baracskay P, O'Neill J, Csicsvari J: **Hippocampal reactivation of random trajectories resembling brownian diffusion**. *Neuron* 2019, **102**:450-461.e7.

106. Schuck NW, Niv Y: **Sequential replay of nonspatial task states in the human hippocampus**. *Science* 2019, **364**.

107. Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TEJ: **Human replay spontaneously reorganizes experience**. *Cell* 2019, **178**:640-652.e14.

108. Zielinski MC, Tang W, Jadhav SP: **The role of replay and theta sequences in mediating hippocampal-prefrontal interactions for memory and cognition**. *Hippocampus* 2017, **30**:60-72.

109. Stoianov IP, Pennartz CMA, Lansink CS, Pezzulo G: **Model-based spatial navigation in the hippocampus-ventral striatum circuit: a computational analysis**. *PLoS Comput Biol* 2018, **14**:e1006316.

110. Wikenheiser AM, Schoenbaum G: **Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex**. *Nat Rev Neurosci* 2016, **17**:513-523.

111. Shin JD, Tang W, Jadhav SP: **Dynamics of awake hippocampal-prefrontal replay for spatial learning and memory-guided decision making**. *Neuron* 2019, **104**:1110-1125.e7.

112. Schmidt B, Duin AA, Redish AD: **Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision making**. *J Neurophysiol* 2019, **121**:1981-2000.

113. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND: **Model-based choices involve prospective neural activity**. *Nat Neurosci* 2015, **18**:767-772.

114. •• Momennejad I, Otto AR, Daw ND, Norman KA: **Offline replay supports planning in human reinforcement learning**. *eLife* 2018, **7**.

Reports evidence of offline reactivation in human fMRI data, and that this reactivation correlates with behavioral measures of planning. Provides evidence for offline planning in humans, and identifies specific correlates in several brain regions, including ACC, OFC, hippocampus, is preceded by uncertainty, and is coupled with activity in hippocampus and anterior cingulate.

115. •• Kurth-Nelson Z, Economides M, Dolan RJ, Dayan P: **Fast sequences of non-spatial state representations in humans**. *Neuron* 2016, **91**:194-204.

Introduces a novel approach to measuring replay in humans using MEG, and uses it to reveal backwards replay of abstract task states in a multi-step planning task. This suggests that humans deploy backward planning to solve this task.

116. •• Eldar E, Lièvre G, Dayan P, Dolan RJ: **The roles of online and offline replay in planning**. *eLife* 2020, **9**:e56911.

Measures replay using MEG during a multi-step task, identifying replay event both during task performance and during rest periods. Identifies separate behavioral correlates of each of these, which suggest roles in online and offline planning, respectively.

117. •• Stachenfeld KL, Botvinick MM, Gershman SJ: **The hippocampus as a predictive map**. *Nat Neurosci* 2017, **20**:1643-1653.

Relates the successor representation strategy for computationally-efficient approximate planning (see box) to firing patterns of neurons in the hippocampus and entorhinal cortex. Makes the case that the brain's navigation system may use this type of predictive code.

118. Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM: **Statistical learning of temporal community structure in the hippocampus**. *Hippocampus* 2016, **26**:3-8.

119. Garvert MM, Dolan RJ, Behrens TEJ: **A map of abstract relational knowledge in the human hippocampal–entorhinal cortex**. *eLife* 2017, **6**:e17086.

120. Baram AB, Muller TH, Nili H, Garvert M, Behrens TEJ: **Entorhinal and ventromedial prefrontal cortices abstract and generalise the structure of reinforcement learning problems**. *bioRxiv* 2019. Available: https://www.biorxiv.org/content/10.1101/827253v1.abstract.