

Planning at decision time and in the background during spatial navigation

Giovanni Pezzulo, Francesco Donnarumma, Domenico Maisto and Ivilin Stoianov



Planning is the model-based approach to solving control problems. The hallmark of planning is the endogenous generation of dynamical representations of future states, like goal locations, or state sequences, like trajectories to the goal location, using an internal model of the task. We review recent evidence of model-based planning processes and the representation of future goal states in the brain of rodents and humans engaged in spatial navigation tasks. We highlight two distinct but complementary usages of planning as identified in artificial intelligence: ‘at decision time’, to support goal-directed choices and sequential memory encoding, and ‘in the background’, to learn behavioral policies and to optimize internal models. We discuss how two kinds of internally generated sequences in the hippocampus – theta and SWR sequences – might participate in the neuronal implementation of these two planning modes, thus supporting a flexible model-based system for adaptive cognition and action.

Address

Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

Corresponding author: Pezzulo, Giovanni (giovanni.pezzulo@istc.cnr.it)

Current Opinion in Behavioral Sciences 2019, 29:69–76

This review comes from a themed issue on **Artificial intelligence**

Edited by **Matt Botvinick** and **Sam Gershman**

<https://doi.org/10.1016/j.cobeha.2019.04.009>

2352-1546/© 2018 Elsevier Inc. All rights reserved.

Introduction

The ability to imagine and plan for the future is increasingly recognized as a central aspect of brain function; yet, its computational and neuronal principles are poorly understood [1–3]. Artificial Intelligence (AI) theories describe planning as the *model-based* approach to solving control problems [4–7]. Model-based control methods select actions on the basis of predictions of action outcomes, which are generated using an internal model of the task; and are distinct from model-free control methods, which select actions based on learned stimulus-response mechanisms, without looking at the future [8–11].

A hallmark of planning is using an internal model – akin to what Tolman [12] called a ‘cognitive map’ – to generate sequences of future actions or states, for example, sequences of actions that achieve an explicit goal like ‘being in location X’ or an implicit goal like ‘maximizing reward over some time horizon’. These cognitive operations are quite elusive to study, as they are largely endogenous (i.e. not stimulus based) and covert (i.e. not immediately apparent from behavior).

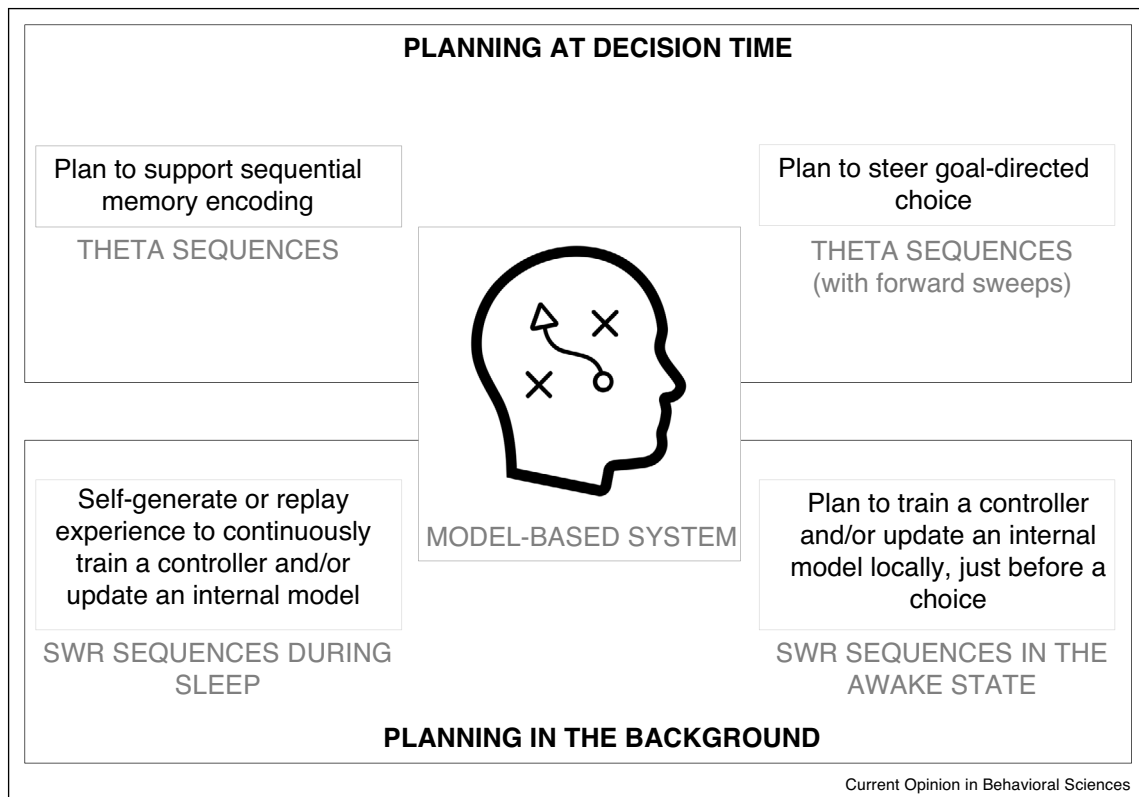
Here, we review recent studies that have unveiled planning dynamics and representations of future states and goals when rodents are engaged in spatial navigation; while also briefly mentioning analogous human studies. We organize the review around two main planning modes that have been identified in AI research — planning ‘at decision time’ versus ‘in the background’ (see [Figure 1](#)).

Planning at decision time: supporting goal-directed choices and sequence encoding

In AI, the most classical use of model-based planning is supporting action selection at decision time [10,13]. Planning at decision time permits agents to predict the outcomes of alternative action courses, thus supporting the flexible deliberation between them. This form of deliberative, goal-directed choice is especially effective when an animal faces novel situations and needs to generalize beyond its previous experience [14,15]. Long ago, Tolman described an intriguing form of deliberation during rodent spatial navigation: ‘vicarious trial and error’ (VTE) behavior [16]. When rodents reach decision points, such as junction points in a maze, they sometimes pause, orient and reorient back and forth, between alternative spatial trajectories or goals, as if they were deliberating between them (see also Ref. [17] for evidence of VTE in humans). VTE occurs during the first trials of exposure to environments whose reward contingencies are unknown, and tends to disappear after training, consistent with a gradual shift from deliberative to routine strategies.

Recent rodent studies have addressed the neuronal dynamics accompanying deliberative and VTE behavior in neural structures that are key for goal-directed spatial navigation, and most prominently the hippocampal-entorhinal system [18•]. This system may support an allocentric spatial map, in which functionally specialized cell types provide metric elements (e.g. place cells that encode specific places within an environment, border cells, head direction cells and

Figure 1



The functions of planning 'at decision time' and 'in the background' discussed in this review; and their putative links to hippocampal theta and SWR sequences.

vectorial representations of spatial goals) and a reference frame (e.g. grid cells) [19–23].

An especially important role for spatial navigation and planning may be played by *sequential assemblies* of hippocampal cells [24,25^{*}]. When a rodent is actively engaged in a task, its hippocampal place cells fire in sequential population patterns called *theta sequences* [26]. These are endogenously generated (but time-compressed) sequences of place cell activation, where individual cells typically code sequentially for the animal's previous location, its current location and a few future locations. The 'theta' refers to the fact that they are locked to specific phases of the hippocampal theta rhythm (yet in bats the coupling can occur outside theta [27]). Theta sequences have been interpreted as mechanisms of memory encoding [28] but – as they code for future locations – also of short-term sequence prediction [29] or predictive coding [24]. In this perspective, theta sequences may be manifestations of a model-based mechanism that continuously estimates the present (e.g. self-localizes) and predicts the future (e.g. paths to a goal location) [25^{*},27,30^{*}]; and these two operations may happen at distinct parts of each theta cycle [22].

Interestingly, when a rodent approaches a difficult decision point, theta sequences can 'sweep forward' from the animal's current location, to serially represent the choice alternatives, such as the branches of a maze [31^{*}]. The causal involvement of forward sweeps in deliberative choice remains to be firmly demonstrated. However, several findings point in that direction. First, forward sweeps seem to occur preferentially during VTE behavior and only during early learning phases [18^{**}]. Second, in a task where rodents had to select multiple goal locations, theta sequences often stretched towards the to-be-selected goal location [32]. Third, during forward sweeps, neurons in the ventral striatum (an area that forms a circuit with the hippocampus and which encodes motivational and reward contingencies) fire in a coordinated manner with the hippocampal theta rhythm [33].

Taken together, these findings are evocative of a model-based control system, in which theta sequences would serially code for possible plans (e.g. spatial trajectories to goal locations), to be evaluated by other brain structures including the ventral striatum [34–36]. The fact that theta sequences still occur after training (yet without 'sweeping forward' at decision points) suggests that a predictive

control scheme remains always in place, but its depth can be adjusted [35,37]. Note that the interpretation of theta sequences in terms of prediction and planning is not necessarily incompatible with their widely assumed role in sequential memory encoding [28]. Rather, predictive processes may support the detection of event boundaries [39] and the segmentation of experiences into behaviorally significant sequences (e.g. of spatial or other events in time [38]), which then form the building blocks of memories.

Other brain areas that form circuits with the hippocampus may participate in the model-based control scheme described here. For example, prefrontal areas may gate deliberative, control planning depth and represent spatial goals [40–44] and grid-cell representations may participate in forward spatial simulations [45]; see Refs. [25[•],30[•],46,47] for computational proposals.

Another rodent study focused on *sequential* choices during spatial navigation [48], by adapting a two-step decision task that was designed to expose (non-spatial) planning in humans [49–51]. It found that VTE behavior occurred at both the successive decision points, but was more sensitive to task contingencies in the earlier of the two, suggesting that this is the phase where the deliberative process is more flexible. A lever pressing version of the same two-step decision task provided strong evidence that the hippocampus supports planning computations, by showing that its inactivation produces planning impairments [52[•]]. Furthermore, it showed that sequential hippocampal dynamics (with neurons encoding relevant task segments rather than locations) may support non-spatial deliberation [53], suggesting the generality of the model-based scheme discussed here.

Human studies of planning during navigation revealed a similar network as the rodent studies, with a more diffuse involvement of the medial temporal lobe (not just the hippocampus) and of prefrontal areas [54–58]. Interestingly, compared to rodent studies, human studies targeted more complex forms of spatial planning involving multiple goals and subgoals — which may explain the involvement of prefrontal mechanisms for cognitive control and temporal abstraction. A study in which participants navigated in a circular arena described network-level interactions between hippocampus and prefrontal cortex during navigation, which flexibly codes for routes and subgoal locations; while other brain regions (*the parahippocampal cortex, perirhinal cortex, and retrosplenial complex*) putatively supporting the (fictive) visualization of future spatial contexts [59^{••}]. Another study revealed a flexible circuit for replanning, in which the lateral prefrontal cortex provides prediction error signals about path changes, frontopolar and superior prefrontal areas reformulate the plan or subgoal, and the hippocampus simulates the new path [60]. Finally, another study revealed a network of caudal prefrontal structures that permits

humans to address rather complex planning tasks, by using a hierarchical strategy that focuses on relevant subgoals (i.e. junction stations in a subway), while ignoring irrelevant spatial-temporal details [61^{••}]. (See also Refs. [62,63] for evidence of sequential goal and subgoal coding in monkey prefrontal areas, when the animals are engaged in a multi-step decision task in a ‘virtual’ navigation.) These findings parallel the well-recognized importance of hierarchical planning, abstraction and sub-goaling in AI [64,65] and underlie the quintessentially goal-based structure of higher cognition [66,67]. They are important to test scalable planning algorithms that we use to address large-scale, real-world problems, beyond the rather restricted experimental situations that are more commonly studied.

Planning in the background: training controllers and optimizing internal models

Planning processes can also operate ‘in the background’, to support learning and *future* action selection. As illustrated by the *Dyna* model, a (model-based) planning process can self-generate or replay ‘fictive’ experience to train a (model-free) controller, which is ultimately responsible for action selection at some later time [68–70]. Similarly, several recent reinforcement learning algorithms replay experience off-line (from a model or a memory pool) to decorrelate inputs, speed up the learning process and bias it to over represent the most important contingencies [71]. Apart for training a controller, one can use planning ‘in the background’ to continuously update an internal model or cognitive map, to generalize or integrate fragmented experiences [25[•],72,73]; and then use this always-updated model for model-based decisions. All these forms of planning ‘in the background’ usually operate when the agent is not engaged in a task and use large batches of experiences to support learning of global task characteristics. However, they can also operate just before a choice, using contextually appropriate sets of experiences to support the current action selection problem. For example, a planning process can self-generate or replay those experiences (fictive or real) that are most relevant for the current choice situation, to train a controller that has to solve it or to calculate on the fly a *local* value function. This is similar to planning ‘at decision time’, but the actual action or policy can be selected by a controller (e.g. model-free) that is distinct from the planner itself.

Sequential neuronal dynamics reminiscent of ‘planning in the background’ and experience replay have been long observed in the rodent hippocampus. Like theta sequences, they resemble (time-compressed) sequential experiences. However, they are generated during *sharp wave ripples* (SWR): very fast neural population patterns which occur during off-line states of the brain, associated with consummatory behavior and non-REM sleep [74[•],75], see Ref. [25[•]]. SWR are typically non-local

and encode trajectories or events that are distinct from the animal's current location (but their statistics during sleep and the awake state are different, see Ref. [25^{*}]). For example, when an animal sleeps or rests, the hippocampus can 'replay' sequences of place cells in the same (or the opposite) order as were encoded when the animal navigated an environment, thus coding for a traversed spatial trajectory, or its reverse [76]. This suggests that the hippocampus stores and 'replays' recent (episodic) memories during off-line periods and sleep, to consolidate (semantic) memories in the neocortex [77]. The finding that (reverse) replay statistics are biased by motivational events such as reward delivery [78] suggests an alternative interpretation in terms of reinforcement learning, where a Dyna-like system biases the replays towards the more (reward-related) salient locations [79^{*},80] — a mechanism that also improves performance in machine learning [81].

Another set of studies link SWR sequences with planning processes just before a choice, when the animal is in the awake state. A foraging study provided intriguing evidence that SWR sequences preferentially depict future paths to goal locations, suggesting an involvement in prospective planning [82^{**}]. Note that unlike theta sequences, these SWR sequences occur when the animal is in the awake state but not directly engaged in navigation. A causal study showed that disrupting SWR sequences (using optogenetics) immediately before difficult choices impairs them (while leaving easier choices intact), which suggests that SWR may support working memory [83] or train/fine-tune a controller immediately before a difficult choice. Another study showed that replays can preferentially depict paths to be avoided (e.g. shock zones) and support inhibitory avoidance behavior [84].

A third set of studies show that the idea that SWR sequences 'replay' recent experience, or the most pressing motivational events, may be too restrictive. One study where rodents were trained in two distinct portions of a maze in different sessions shows preferential replays of *the other* and not *the current* portion; and also valid but unexperienced trajectories, such as shortcuts that the animal had not experienced [85]. Another study shows that a reward revaluation procedure, in which one goal type is made more salient, shifts SWR sequences *away* from that goal location [86]. Other rodent studies found close resemblances between hippocampal sequences *before* and *after* actually experiencing a novel environment, lending support for forms of prospective coding or preplay, whose interpretation is currently debated [24,87–89]. All these studies suggest that SWR sequences may update an internal model by performing internal operations (e.g. re-consolidate old experiences, model reduction) that go beyond the immediate necessities of a controller [25^{*},47,72,90,91]. In this perspective, the 'replays' may be seen as manifestations of sampling dynamics [92–94] or variational iterations [90] of the probabilistic inference implemented by an internal (generative) model.

Recent human studies devised ways to decode fast sequential neural activity from neuroimaging recordings, while humans solve (spatial and non-spatial) planning tasks. These studies revealed fast sequences of task-related state representations, which are putatively analogous to SWR sequences in rodents [95,96^{*}]. Interestingly, one study found that when state representations are experienced in a semantically incorrect order, they are neurally 'replayed' in the semantically correct order [97], lending support to the idea that they stem from an internal model and contribute to updating it [25^{*}].

Summing up, and open issues

Our review exemplified how two kinds of internally generated hippocampal sequences (theta and SWR sequences) may participate in the neural implementation of two planning modes ('at decision time' and 'in the background') that use 'fictive' experience from a model (or more weakly, a memory) for decision-making, memory consolidation, and learning. This model-based control perspective may help clarifying the multiple, sometimes disconnected roles attributed to hippocampal sequences, from memory to prospective functions [18^{**},25^{*},91]. Future study will need to assess whether and how the two modes may interact synergistically; for example, by shifting from short-term (theta) to long-term (SWR) prediction and planning, or encoding (theta) and simulation (SWR) [25^{*},83,98].

Our review has highlighted the prominence of *sequential* brain dynamics to support navigational planning. While we mostly focused on the hippocampus, many other brain areas such as the entorhinal cortex [99], the ventral striatum [100] and the prefrontal cortex [101] show replay-like sequential activity, sometimes in a coordinate manner; whether and how they implement a systems-level planning architecture remains an open question.

Another aspect that is incompletely known is how these sequential assemblies act in a coordinated manner with other (not necessarily sequential) neural codes for goal locations or directions to goals (e.g. vectorial representations) to jointly support spatial planning. The studies we reviewed suggest that the coding of spatial trajectories to goal locations (using sequential dynamics but also static codes, for example, cells for goal-specific trajectories [102,103]) is more prominent – or at least easier for us to study in rodents – compared to the coding of goal locations proper. Yet some salient or structural aspects of the environment, such as reward locations or passages, are preferentially coded, for example, with a higher density of place cells [104].

This points to the fact that efficient model-based computations benefit from optimizing the model (or cognitive map) itself. This can be done, for example, by learning state representations that support efficient planning and prediction [35,105,106]; or that automatically include

(predictive) information about their possible successors, for example, successor representations [107–109]. What the content (and the ‘metric’) of the cognitive maps that we use for spatial navigation in the hippocampus and surrounding structures is remains a heavily discussed topic. Proposals vary, suggesting that they encode Euclidean distance and the path distance to goals [110], topological [111] or graph-like representations [112], patches of 2D and 3D codes [27], deep structural aspects of a problem, such as its temporal community structure [113] that afford efficient hierarchical planning [64,65,114], or global (relational) codes that support long-term prediction, without chaining multiple short-term predictions [114–117] (see also Refs. [19,118] for a discussion of how global codes may explain cognitive maps at large).

Finally, another limitation of current planning studies is that we have limited understanding of the specific (perhaps approximate) planning algorithms that allow us to face the complexities of the real world, where methods like exhaustive tree search would be intractable. Some candidate approximations have been tested (yet mostly at the behavioral level), including pruning unpromising branches, search only until a reliable (cached) habitual pattern can be retrieved or until sufficient discriminative evidence for a plan is gathered [35,119*,120–122]; hybrids of model-based and model-free or habitual controllers [37,123–125]; and hierarchical schemes [65,126–128]. AI research has devised many other planning approximations, such as heuristic planning, various way to transform the problem specification to render planning more tractable [9], and combinations of sample-based planners (e.g. Monte Carlo Tree Search) and deep nets that can solve complex games [119*,129], whose plausibility remains to be tested. More broadly, testing-specific algorithmic implementations of planning in the brain is a task for which a strict collaboration between AI and neuroscience is indispensable [130].

Conflict of interest statement

Nothing declared.

Acknowledgements

This research has received funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2 to GP). We thank Chris Summerfield, Matt Sims and Hugo Spiers for valuable comments and suggestions.

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. Schacter DL, Addis DR, Buckner RL: **Remembering the past to imagine the future: the prospective brain.** *Nat Rev Neurosci* 2007, **8**:657-661.
2. Friston K: **The free-energy principle: a unified brain theory?** *Nat Rev Neurosci* 2010, **11**:127-138.
3. Bar M: **The proactive brain: memory for predictions.** *Philos Trans R Soc Lond B Biol Sci* 2009, **364**:1235-1243.
4. Newell A, Simon HA: *Human Problem Solving.* Prentice-Hall; 1972.
5. Miller GA, Galanter E, Pribram KH: *Plans and the Structure of Behavior.* Holt, Rinehart and Winston; 1960.
6. Anderson JR, Lebiere C: *The Atomic Components of Thought.* Mahwah, NJ: Erlbaum; 1998.
7. Rosenbloom PS, Laird JE, Newell A: *The Soar Papers: Research on Integrated Intelligence.* Cambridge, MA: MIT Press; 1992.
8. Dayan P, Daw ND: **Decision theory, reinforcement learning, and the brain.** *Cogn Affect Behav Neurosci* 2008, **8**:429-453.
9. Geffner H: **Model-free, model-based, and general intelligence.** *ArXiv Prepr ArXiv* 2018. 180602308.
10. Russell SJ, Norvig P: *Artificial intelligence: a modern approach (International Edition).* 2002.
11. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction.* MIT Press; 1998.
12. Tolman EC: **Cognitive maps in rats and men.** *Psychol Rev* 1948, **55**:189-208.
13. Korf RE: **Planning as search: a quantitative approach.** *Artif Intell* 1987, **33**:65-88.
14. Balleine BW, Dickinson A: **Goal-directed instrumental action: contingency and incentive learning and their cortical substrates.** *Neuropharmacology* 1998, **37**:407-419.
15. Daw ND, Dayan P: **The algorithmic anatomy of model-based evaluation.** *Philos Trans R Soc B Biol Sci* 2014, **369**:20130478.
16. Tolman EC: **The determiners of behavior at a choice point.** *Psychol Rev* 1938, **46**:318-336.
17. Santos-Pata D, Verschure PFMJ: **Human vicarious trial and error is predictive of spatial navigation performance.** *Front Behav Neurosci* 2018, **12**:237 <http://dx.doi.org/10.3389/fnbeh.2018.00237>.
18. Redish AD: **Vicarious trial and error.** *Nat Rev Neurosci* 2016, **17**:147-159.
A comprehensive review of the behavioral and neuronal mechanisms supporting vicarious trial and error in rodents.
19. Buzsáki G, Moser EI: **Memory, navigation and theta rhythm in the hippocampal-entorhinal system.** *Nat Neurosci* 2013, **16**:130-138.
20. Sarel A, Finkelstein A, Las L, Ulanovsky N: **Vectorial representation of spatial goals in the hippocampus of bats.** *Science* 2017, **355**:176-180.
21. Grieves RM, Jeffery KJ: **The representation of space in the brain.** *Behav Processes* 2017, **135**:113-131.
22. Sanders H, Rennó-Costa C, Idiart M, Lisman J: **Grid cells and place cells: an integrated view of their navigational and memory function.** *Trends Neurosci* 2015, **38**:763-775.
23. Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI: **Microstructure of a spatial map in the entorhinal cortex.** *Nature* 2005, **436**:801-806.
24. Dragoi G, Buzsáki G: **Temporal encoding of place sequences by hippocampal cell assemblies.** *Neuron* 2006, **50**:145-157.
25. Pezzulo G, Kemere C, van der Meer MAA: **Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition.** *Ann N Y Acad Sci* 2017, **1396**:144-165.
A computational theory of internally generated hippocampal sequences in terms of internal generative models and probabilistic planning-as-inference.
26. Foster DJ, Wilson MA: **Hippocampal theta sequences.** *Hippocampus* 2007, **17**:1093-1099.
27. Finkelstein A, Las L, Ulanovsky N: **3-D maps and compasses in the brain.** *Annu Rev Neurosci* 2016, **39**:171-196.

28. Lisman JE, Idiart MA: **Storage of 7 +/- 2 short-term memories in oscillatory subcycles.** *Science* 1995, **267**:1512.
29. Lisman J, Redish AD: **Prediction, sequences and the hippocampus.** *Philos Trans R Soc B Biol Sci* 2009, **364**:1193-1201.
30. Penny WD, Zeidman P, Burgess N: **Forward and backward inference in spatial cognition.** *PLoS Comput Biol* 2013, **9**: e1003383.
- A systems-level computational model casting spatial cognition and planning in terms of model-based computations.
31. 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.
- A demonstration that hippocampal theta sequences can sequentially encode multiple future trajectories at difficult choice points.
32. Wikenheiser AM, Redish AD: **Hippocampal theta sequences reflect current goals.** *Nat Neurosci* 2015, **18**:289-294.
33. Meer MAA, van der, Redish AD: **Theta phase precession in rat ventral striatum links place and reward information.** *J Neurosci* 2011, **31**:2843-2854.
34. Pezzulo G, van der Meer MAA, Lansink CS, Pennartz CMA: **Internally generated sequences in learning and executing goal-directed behavior.** *Trends Cogn Sci* 2014, **18**:647-657.
35. 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.
36. van der Meer M, Kurth-Nelson Z, Redish AD: **Information processing in decision-making systems.** *Neuroscientist* 2012, **18**:342-359.
37. Pezzulo G, Rigoli F, Chersi F: **The mixed instrumental controller: using value of information to combine habitual choice and mental simulation.** *Front Cogn* 2013, **4**:92.
38. MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum H: **Hippocampal "Time Cells" bridge the gap in memory for discontinuous events.** *Neuron* 2011, **71**:737-749.
39. Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR: **Event perception: a mind-brain perspective.** *Psychol Bull* 2007, **133**:273-293.
40. Gremel CM, Costa RM: **Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions.** *Nat Commun* 2013, **4**:2264.
41. Jones JL, Esber GR, McDannald MA, Gruber AJ, Hernandez A, Mireni A, Schoenbaum G: **Orbitofrontal cortex supports behavior and learning using inferred but not cached values.** *Science* 2012, **338**:953-956.
42. McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G: **Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning.** *J Neurosci* 2011, **31**:2700-2705.
43. Hok V, Chah E, Save E, Poucet B: **Prefrontal cortex focally modulates hippocampal place cell firing patterns.** *J Neurosci* 2013, **33**:3443-3451.
44. Ito HT, Zhang S-J, Witter MP, Moser EI, Moser M-B: **A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation.** *Nature* 2015, **522**:50-55.
45. Bellmund JL, Deuker L, Navarro Schröder T, Doeller CF: **Grid-cell representations in mental simulation.** *eLife* 2016, **5**:e17089.
46. Bicanski A, Burgess N: **A neural-level model of spatial memory and imagery.** *eLife* 2018, **7**:e33752.
47. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G: **Active inference: a process theory.** *Neural Comput* 2017, **29**:1-49.
48. Hasz BM, Redish AD: **Deliberation and procedural automation on a two-step task for rats.** *Front Integr Neurosci* 2018, **12**.
49. 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.
50. Deserno L, Huys QJ, Boehme R, Buchert R, Heinze H-J, Grace AA, Dolan RJ, Heinz A, Schlagenhaut F: **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.
51. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND: **Model-based choices involve prospective neural activity.** *Nat Neurosci* 2015, **18**:767-772.
52. Miller KJ, Botvinick MM, Brody CD: **Dorsal hippocampus contributes to model-based planning.** *Nat Neurosci* 2017, **20**:1269-1276.
- A demonstration that the hippocampus is required for model-based planning, outside the spatial navigation domain.
53. Miller K, Daw N, Botvinick M, Brody CD: **Hippocampal sequences and model-based planning in the rat.** *Computational and Systems Neuroscience (CoSyNe) Conference; 28 February-03 March 2019 in Lisbon, Portugal*: 2019.
54. Javadi A-H, Emo B, Howard LR, Zisch FE, Yu Y, Knight R, Pinelo Silva J, Spiers HJ: **Hippocampal and prefrontal processing of network topology to simulate the future.** *Nat Commun* 2017, **8**:14652.
55. Simon DA, Daw ND: **Neural correlates of forward planning in a spatial decision task in humans.** *J Neurosci* 2011, **31**:5526-5539.
56. 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.
57. Spiers HJ, Maguire EA: **A navigational guidance system in the human brain.** *Hippocampus* 2007, **17**:618-626.
58. Deuker L, Bellmund JL, Navarro Schröder T, Doeller CF: **An event map of memory space in the hippocampus.** *eLife* 2016, **5**: e16534.
59. Brown TI, Carr VA, LaRocque KF, Favila SE, Gordon AM, Bowles B, Bailenson JN, Wagner AD: **Prospective representation of navigational goals in the human hippocampus.** *Science* 2016, **352**:1323-1326.
- A detailed study of the neuronal architecture supporting spatial planning and prospective goal and subgoal representations in the human brain.
60. Spiers HJ, Gilbert SJ: **Solving the detour problem in navigation: a model of prefrontal and hippocampal interactions.** *Front Hum Neurosci* 2015, **9**.
61. Balaguer J, Spiers H, Hassabis D, Summerfield C: **Neural mechanisms of hierarchical planning in a virtual subway network.** *Neuron* 2016, **90**:893-903.
- A nicely designed human study that provides behavioral evidence for hierarchical planning and illustrates its neuronal underpinnings.
62. Mushiaki H, Saito N, Sakamoto K, Itoyama Y, Tanji J: **Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans.** *Neuron* 2006, **50**:631-641.
63. Saito N, Mushiaki H, Sakamoto K, Itoyama Y, Tanji J: **Representation of immediate and final behavioral goals in the monkey prefrontal cortex during an instructed delay period.** *Cereb Cortex* 2005, **15**:1535-1546.
64. Botvinick M, Weinstein A: **Model-based hierarchical reinforcement learning and human action control.** *Philos Trans R Soc Lond B Biol Sci* 2014, **369**:20130480.
65. Donnarumma F, Maisto D, Pezzulo G: **Problem solving as probabilistic inference with subgoal: explaining human successes and pitfalls in the tower of Hanoi.** *PLoS Comput Biol* 2016, **12**:e1004864.
66. Pezzulo G, Castelfranchi C: **Thinking as the control of imagination: a conceptual framework for goal-directed systems.** *Psychol Res* 2009, **73**:559-577.
67. O'Reilly RC, Hazy TE, Mollick J, Mackie P, Herd S: **Goal-driven cognition in the brain: a computational framework.** *ArXiv14047591 Q-Bio* 2014.

68. Sutton RS: **Integrated architectures for learning, planning, and reacting based on approximating dynamic programming.** In *Proceedings of the Seventh International Conference on Machine Learning*. 1990:216-224. Morgan Kaufmann.
69. Moore AW, Atkeson CG: **Prioritized sweeping: reinforcement learning with less data and less time.** *Mach Learn* 1993, **13**:103-130.
70. Gershman SJ, Markman AB, Otto AR: **Retrospective revaluation in sequential decision making: a tale of two systems.** *J Exp Psychol Gen* 2014, **143**:182.
71. Kumaran D, Hassabis D, McClelland JL: **What learning systems do intelligent agents need? complementary learning systems theory updated.** *Trends Cogn Sci* 2016, **20**:512-534.
72. Penagos H, Varela C, Wilson MA: **Oscillations, neural computations and learning during wake and sleep.** *Curr Opin Neurobiol* 2017, **44**:193-201.
73. Redish AD, Touretzky DS: **The role of the hippocampus in solving the Morris water maze.** *Neural Comput* 1998, **10**:73-111.
74. Buzsáki G: **Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning.** *Hippocampus* 2015, **25**:1073-1188.
An extremely detailed and comprehensive review of hippocampal sharp-wave ripples.
75. Diba K, Buzsáki G: **Forward and reverse hippocampal place-cell sequences during ripples.** *Nat Neurosci* 2007, **10**:1241-1242.
76. Foster DJ, Wilson MA: **Reverse replay of behavioural sequences in hippocampal place cells during the awake state.** *Nature* 2006, **440**:680-683.
77. McClelland JL, McNaughton BL, O'Reilly RC: **Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory.** *Psychol Rev* 1995, **102**:419-457.
78. Ambrose RE, Pfeiffer BE, Foster DJ: **Reverse replay of hippocampal place cells is uniquely modulated by changing reward.** *Neuron* 2016, **91**:1124-1136.
79. Mattar MG, Daw ND: **Prioritized memory access explains planning and hippocampal replay.** *Nat Neurosci* 2018, **21**:1609.
A computational theory of hippocampal replays in terms of reinforcement learning and prioritized replay, within a Dyna-like architecture.
80. Cazé R, Khamassi M, Aubin L, Girard B: **Hippocampal replays under the scrutiny of reinforcement learning models.** *J Neurophysiol* 2018, **120**:2877-2896.
81. Schaul T, Quan J, Antonoglou I, Silver D: *Prioritized Experience Replay*. 2015.
82. Pfeiffer BE, Foster DJ: **Hippocampal place-cell sequences depict future paths to remembered goals.** *Nature* 2013, **497**:74-79.
A study using a well-designed paradigm to link explicitly hippocampal SWR sequences and model-based planning in rodents.
83. Jadhav SP, Kemere C, German PW, Frank LM: **Awake hippocampal sharp-wave ripples support spatial memory.** *Science* 2012, **336**:1454-1458.
84. Wu C-T, Haggerty D, Kemere C, Ji D: **Hippocampal awake replay in fear memory retrieval.** *Nat Neurosci* 2017, **20**:571-580.
85. Gupta AS, van der Meer MAA, Touretzky DS, Redish AD: **Hippocampal replay is not a simple function of experience.** *Neuron* 2010, **65**:695-705.
86. Carey AA, Tanaka Y, van der Meer M: **Reward revaluation biases hippocampal sequence content away from the preferred outcome.** *bioRxiv* 2018:397950 <http://dx.doi.org/10.1101/397950>.
87. Dragoi G, Tonegawa S: **Preplay of future place cell sequences by hippocampal cellular assemblies.** *Nature* 2011, **469**:397-401.
88. Liu K, Sibille J, Dragoi G: **Generative predictive codes by multiplexed hippocampal neuronal tuples.** *Neuron* 2018, **99**:1329-1341.e6.
89. Ólafsdóttir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ: **Hippocampal place cells construct reward related sequences through unexplored space.** *eLife* 2015, **4**:e06063.
90. Friston KJ, Lin M, Frith CD, Pezzulo G, Hobson JA, Ondobaka S: **Active inference, curiosity and insight.** *Neural Comput* 2017, **29**:2633-2683 http://dx.doi.org/10.1162/neco_a_0099.
91. Ólafsdóttir HF, Bush D, Barry C: **The role of hippocampal replay in memory and planning.** *Curr Biol* 2018, **28**:R37-R50.
92. Friedrich J, Lengyel M: **Goal-directed decision making with spiking neurons.** *J Neurosci* 2016, **36**:1529-1546.
93. Buesing L, Bill J, Nessler B, Maass W: **Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons.** *PLoS Comput Biol* 2011, **7**:e1002211.
94. Rueckert E, Kappel D, Tanneberg D, Pecevski D, Peters J: **Recurrent spiking networks solve planning Tasks.** *Sci Rep* 2016, **6**:21142.
95. Schuck NW, Niv Y: **Sequential replay of non-spatial task states in the human hippocampus.** *bioRxiv* 2018:315978 <http://dx.doi.org/10.1101/315978>.
96. Kurth-Nelson Z, Economides M, Dolan RJ, Dayan P: **Fast sequences of non-spatial state representations in Humans.** *Neuron* 2016, **91**:194-204.
Fast, replay-like activity decoded from the human brain during non-spatial planning.
97. Liu Y, Kurth-Nelson Z, Behrens T: *Non-spatial Neural Replay in Building and Updating World Models in Humans*. Cosyne; 2018.
98. Drieu C, Todorova R, Zugaro M: **Nested sequences of hippocampal assemblies during behavior support subsequent sleep replay.** *Science* 2018, **362**:675-679.
99. O'Neill J, Boccaro CN, Stella F, Schoenenberger P, Csicsvari J: **Superficial layers of the medial entorhinal cortex replay independently of the hippocampus.** *Science* 2017, **355**:184-188.
100. Lansink CS, Goltstein PM, Lankelma JV, McNaughton BL, Pennartz CM: **Hippocampus leads ventral striatum in replay of place-reward information.** *PLoS Biol* 2009, **7**:e1000173.
101. Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP: **Replay of rule-learning related neural patterns in the prefrontal cortex during sleep.** *Nat Neurosci* 2009, **12**:919-926.
102. Grieves RM, Wood ER, Dudchenko PA: **Place cells on a maze encode routes rather than destinations.** *eLife* 2016, **5**:e15986.
103. Frank LM, Brown EN, Wilson M: **Trajectory encoding in the hippocampus and entorhinal cortex.** *Neuron* 2000, **27**:169-178.
104. Spiers HJ, Hayman RMA, Jovalekic A, Marozzi E, Jeffery KJ: **Place field repetition and purely local remapping in a multicompartiment environment.** *Cereb Cortex N Y N* 1991 2015, **25**:10-25.
105. Wayne G, Hung C-C, Amos D, Mirza M, Ahuja A, Grabska-Barwinska A, Rae J, Mirowski P, Leibo JZ, Santoro A et al.: *Unsupervised Predictive Memory in a Goal-Directed Agent*. 2018.
106. Tishby N, Pereira FC, Bialek W: **The information bottleneck method.** *ArXiv Prepr Physics* 2000. 0004057.
107. Stachenfeld KL, Botvinick M, Gershman SJ: **Design principles of the hippocampal cognitive map.** *Advances in Neural Information Processing Systems*. 2014:2528-2536.
108. Stachenfeld KL, Botvinick MM, Gershman SJ: **The hippocampus as a predictive map.** *Nat Neurosci* 2017, **20**:1643-1653.
109. Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw ND, Gershman SJ: **The successor representation in human reinforcement learning.** *Nat Hum Behav* 2017, **1**:680.
110. Howard LR, Javadi AH, Yu Y, Mill RD, Morrison LC, Knight R, Loftus MM, Staskute L, Spiers HJ: **The hippocampus and**

- entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Curr Biol CB* 2014, **24**:1331-1340.**
111. Dabaghian Y, Mémoli F, Frank L, Carlsson G: **A topological paradigm for hippocampal spatial map formation using persistent homology.** *PLoS Comput Biol* 2012, **8**:e1002581.
 112. Chrastil ER, Warren WH: **From cognitive maps to cognitive graphs.** *PLoS One* 2014, **9**:e112544.
 113. Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM: **Statistical learning of temporal community structure in the hippocampus.** *Hippocampus* 2016, **26**:3-8.
 114. Maisto D, Donnarumma F, Pezzulo G: **Nonparametric problem-space clustering: learning efficient codes for cognitive control tasks.** *Entropy* 2016, **18**:61.
 115. Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z: **What is a cognitive map? Organizing knowledge for flexible behavior.** *Neuron* 2018, **100**:490-509.
 116. Baram AB, Muller TH, Whittington JCR, Behrens TEJ: **Intuitive planning: global navigation through cognitive maps based on grid-like codes.** *bioRxiv* 2018:421461 <http://dx.doi.org/10.1101/421461>.
 117. Collin SH, Milivojevic B, Doeller CF: **Hippocampal hierarchical networks for space, time, and memory.** *Curr Opin Behav Sci* 2017, **17**:71-76.
 118. Bellmund JLS, Gärdenfors P, Moser EI, Doeller CF: **Navigating cognition: Spatial codes for human thinking.** *Science* 2018, **362**:eaat6766.
 119. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M *et al.*: **Mastering the game of Go with deep neural networks and tree search.** *Nature* 2016, **529**:484-489.
- A deep reinforcement learning algorithm, which achieves state-of-the-art performance in games using a wise combination of planning and learning processes.
120. Huys QJM, Lally N, Faulkner P, Eshel N, Seifritz E, Gershman SJ, Dayan P, Roiser JP: **Interplay of approximate planning strategies.** *Proc Natl Acad Sci U S A* 2015, **112**:3098-3103.
 121. 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.
 122. 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 U S A* 2016, **113**:12868-12873.
 123. Economides M, Kurth-Nelson Z, Lübbert A, Guitart-Masip M, Dolan RJ: **Model-based reasoning in humans becomes automatic with training.** *PLoS Comput Biol* 2015, **11**:e1004463.
 124. Cushman F, Morris A: **Habitual control of goal selection in humans.** *Proc Natl Acad Sci U S A* 2015, **112**:13817-13822.
 125. Keramati M, Dezfouli A, Piray P: **Speed/accuracy trade-off between the habitual and the goal-directed processes.** *PLoS Comput Biol* 2011, **7**:e1002055.
 126. Solway A, Diuk C, Córdoba N, Yee D, Barto AG, Niv Y, Botvinick MM: **Optimal behavioral hierarchy.** *PLoS Comput Biol* 2014, **10**:e1003779.
 127. Tomov M, Yagati S, Kumar A, Yang W, Gershman S: **Discovery of hierarchical representations for efficient planning.** *bioRxiv* 2018:499418 <http://dx.doi.org/10.1101/499418>.
 128. McNamee D, Wolpert DM, Lengyel M: **Efficient state-space modularization for planning: theory, behavioral and neural signatures.** In *Advances in Neural Information Processing Systems 29*. Edited by Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R. Curran Associates, Inc.; 2016:4511-4519.
 129. Anthony T, Tian Z, Barber D: **Thinking fast and slow with deep learning and tree search.** *ArXiv170508439 Cs* 2017.
 130. Hassabis D, Kumaran D, Summerfield C, Botvinick M: **Neuroscience-inspired artificial intelligence.** *Neuron* 2017, **95**:245-258.