Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Predicting the past, remembering the future

Samuel J Gershman

Rational analyses of memory suggest that retrievability of past experience depends on its usefulness for predicting the future: memory is adapted to the temporal structure of the environment. Recent research has enriched this view by applying it to semantic memory and reinforcement learning. This paper describes how multiple forms of memory can be linked via common predictive principles, possibly subserved by a shared neural substrate in the hippocampus. Predictive principles offer an explanation for a wide range of behavioral and neural phenomena, including semantic fluency, temporal contiguity effects in episodic memory, and the topological properties of hippocampal place cells.

Address
Department of Psychology and Center for Brain Science, Harvard University, United States

Corresponding author: Gershman, Samuel J (gershman@fas.harvard.edu)

Current Opinion in Behavioral Sciences 2017, **17**:7–13

This review comes from a themed issue on **Memory in time and space**

Edited by **Lila Davachi** and **Neil Burgess**

http://dx.doi.org/10.1016/j.cobeha.2017.05.025

Introduction

Why remember the past? George Santayana famously remarked that 'Those who cannot remember the past are condemned to repeat it' [1], and William Faulkner expressed a similar sentiment in *Requiem for a Nun*: 'The past is never dead. It isn't even the past.' Patterns of recurrence have fascinated historians from Polybius to Arnold Toynbee [2], and some have tried to harvest lessons from these patterns to guide political decision making [3]. More mundanely, the daily lives of most organisms are structured by cycles of sleeping, eating and other routine activities. At shorter timescales, many events can be anticipated using sensory information from the recent past, giving organisms with memory an adaptive advantage. Thus, memory of the past, spanning vastly different timescales, is relevant for predicting and controlling the future [4•,5–7].

Most research on human memory has focused on understanding, at a descriptive level, how information from the past is stored and retrieved, without contemplating the usefulness of this information for future action. However, researchers have increasingly come to appreciate the fact that memory is organized around predictive design principles [8], evident in multiple forms of memory (semantic, episodic, short-term and procedural). Computational models have formalized these principles mathematically, drawing upon ideas from library science, search engine algorithms, probability theory, and reinforcement learning. Despite this diversity of approaches, we will see that several of them can be unified in terms of a single predictive representation, repurposed (with slight modifications) for the needs of different memory systems. Neural correlates of this predictive representation have been observed in the hippocampus, suggesting a functional explanation for the region's involvement in both retrospective and prospective cognition.

The rational analysis of memory

John Anderson's 'rational analysis of memory' was the first theoretical attempt to explain the structure of memory in terms of beliefs about the future [9,10]. Anderson conceptualized the problem facing memory as one of determining *need probability*: the likelihood that a particular piece of information will be needed in the future. On the assumption that past need predicts future need, memory can used to produce a forecast. Importantly, because need probability may change over time, this forecast should adapt to the statistics of past experience.

It is worth highlighting here the extent to which this view represents a significant departure from prevailing ideas about episodic memory. The standard view (see [11•] for a summary) holds that there is a fundamental tension between memory for specific instances (episodic memory) and memory for statistical regularities (semantic memory). In contrast, Anderson's rational analysis posited that episodic memory is structured by statistical regularities. To formalize this idea, Anderson adapted a model of library borrowing, likening memories to books in a library system. The usage of each book is subject to fluctuations, and the task facing the library system is to track these usage statistics in order to anticipate the probability that a book will be needed in the near future. Anderson showed that this model could capture many basic properties of memory, such as spacing, recency, fan and word frequency effects.

Several authors have explored different assumptions about the statistical regularities governing episodic memory. For example, Gershman and colleagues [12] developed a Bayesian nonparametric model in which the environmental dynamics can switch between an unknown number of 'modes.' This model makes the prediction (confirmed experimentally) that abrupt changes will result in the inference that a new mode has been encountered, effectively creating an event boundary that reduces interference between items on either side of the boundary. The importance of event boundaries in memory formation and retrieval has been highlighted by a number of other recent studies [13–18].

Rational analysis has also been applied to semantic memory. In this case, the need probability is dictated by the long-run marginal probability of items, rather than the temporally specific need probability used in the analysis of episodic memory. Griffiths and colleagues [19] showed that the semantic need probability is precisely what is computed by Google's PageRank search algorithm [20]. They conceptualized semantic memory as a directed network (a kind of mental World Wide Web) over which semantic processing flows according to a random walk. If we let $\bar{L}_{ij} = L_{ij}/\sum_k L_{ik}$ denote the normalized link matrix (where $L_{ij} = 1$ if there is a link from item i to item j, and 0 otherwise), then PageRank computes the vector of item ranks $\mathbf{r}$ according to the recursive definition $\mathrm{r} = \bar{\mathrm{L}}\mathrm{r}$.

We can understand the rank of item i as proportional to its marginal probability $P(i)$ after running the random walk on the semantic network for a long time. Intuitively, this means that the proportion of time a person spends thinking about item i scales with its rank. It can also be interpreted as a probabilistic model of environmental structure: if the need for an item over time can be described by a random walk, then its rank reflects its long-run need probability. Griffiths and colleagues asked participants to generate words beginning with a particular letter, and showed that the number of times a word was retrieved (a measure of semantic fluency) was well-predicted by its rank. This suggests that human semantic memory is structured to make predictions of future need probability over long timescales.

A predictive substrate

Another way to derive the PageRank algorithm is to construct a predictive representation for each item i that encodes the expected discounted number of times item j will be needed on a random walk initiated at i:

$$M_{ij} = E\left[\sum_{t=1}^{\infty} \gamma^t I(s_t = j) | s_0 = i\right], \tag{1}$$

where s_t denotes the item needed at time t, $I(\cdot) = 1$ if its argument is true (0 otherwise), and γ is a discount factor controlling the effective time horizon over which item counts are accumulated. The expectation is taken over randomness in state transitions and rewards. One can interpret $1 - \gamma$ as the constant probability that the random walk will terminate at any given time. An item's rank can be computed from the predictive representation simply by summing the rows of $\mathbf{M}: r_j \propto \sum_i M_{ij}$. In other words, the long-run probability of visiting an item is obtained by summing its cumulative expected need across the collection of random walks initiated at different items.

This way of deriving PageRank is intriguing because the matrix $\mathbf{M}$ has another life in the reinforcement learning literature, where it is known as the *successor representation* (SR) [21]. The central problem in reinforcement learning is the computation of *value*, the discounted cumulative reward expected in the future upon entering a state:

$$V_i = E\left[\sum_{t=1}^{\infty} \gamma^t R(s_t) | s_0 = i\right], \tag{2}$$

where $R(s)$ is the immediate reward expected upon entering state s. The SR renders value computation a linear operation, due to the identity $V_i = \sum_j M_{ij} R(j)$. It therefore offers a significant computational advantage over traditional model-based planning algorithms that typically scale superlinearly with the number of states [22].

Further insight can be gleaned by considering the long-run average reward. It can be shown that the long-run average reward is a reward-weighted version of the rank (and hence need probability) returned by the PageRank algorithm, revealing a deep and surprising formal connection between reinforcement learning, information retrieval, and the rational analysis of memory. It suggests a way in which reward associations can reweight need probabilities, in accordance with well-known motivational influences on memory [23–25].

Learning the successor representation with temporal context

The SR is defined as an expectation over infinitely long trajectories, raising the question of how it can be tractably computed. One possibility, borrowing ideas from reinforcement learning, is to directly update an estimate the SR ($\hat{\mathrm{M}}$) from state transitions. Specifically, the SR can be updated incrementally using a form of temporal difference learning [26]:

$$\Delta\hat{M}(i,j) = \alpha[I(s_{t+1} = j) + \gamma\hat{M}(s_{t+1},j) - \hat{M}(s_t,j)]e_t(i), \tag{3}$$

where α is a learning rate and $e_t(i)$ is an exponentially decaying memory trace of recent states that specifies the ‘eligibility’ of state i for updating. The term in brackets represents a ‘prediction error’ analogous to the reward prediction error thought to be conveyed by phasic dopamine signaling [27], but in this case the prediction error is defined on state expectations rather than on reward expectations.

The temporal difference update described above is nearly identical to an influential theory of episodic memory, the temporal context model (TCM; [28,29]). TCM was originally developed to explain the temporal structure of memory retrieval in free recall experiments, but has more recently been applied to many other phenomena, ranging from false memory [30,31] to spacing and repetition effects [32]. According to TCM, items are bound in memory to a slowly drifting representation of temporal context (a recency-weighted average of previous items), and at test the temporal context acts as a retrieval cue, preferentially drawing items based on their strength of association. It can be shown (see [26] for details) that the temporal context representation corresponds to the eligibility trace e_t, and the matrix of item-context associations corresponds to $\hat{M}$.

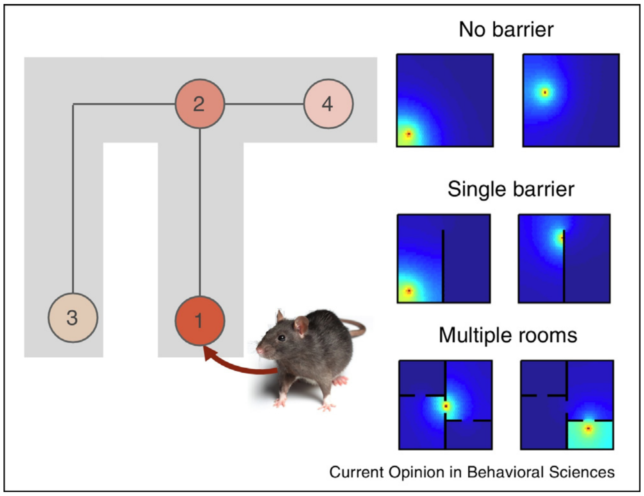
The connection between TCM and the SR suggests another way in which episodic memory can be cast in terms of predictive mechanisms. However, the temporal difference learning update is equivalent to the originally proposed Hebbian update rule [28] only when items are presented once per list. In this case, the second term in the temporal difference update is always zero, because no predictions for novel items have been formed yet. When items are repeated, the second term comes into play, altering the theory’s predictions. An elegant study [33••] drew out these implications, testing the prediction that the temporal difference update will produce a *context repetition effect*: repeating the temporal context of a particular item will strengthen memory for that item, even if the item itself was not repeated. This prediction follows from the fact that the second term in the temporal difference update will be positive whenever an item’s temporal context recurs. The study found support for this prediction across several experiments, as well as a boundary condition, whereby the effect only occurs when the item is strongly predicted by the temporal context. This again is entirely consistent with the theory, since the second term in the update will be smaller when the predictive relationship between context and item is weak. The context repetition effect was also recently demonstrated in rats using an object recognition task [34].

The role of the hippocampus

The multi-faceted role of predictive representations in memory suggests that they may have a dedicated neural substrate. Several lines of evidence suggest that the hippocampus, although traditionally viewed as a repository for episodic memory and spatial representation, may in fact be organized around predictive principles [35,36]. First, place cells in the hippocampus sweep ahead of an animal’s current position when it is at a choice point [37]; these forward sweeps may arise from phase precession, the progressive shift in spike timing relative to the ongoing theta oscillation as an animal moves through a place field [38,39,4•]. Second, when an animal repeatedly runs on a particular trajectory, place fields tend to expand opposite the direction of travel [40], consistent with the idea that earlier place cells learn to predict upcoming locations. Third, functional MRI experiments in humans have indicated that the hippocampus is sensitive to the predictability of upcoming stimuli [41–43,44•]. Fourth, the hippocampus is activated when humans engage in episodic future-thinking [45], and damage to the hippocampus appears to severely impair this ability [46] (though this finding is controversial; see [47]).

Recent theoretical work has formalized the predictive function of the hippocampus in terms of the SR [48]. According to this view, place cells do not actually encode spatial location; they encode expectations about future locations (though see [49] for evidence that some place cells have no predictive properties during immobility). In an open field, these future locations will tend to be in the vicinity of the animal’s current location, yielding classical radially symmetric place fields. However, predictive fields will become distorted when obstacles are introduced or the environmental topology is altered (Figure 1) — just as observed experimentally [50–52]. Likewise, replay of place cell sequences appear to respect the

Figure 1



Current Opinion in Behavioral Sciences

The successor representation defined over space. (Left) Neurons tuned to different locations encode prediction about future locations. Shading corresponds to expected number of times a neuron’s preferred location will be visited along a trajectory initiated at the animal’s current location. (Right) Simulated predictive place fields for different environments.

environmental topology [53]. The topological sensitivity of the SR is a natural consequence of the fact that an animal's future trajectory is constrained by barriers to movement.

Even when there are not physical barriers to movement, the SR may still distort along certain paths, for example, when rewards occur in reliable locations. This follows from the fact that an animal's policy tends to be reward-seeking. Consistent with this hypothesis, place fields tend to cluster around rewarded locations [54]. Moreover, a recent experiment found that forward sweeps preferentially visit rewarded locations [55••].

Early during learning, the SR may resemble a relatively 'pure' representation of space, since the predictive relationships between states have not yet been learned. The fact that some place cells show no predictive properties during early exposure to an environment [49] aligns well with the classic finding that contextual fear conditioning is more effective if the animal is first pre-exposed to the environment (an effect that is hippocampus-dependent [56]). According to the SR theory, predictive relationships between states allow the shock association to propagate to other states (i.e. locations within the environment), leading to a more robust fear memory.

It is important to note that the SR theory does not entirely dispense with some notion of purely spatial representation, as posited by the classical place cell literature. In fact, the SR, when defined over space, is predicated on such a representation, since this is the only way to index spatially distinct states. Likewise, place cells could also index local perceptual features [57], allowing the SR to be defined over this feature space.

Another point of contact between the classical place cell literature and the SR theory concerns the mechanisms of SR updating. The previous section described a temporal difference learning algorithm for updating the SR based on experienced trajectories. However, it is well-known that animals can rapidly change their behavior in response to structural changes in the environment without directly experiencing those changes, such as the detour behavior observed by Tolman [58]. These rapid behavioral changes seem to require an internal model or 'cognitive map' that can be updated in the absence of direct experience. One way to accomplish this within the SR theory is to posit a simulation mechanism that can use an internal model to update the SR [59,60], possibly implemented by forward sweeps of place cells [37].

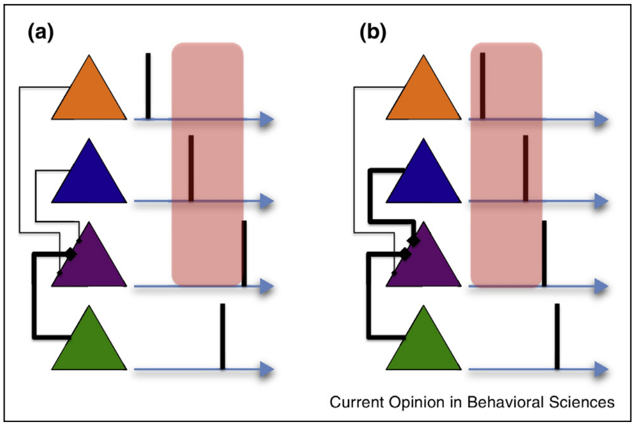
One of the strengths of the SR theory is that it provides a framework for understanding the origin of non-spatial, relational representations in the hippocampus. For example, the hippocampus is involved in the computation of transitive inferences [61], and damage to the hippocampus impairs transitive inference [62]. Recently, Garvert and colleagues [63•] used functional MRI to directly examine the SR account of relational representation. Subjects in their study were exposed to sequences of items generated by a random walk on a graph. Representational similarity in the hippocampus (as measured by adaptation of the neural response across random pairs of items) scaled inversely with distance on the graph, and quantitative analyses revealed that this similarity structure was best described by the SR compared to other graph-theoretic measures.

Predictive spiking neurons

How does the SR arise in biologically realistic neural circuits? Neurons must spike within a few tens of milliseconds of one another for their synapses to be strengthened, much shorter than the temporal horizon of predictive codes like the SR (on the order of seconds). Brea and colleagues [64••] have developed a framework for resolving this puzzle, using a variant of traditional spike timing-dependent plasticity [65].

The basic idea is illustrated in Figure 2. When a synapse between two neurons is potentiated, the presynaptic neuron can cause the postsynaptic neuron to fire earlier. This in turn allows the postsynaptic neuron to enter into plasticity with other neurons that are spiking even earlier. Thus, the membrane potential of the postsynaptic neuron comes to reflect the anticipatory drive from presynaptic

Figure 2



Prospective coding by spiking neurons. **(a)** A network of spiking neurons (colored triangles), with spike times indicated by vertical lines. The green neuron drives spiking of the purple neuron, opening a potentiation window (red box) during which other spiking neurons can strengthen their synaptic connections with the purple neuron. **(b)** As a consequence of plasticity, the purple neuron can now be driven by the blue neuron, shifting the potentiation window earlier and allowing the orange neuron to strengthen its connection with the purple neuron. The activity of the green neuron thus becomes predicted by neurons spiking earlier in time.
Adapted from [64••].

inputs; in effect, the postsynaptic neuron is anticipating its own future spiking. This is the same kind of bootstrapping that underlies the temporal difference learning algorithm described above.

Brea and colleagues [64••] showed that a form of spike timing-dependent plasticity can be used to implement predictive spiking. In particular, under some fairly general assumptions their plasticity rule enables the dendritic potential to directly encode the expected discounted future spike rate at the soma — that is, one column of the SR matrix $\mathbf{M}$. Predictive spiking offers an explanation for widespread anticipatory responses in the brain [66,43,67].

Conclusions

Starting from the principle that memory is designed to serve a predictive function, a rich web of theoretical insights can be derived. Semantic and episodic memory can be linked via a common predictive representation initially studied in the reinforcement learning literature, allowing us to contemplate the computational properties that these different forms of memory may have in common. These commonalities may derive from a shared neural substrate, as suggested by the widespread involvement of the hippocampus across domains [68].

Despite the appeal of a unifying neural substrate, this viewpoint is in tension with taxonomies of memory that stipulate dissociable systems in the brain [69]. For example, transformational consolidation theories hold that episodic memories stored initially in the hippocampus are gradually transformed into semantic memories stored in neocortex [70,11•]. This viewpoint is supported by evidence that hippocampal amnesics show spared long-term semantic memory [71]. However, more recent experiments have found semantic impairments in amnesia (see [72] for a review), and amnesics exhibit semantically impoverished episodic future-thinking [73]. Beyond semantic memory, there is now considerable evidence that the hippocampus (and episodic memory more generally) plays an important role in reinforcement learning as well [7]. Collectively, these observations lend credence to the idea that multiple forms of memory may draw upon a common predictive representation in the hippocampus.

Another tension derives from dissociations between different forms of memory that support behavioral control in reinforcement learning tasks. For example, stimulus-response strategies depend on subregions of the striatum but not on the hippocampus, whereas allocentric ‘place’ strategies depend on the hippocampus but not the striatum [74]. These dissociations indicate that a predictive representation in the hippocampus can only be one part of a larger multi-system architecture, whose components interact both competitively and cooperatively [7].

Conflict of interest statement

Nothing declared.

Acknowledgements

I am grateful to Matthew Botvinick and Nathaniel Daw for helpful discussions, and to Dan Schacter for comments on a previous draft. This work was made possible through grant support from the National Institutes of Health.

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. Santayana G: *The Life of Reason* . Constable and Company Limited; 1905.
2. Trompf GW: *The Idea of Historical Recurrence in Western Thought: From Antiquity to the Reformation*. University of California Press; 1979.
3. Neustadt RE, Ernest R: *Thinking in Time: The Uses of History for Decision-Makers*. New York: Free Press; 1986.
4. • Shankar KH, Singh I, Howard MW: **Neural mechanism to simulate a scale-invariant future**. *Neural Comput* 2016, **28**:2594-2627.

A computational model of how hippocampal theta oscillations can be used to predict the future, by translating a representation of stimulus history forward in time.

5. Heeger DJ: **Theory of cortical function**. *Proc Natl Acad Sci U S A* 2017, **114**:1773-1782.
6. Palmer SE, Marre O, Berry MJ, Bialek W: **Predictive information in a sensory population**. *Proc Natl Acad Sci U S A* 2015, **112**:6908-6913.
7. Gershman SJ, Daw ND: **Reinforcement learning and episodic memory in humans and animals: an integrative framework**. *Annu Rev Psychol* 2017, **68**:101-128.
8. Schacter DL: **Adaptive constructive processes and the future of memory**. *Am Psychol* 2012, **67**:603-613.
9. Anderson JR, Milson R: **Human memory: an adaptive perspective**. *Psychol Rev* 1989, **96**:703-719.
10. Anderson JR, Schooler LJ: **Reflections of the environment in memory**. *Psychol Sci* 1991, **2**:396-408.
11. • 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.

A theoretical review linking the neurocognitive mechanisms of episodic and semantic memory to machine learning algorithms.

12. Gershman SJ, Radulescu A, Norman KA, Niv Y: **Statistical computations underlying the dynamics of memory updating**. *PLOS Comput Biol* 2014, **10**:e1003939.
13. Swallow KM, Zacks JM, Abrams RA: **Event boundaries in perception affect memory encoding and updating**. *J Exp Psychol: Gen* 2009, **138**:236-257.
14. Ezzyat Y, Davachi L: **What constitutes an episode in episodic memory?** *Psychol Sci* 2011, **22**:243-252.
15. Ezzyat Y, Davachi L: **Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity**. *Neuron* 2014, **81**:1179-1189.
16. DuBrow S, Davachi L: **The influence of context boundaries on memory for the sequential order of events**. *J Exp Psychol: Gen* 2013, **142**:1277-1286.
17. Horner AJ, Bisby JA, Wang A, Bogus K, Burgess N: **The role of spatial boundaries in shaping long-term event representations**. *Cognition* 2016, **154**:151-164.

18. Pettijohn KA, Thompson AN, Tamplin AK, Krawietz SA, Radvansky GA: **Event boundaries and memory improvement**. *Cognition* 2016, **148**:136-144.

19. Griffiths TL, Steyvers M, Firl A: **Google and the mind predicting fluency with PageRank**. *Psychol Sci* 2007, **18**:1069-1076.

20. Brin S, Page L: **The anatomy of a large-scale hypertextual web search engine**. *Comput Netw ISDN Syst* 1998, **30**:107-117.

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

22. Daw ND, Dayan P: **The algorithmic anatomy of model-based evaluation**. *Philos Trans R Soc B* 2014, **369**:20130478.

23. Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze H-J, Düzel E: **Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation**. *Neuron* 2005, **45**:459-467.

24. Alison Adcock R, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JDE: **Reward-motivated learning: mesolimbic activation precedes memory formation**. *Neuron* 2006, **50**:507-517.

25. Miendlarzewska EA, Bavelier D, Schwartz S: **Influence of reward motivation on human declarative memory**. *Neurosci Biobehav Rev* 2016, **61**:156-176.

26. Gershman SJ, Moore CD, Todd MT, Norman KA, Sederberg PB: **The successor representation and temporal context**. *Neural Comput* 2012, **24**:1553-1568.

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

28. Howard MW, Kahana MJ: **A distributed representation of temporal context**. *J Math Psychol* 2002, **46**:269-299.

29. Sederberg PB, Howard MW, Kahana MJ: **A context-based theory of recency and contiguity in free recall**. *Psychol Rev* 2008, **115**:893-912.

30. Sederberg PB, Gershman SJ, Polyn SM, Norman KA: **Human memory reconsolidation can be explained using the temporal context model**. *Psychonom Bull Rev* 2011, **18**:455-468.

31. Gershman SJ, Schapiro AC, Hupbach A, Norman KA: **Neural context reinstatement predicts memory misattribution**. *J Neurosci* 2013, **33**:8590-8595.

32. Siegel LL, Kahana MJ: **A retrieved context account of spacing and repetition effects in free recall**. *J Exp Psychol: Learn Mem Cogn* 2014, **40**:755-764.

33. •• Smith TA, Hasinski AE, Sederberg PB: **The context repetition effect: predicted events are remembered better, even when they don't happen**. *J Exp Psychol: Gen* 2013, **142**:1298-1308.

An elegant study of recognition memory that directly supports a temporal difference learning update rule for the successor representation.

34. Manns JR, Galloway CR, Sederberg PB: **A temporal context repetition effect in rats during a novel object recognition memory task**. *Anim Cogn* 2015, **18**:1031-1037.

35. Levy WB, Hocking AB, Wu X: **Interpreting hippocampal function as recoding and forecasting**. *Neural Netw* 2005, **18**:1242-1264.

36. Lisman J, David Redish A: **Prediction, sequences and the hippocampus**. *Philos Trans R Soc Lond B: Biol Sci* 2009, **364**:1193-1201.

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

38. Jensen O, Lisman JE: **Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells**. *Learn Memory* 1996, **3**:279-287.

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

40. Mehta MR, Quirk MC, Wilson MA: **Experience-dependent asymmetric shape of hippocampal receptive fields**. *Neuron* 2000, **25**:707-715.

41. Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ: **Information theory, novelty and hippocampal responses: unpredicted or unpredictable?** *Neural Netw* 2005, **18**:225-230.

42. Bornstein AM, Daw ND: **Dissociating hippocampal and striatal contributions to sequential prediction learning**. *Eur J Neurosci* 2012, **35**:1011-1023.

43. Schapiro AC, Kustner LV, Turk-Browne NB: **Shaping of object representations in the human medial temporal lobe based on temporal regularities**. *Curr Biol* 2012, **22**:1622-1627.

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

A functional MRI study demonstrating that the hippocampus is sensitive to long-range temporal clustering of states.

45. Schacter DL, Addis DR, Buckner RL: **Remembering the past to imagine the future: the prospective brain**. *Nat Rev Neurosci* 2007, **8**:657-661.

46. Hassabis D, Kumaran D, Vann SD, Maguire EA: **Patients with hippocampal amnesia cannot imagine new experiences**. *Proc Natl Acad Sci U S A* 2007, **104**:1726-1731.

47. Squire LR, van der Horst AS, McDuff SGR, Frascino JC, Hopkins RO, Mauldin KN: **Role of the hippocampus in remembering the past and imagining the future**. *Proc Natl Acad Sci U S A* 2010, **107**:19044-19048.

48. Stachenfeld KL, Botvinick M, Gershman SJ: **Design principles of the hippocampal cognitive map**. *Advance Neural Information Processing Systems*. 2014:2528-2536.

49. Kay K, Sosa M, Chung JE, Karlsson MP, Larkin MC, Frank LM: **A hippocampal network for spatial coding during immobility and sleep**. *Nature* 2016, **531**:185-190.

50. Muller RU, Kubie JL: **The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells**. *J Neurosci* 1987, **7**:1951-1968.

51. Skaggs WE, McNaughton BL: **Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions**. *J Neurosci* 1998, **18**:8455-8466.

52. Alvernhe A, Save E, Poucet B: **Local remapping of place cell firing in the Tolman detour task**. *Eur J Neurosci* 2011, **33**:1696-1705.

53. Wu X, Foster DJ: **Hippocampal replay captures the unique topological structure of a novel environment**. *J Neurosci* 2014, **34**:6459-6469.

54. Hollup SA, Molden S, Donnett JG, Moser MB, Moser EI: **Accumulation of hippocampal place fields at the goal location in an annular watermaze task**. *J Neurosci* 2001, **21**:1635-1644.

55. •• Freyja Olafsdottir H, Barry C, Saleem AB, Hassabis D, Spiers HJ: **Hippocampal place cells construct reward related sequences through unexplored space**. *eLife* 2015, **4**:e06063.

Electrophysiological evidence that place cells are sensitive to reward predictions.

56. Rudy JW, Barrientos RM, O'reilly RC: **Hippocampal formation supports conditioning to memory of a context**. *Behav Neurosci* 2002, **116**:530-538.

57. Byrne P, Becker S, Burgess N: **Remembering the past and imagining the future: a neural model of spatial memory and imagery**. *Psychol Rev* 2007, **114**:340-375.

58. Tolman EC: *Cognitive Maps in Rats and Men*. 1948.

59. Russek EM, Momennejad I, Botvinick MM, Gershman SJ, Daw ND: **Predictive representations can link model-based reinforcement learning to model-free mechanisms**. *bioRxiv* 2016:083857.

60. Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw N, Gershman SJ: **The successor representation in human reinforcement learning**. *bioRxiv* 2016:083824.

61. Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D: **Hippocampal activation during transitive inference in humans**. *Hippocampus* 2004, **14**:153-162.

62. Dusek JA, Eichenbaum H: **The hippocampus and memory for orderly stimulus relations**. *Proc Natl Acad Sci U S A* 1997, **94**:7109-7114.

63. Garvert MM, Dolan RJ, Behrens TEJ: **A map of abstract relational**
• **knowledge in the human hippocampal-entorhinal cortex**. *eLife* 2017, **6**:e17086.
Evidence from functional MRI adaptation that the similarity structure of hippocampal representations is sensitive to predictive relationships between stimuli.

64. Brea J, Tamás GA, Urbanczik R, Senn W: **Prospective coding by**
•• **spiking neurons**. *PLOS Comput Biol* 2016, **12**:e1005003.
A spiking neuron implementation of the successor representation.

65. Caporale N, Dan Y: **Spike timing-dependent plasticity: a Hebbian learning rule**. *Annu Rev Neurosci* 2008, **31**:25-46.

66. Rainer G, Chenchal Rao S, Miller EK: **Prospective coding for objects in primate prefrontal cortex**. *J Neurosci* 1999, **19**:5493-5505.

67. Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM: **Neural representations of events arise from temporal community structure**. *Nat Neurosci* 2013, **16**:486-492.

68. Shohamy D, Turk-Browne NB: **Mechanisms for widespread hippocampal involvement in cognition**. *J Exp Psychol: Gen* 2013, **142**:1159-1170.

69. Squire LR: **Memory systems of the brain: a brief history and current perspective**. *Neurobiol Learn Memory* 2004, **82**:171-177.

70. Winocur G, Moscovitch M, Bontempi B: **Memory formation and long-term retention in humans and animals: convergence towards a transformation account of hippocampal-neocortical interactions**. *Neuropsychologia* 2010, **48**:2339-2356.

71. Manns JR, Hopkins RO, Squire LR: **Semantic memory and the human hippocampus**. *Neuron* 2003, **38**:127-133.

72. Greenberg DL, Verfaellie M: **Interdependence of episodic and semantic memory: evidence from neuropsychology**. *J Int Neuropsychol Soc* 2010, **16**:748-753.

73. Race E, Keane MM, Verfaellie M: **Losing sight of the future: impaired semantic prospection following medial temporal lobe lesions**. *Hippocampus* 2013, **23**:268-277.

74. Packard MG, McGaugh JL: **Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning**. *Neurobiol Learn Memory* 1996, **65**:65-72.