Integrating memories to guide decisions

Daphna Shohamy[1,2] and Nathaniel D Daw[3]

Adaptive decisions are guided by past experience. Yet, decisions are often made between alternatives that have not been directly experienced before, requiring the integration of memories across multiple past events. We review emerging findings supporting at least two seemingly distinct mechanisms for how the brain draws on memories in the service of choice. Prospective integration is triggered when a new decision is faced, allowing inferences to be drawn on the spot. A complementary retrospective mechanism integrates existing memories into a network of related experiences before a decision is actually faced. We discuss evidence supporting each of these mechanisms and the implications for understanding the role of memory in decision-making.

Addresses
[1] Department of Psychology, Columbia University, 1190 Amsterdam Avenue, 406 Schermerhorn Hall, MC5501, New York, NY, United States
[2] Kavli Institute for Brain Science, Columbia University, United States
[3] Center for Neural Science & Department of Psychology, New York University, 4 Washington Place, Room 809, New York, NY, United States

Corresponding author: Shohamy, Daphna (ds2619@columbia.edu)

Introduction

Memory is central to adaptive behavior. To make flexible decisions, organisms must draw on past experiences to anticipate and evaluate the outcomes of different candidate courses of action. In short, choice depends on memory. Here we review a range of research in humans and animals concerning how memories are retrieved and used to guide value-based decisions. We focus particularly on questions about when and in what order representations of previous events are built and accessed, and how this subserves the computation of decision variables to guide flexible choice.

A key issue in decision making has been distinguishing different systems for evaluating options. It is now widely appreciated that seemingly the same behaviour — a lever press or a turn in a maze — may in different circumstances arise from a number of different systems that are psychologically, neurally, and computationally distinct. Early research focused on simple ''model-free'' incremental learning, associated with the midbrain dopamine system [1,2]. This mechanism does not rely on memories for individual past events; instead it tracks a summary, like a running average, of the net rewards obtained following different actions.

This review concerns a different — though itself not necessarily unitary — class of decisions, which has been defined, operationally, by challenging organisms with choices that cannot be solved by simple model-free learning [3,4]. Such problems present novel choice options or changed circumstances, which require the organism to flexibly draw on memories of past experiences and generalize them to the decision at hand. The key requirement is *integrating* information across distinct past experiences. This is often necessary because the consequences of most decisions unfold over multiple steps — as in a chess game, or a maze. Formally, computing the value of an action in such a circumstance involves piecing together the series of consequences expected at each step [5]. Even evaluating the simplest choices–a lever press for food–requires integrating two pieces of information, about the outcome identity and its value. Oftentimes, we may not gain all this information at the same time, which can make it impossible for simple incremental learning mechanisms to track a summary.

There are many examples of humans and animals demonstrating such integrative reasoning in the laboratory. This is the defining feature of a category of behavior known as goal-directed action [3]. Neurally, there is evidence that such behaviors depend on the hippocampus and ventromedial prefrontal cortex, regions known to be essential for the formation and retrieval of memories [6]. Computationally, these behaviors have been associated with a family of theories known as model-based reinforcement learning, which evaluate actions by putting together information from a learned map or model of the different associations in the task [7]. However, the set of past experiences and possible future trajectories is vast, and there exist many different ways of solving the difficult problem of accessing some manageable subset of them to guide choice efficiently [5]. Accordingly, there are many open questions about how and when the brain puts together the pieces to produce an integrated decision variable, and what this means for the different ways in which memories can drive decisions.

For concreteness, consider a representative task, known as sensory preconditioning (Figure 1). Experience proceeds in several phases. In Phase I, subjects learn to associate two neutral stimuli, A and B. In Phase II, subjects learn that stimulus B predicts a valuable outcome, such as money. In a final decision phase, subjects are offered a choice between A and another equally familiar control stimulus. Animals and humans tend to prefer A, indicating that they have integrated information across the two distinct sets of associations (A → B, B → reward) so that they come to attribute the reward value of B to A as well, despite the fact that A was never actually rewarded.

The essential feature of this task, and many others like it, is that subjects integrate information about A's and B's associations that were obtained in separate episodes. Participants' behavior on these sorts of decisions clearly reflects processes that are beyond the simple stamping-in of any previous choice. Instead, they reflect what appear to be inferences based upon integration of the different elements.

A similar structure and logic are characteristic of a variety of other experiments, from spatial navigation (latent learning, shortcuts) [4] and instrumental conditioning (reward devaluation) [8] to learning in multiplayer games [9] and multistep decision tasks [10]. Research using all these tasks demonstrates that, at least in some circumstances, subjects can integrate multiple distinct pieces of evidence into a decision variable.

Figure 1

| Association Phase | Reward Phase | Decision phase |

Current Opinion in Behavioral Sciences
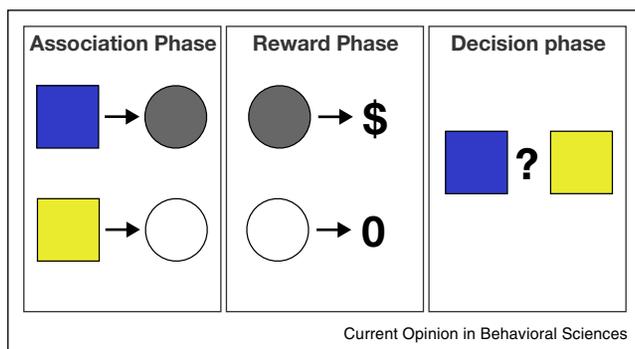
Task for measuring flexible use of memory to guide value-based decisions. In this "Sensory Preconditioning" task, participants first learn to associate pairs of stimuli with each other (e.g., squares with circles of different colors), without any rewards (association phase). Next, they learn that one stimulus (e.g., the grey circle) leads to monetary reward, while another (e.g., the white circle), leads to no reward (reward phase). Finally, participants are asked to make a decision between two neutral stimuli (blue vs. yellow squares; decision phase). Participants often prefer the blue square to the yellow square or other neutral and equally familiar stimuli, suggesting they have integrated the reward value with the blue square because of the memory associating the blue square with the rewarded grey circle. These sorts of tasks allow experimenters to measure spontaneous memory-based flexibility without requiring or rewarding it. In actual experiments, all stimuli are controlled for familiarity, likeability, value, etc.

But how, exactly, are such inferences produced? It is often assumed that this sort of behavior demonstrates something like inferential reasoning, conducted at the time of the decision. Yet this need not be the case. Instead, it is possible that such behavior reflects integration processes that happened earlier, before a decision was faced. Choice behavior itself does not generally reveal when the computations that produced it took place. Similarly, neuroscientific data — including pretraining lesions [11] or neural correlates of the final, integrated decision variables [10] — do not speak to this question. Recent work, much of it using more finely targeted neural methods, has begun to shed light on questions about how, when, and in what order, memories are accessed to support flexible decisions about novel choice options. We discuss experiments supporting the decision-time account, and then a second set of recent results suggesting that choices may instead reflect integration of memories that had occurred earlier, before a participant was ever asked to make a decision. These processes may reflect different computational strategies for model-based computation of decision variables [5].

Prospective integration

The hypothesis, implicit or explicit, behind much work on flexible decision making, is that the decision itself triggers the computation to evaluate the options, in which subjects combine the relevant associations "just in time". This is a strategy we broadly refer to as prospection (Figure 2), since it involves forecasting the consequences of some imminent potential course of action.

In some cases, it does seem almost a foregone conclusion that subjects must evaluate an option at decision time. For instance, when a truly novel option is introduced — the very first time you face an exotic dish like "tea jelly" or "pea mousse" [12••], whose value might be imagined only from the properties of its components — it seems unlikely that the brain could possibly have precomputed its value. Indeed, upon evaluating tea jelly for the first time, fMRI adaptation effects demonstrate that people access the separate component elements, tea and jelly.

Other examples from both human neuroimaging and rodent neural inactivation show that integrative, inferential reasoning in tasks similar to that of Figure 1 is accompanied or supported by neural events happening at the time decisions are faced [13–18]. In some of these cases, inferential reasoning is accompanied by activity in the hippocampus, presumably due to its role in supporting memory retrieval processes in service of inferences. Other work has highlighted the importance of the orbitofrontal cortex (OFC), consistent with the hypothesis that the OFC supports the construction of value when an organism is confronted with a new decision for which precomputed values are not available [15,19].

Figure 2

Retrospective Integration

Prospective Integration

Current Opinion in Behavioral Sciences
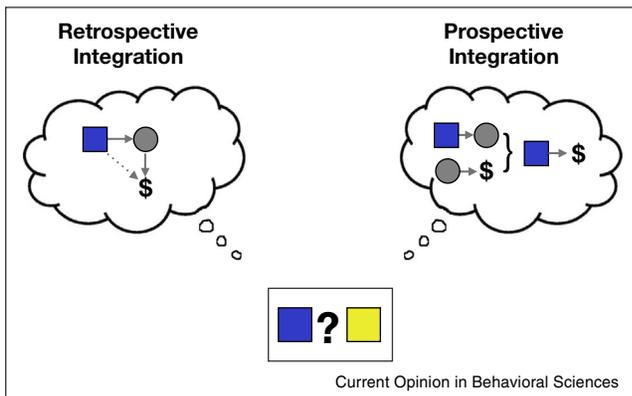
Schematic of possible mechanisms underlying integration of memories to guide decisions. When confronted with a new decision which cannot be wholly based on past rewards, such as predicting whether the blue square will lead to reward or not, participants' behavior tends to reflect the integration of memory for past relevant events. This integration can happen via two distinct mechanisms. **(a)** One possibility is that at the time of making the decision, participants retrieve relevant memories and use them to engage in prospective reasoning about the likely outcomes of their decisions. **(b)** Another possibility is that the overlap in the memories themselves triggers integration of distinct episodes during learning/encoding, before a decision is ever confronted. In this sort of retrospective mechanism, the attribution of reward value to the blue square would have already been in place before a decision was ever required.

But perhaps the most mechanistically suggestive examples arise in rodent spatial navigation, where representations of an animal's position (as represented by patterns of activity in hippocampal place cells) can ''run ahead'' of the animal's position. This can be seen at choice points [20], can predict the direction of the animal's own subsequent locomotion [21], and can be integrative in the sense of knitting together ''subroutes'' that have not actually yet been experienced sequentially [22]. A related oscillation in spatial representation, theta phase procession, was also shown to modulate in a way that reflected the animal's current spatial goal [23]. Similar anticipatory representations in hippocampus and prefrontal cortex have been observed in statistical learning tasks in humans [24,25]. Such anticipatory or ''preplay'' activity is a promising substrate for prospective computation of the value of actions. However, although hippocampal preplay appears to support correct navigational performance on simple spatial tasks [26,27], it has not so far been studied while animals were performing behavioral tasks that specifically require or demonstrate integrative, model-based decisions.

A recent experiment aimed to close this circle by linking prospective neural activity to the flexible behavior it is hypothesized to support in a single task [28]. This study used a multistep reward learning task which, much like sensory preconditioning, examined to what extent subjects, when choosing an option, integrated information about rewards recently received during other, interleaved trials with different choice options. Choices resulted in intermediate stimuli from fMRI-decodable image categories, allowing us to decode the prospective activation of the future path at decision time. Such activity was correlated, across both trials and subjects, with the extent to which choices successfully reflected model-based integration.

Separate research in the cognitive neuroscience of memory has also suggested that there are important links between memory for the past and prospective thoughts about the future. This work started from the observation that amnesic patients with severe memory loss also have difficulty imagining future events [29,30]. Neuroimaging studies in the healthy brain also link both retrieving memory and imagining future events to activity in a common set of structures, the hippocampus and surrounding medial temporal lobe [31,32]. These studies suggest a role for hippocampal memories in prospection for decision making, though this was not tested directly. A related line of evidence does suggest that prospective thinking affects evaluation. Instructing human participants to imagine events in the future attenuates the discounting of delayed rewards in an inter-temporal choice task, an effect that is correlated with activation of the hippocampus at decision time [33]. More recent findings indicate a causal role for the medial temporal lobe in this effect. This study found that, unlike healthy controls, patients with anterograde amnesia following damage to the MTL show similar amounts of discounting regardless of whether they are asked to engage in prospection or not [34••].

Retrospective integration

Despite much evidence for prospective integration mechanisms at decision time, there are also many intriguing hints that seemingly similar behaviors can be produced in a different way. Interestingly — returning to the case of position representations in the rodent hippocampus — representations of positions other than the current one do not always run ahead of the animal. Indeed, similar retrospective ''replay'' phenomena have also been reported, in which place cells represent positions where the animal has been in the past, including reverse replay behind the animal [35], or replaying spatially remote routes while the animal is pausing or asleep [36,37].

In a spatial task, the integration problem (analogous to that of Figure 1) involves, in effect, propagating information about reward availability to other, more distant locations from which the reward might be reached. Since replay events do not represent immediately available trajectories, they could allow such integration to be precomputed, rather than evoked at decision time. Consistent with this suggestion, reward receipt can trigger reverse replay, ''rewinding'' the path that preceded it [35], and this activity can even branch off to traverse an alternative

maze arm [38••]. Even more dramatically, observing reward in an inaccessible, unexplored area triggers activation, during a subsequent rest, of sequences of place cell firing that later represent the path to the reward when the area is explored [39••].

All of these dynamics strongly suggest a mechanism for the integrative propagation of reward information opportunistically and ahead of time, rather than on-demand when facing a decision. However, such retrospective activity in rodents has not directly been linked to the solution of behavioral tasks that demonstrably require integration. Addressing this gap, retrospective integration has also been examined in humans in the context of the sensory preconditioning task of Figure 1. Although it is possible to compute the value of A by evoking B (and then reward) at choice time — a prospective approach — it is equally possible that when B is evoked during Phase 2, this retrieves A, which is associated with reward at that point (Figure 2). fMRI and MEG studies of this task, using decodable image categories for the stimuli, support this mechanism [40,41]. These studies demonstrate that the flexible behavior at decision time — the tendency, for example, to choose A — reflects processes that happened earlier, during the learning phase. In particular, flexible decisions are correlated with activity in the hippocampus in the earlier learning phase, as well as with evidence for specific reactivation of the A stimulus at the same time as B is being rewarded. Earlier studies of several other integration tasks suggested a similar mechanism [42–44].

Similar integration might also be supported by re-activating raw experience from Phases I and II during rest periods or sleep following Phase II, driving new learning about A via offline integration. In machine learning, such replay-driven learning has been proposed in an architecture called DYNA [45], which demonstrates how actual, remembered, and model-simulated experiences can be blended together for learning. Applied to biological learning [46], this mechanism might help explain why model-based and model-free learning seem to share more neural substrates — such as dopamine and reward prediction errors — than previously expected [47–49]. Also, a recent study [50••] showed that manipulating rest periods in a revaluation design analogous to Figure 1 had effects on humans' success solving subsequent integration tests, consistent with the idea that this behavior was supported by processes occurring during rest. Night time sleep also seems to specifically enhance such integration, as shown using a transitive inference task in humans [51].

Conclusions and future questions
Altogether, data support the idea that memories are retrieved and integrated to construct decision variables at a variety of times, ranging from the time of encoding to the time of decision. These mechanisms are clearly not mutually exclusive, and indeed there is good evidence supporting each of them. This raises a larger question, though: how does the brain decide which sorts of strategies to evoke under which circumstances?

Earlier work has considered rational accounts of the tradeoff between model-based and model-free decision making in terms of the relative costs (e.g., delay) and benefits (e.g., better chance of gaining rewards) of prospection at decision time, a tradeoff which will vary depending on issues like the amount of time pressure or training [7,52,53]. Prospective integration is highly flexible and can in principle be applied to any new situation. However, it depends on memory retrieval at the time of choice and therefore can delay execution. Retrospective integration is efficient and fast at decision time, because it allows new decisions to depend on integration that had been previously computed. Indeed, there is some evidence that decisions based on retrospective integration can be computed as quickly as decisions that require no integration at all [43]. This replay mechanism also provides for an ongoing integration of past and present experiences, connecting otherwise discrete experiences into a networked web of memories. However, precomputing all options is not possible and thus this approach is well suited for integration of experiences that already share overlapping elements, but not for integration in the service of truly new decisions.

If replay may also happen between experience and the decisions it supports, then analogous questions — when to replay? which events to replay? — might similarly be understood in terms of rational analysis of costs and benefits. Such prioritization has been the product of some work in computer science [54] though — apart from a finding that place cell sequences favor unexplored locations that contain reward over those without it [39] — there is almost no neuroscientific data addressing this question. More generally, this is a promising instance of a broader trend in cognitive neuroscience toward rational cost-benefit analyses of cognitive control phenomena [55,56].

A second question, which we largely skirted, is what sorts of memory representations are operated on by the replay and preplay operations we have considered. Classic work on model-based decision making envisions that it operates over semantic representations (like maps), which may themselves arise from the integration or average over many distinct experiences [4]. Indeed, a similar replay story — in the form of systems consolidation — has been invoked to explain the formation of semantic from episodic knowledge [57]. However, there has also been increasing interest in the possibility that decisions are themselves derived more directly from episodic information — that is, by retrieving representations of the events on individual trials from working memory or episodic memory [58,59••]. Interestingly, the hippocampal memory system is thought to be involved both with

representations of individual episodes, and with more general relational links derived from multiple experiences (as in sensory preconditioning) [60]. It remains to be understood to what extent the decision phenomena we have examined are supported by one or the other or both of these underlying representations [61].

Conflict of interest
Nothing declared.

Acknowledgements

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. Houk JC, Adams JL, Barto AG: **A model of how the basal ganglia generate and use neural signals that predict reinforcement**. *Models of information processing in the basal ganglia*. 1995:249-270.

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

3. Dickinson A, Balleine B: **The role of learning in the operation of motivational systems**. *Stevens' handbook of experimental psychology*. 2002.

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

5. Daw ND, Dayan P: **The algorithmic anatomy of model-based evaluation**. *Philos Trans R Soc Lond B Biol Sci* 2014:369.

6. Daw ND, O'Doherty JP: **Multiple systems for value learning**. In *Neuroeconomics, decision making and the brain*, 2nd ed.. Edited by Glimcher P, Fehr E.Amsterdam: Academic Press; 2014:393-410.

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

8. 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.

9. Camerer C, Hua Ho T: **Experience-weighted attraction learning in normal form games**. *Econometrica* 1999, **67**:827-874.

10. 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.

11. 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.

12. Barron HC, Dolan RJ, Behrens TE: **Online evaluation of novel
•• choices by simultaneous representation of multiple memories**. *Nat Neurosci* 2013, **16**:1492-1498.
Demonstrates how humans evaluate truly novel options — bizarre molecular gastronomy dishes — by piecing together knowledge about their components. This study is also methodologically noteworthy for its quantitative use of fMRI adaptation to track changes in these representations with experience

13. Bornstein AM, Daw ND: **Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans**. *PLoS Comput Biol* 2013, **9**:e1003387.

14. Greene AJ, Gross WL, Elsinger CL, Rao SM: **An FMRI analysis of the human hippocampus: inference, context, and task awareness**. *J Cogn Neurosci* 2006, **18**:1156-1173.

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

16. Preston AR, Shrager Y, Dudukovic NM, Gabrieli JD: **Hippocampal contribution to the novel use of relational information in declarative memory**. *Hippocampus* 2004, **14**:148-152.

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

18. West EA, DesJardin JT, Gale K, Malkova L: **Transient inactivation of orbitofrontal cortex blocks reinforcer devaluation in macaques**. *J Neurosci* 2011, **31**:15128-15135.

19. Stalnaker TA, Cooch NK, Schoenbaum G: **What the orbitofrontal cortex does not do**. *Nat Neurosci* 2015, **18**:620-627.

20. 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.

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

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

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

24. 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.

25. 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.

26. Jadhav SP, Kemere C, German PW, Frank LM: **Awake hippocampal sharp-wave ripples support spatial memory**. *Science* 2012, **336**:1454-1458.

27. Singer AC, Carr MF, Karlsson MP, Frank LM: **Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task**. *Neuron* 2013, **77**:1163-1173.

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

29. 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.

30. Wheeler MA, Stuss DT, Tulving E: **Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness**. *Psychol Bull* 1997, **121**:331-354.

31. Addis DR, Wong AT, Schacter DL: **Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration**. *Neuropsychologia* 2007, **45**:1363-1377.

32. Summerfield JJ, Hassabis D, Maguire EA: **Differential engagement of brain regions within a 'core' network during scene construction**. *Neuropsychologia* 2010, **48**:1501-1509.

33. Peters J, Buchel C: **Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions**. *Neuron* 2010, **66**:138-148.

34. Palombo DJ, Keane MM, Verfaellie M: **The medial temporal lobes
•• are critical for reward-based decision making under conditions that promote episodic future thinking**. *Hippocampus* 2015, **25(3)**:345-353.
This study of intertemporal choice in mediotemporal-lobe lesion patients provides rare causal evidence for a role of the MTL in value based

decision making, by showing that patients lack a relationship between future thinking and time discount preferences previously shown in healthy people.

35. Foster DJ, Wilson MA: **Reverse replay of behavioural sequences in hippocampal place cells during the awake state**. *Nature* 2006, **440**:680-683.

36. Louie K, Wilson MA: **Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep**. *Neuron* 2001, **29**:145-156.

37. Karlsson MP, Frank LM: **Awake replay of remote experiences in the hippocampus**. *Nat Neurosci* 2009, **12**:913-918.

38. Wu X, Foster DJ: **Hippocampal replay captures the unique
•• topological structure of a novel environment**. *J Neurosci* 2014, **34**:6459-6469.
Although hippocampal place cell ''preplay'' phenomena have gotten more attention recently, this study rekindles interest in replay by showing that it displays a number of similarly sophisticated and computationally tantalizing properties.

39. Olafsdottir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ:
•• **Hippocampal place cells construct reward related sequences through unexplored space**. *Elife* 2015:4.
This study reports replay-like phenomena in which place cell sequences foreshadow their activity in as-yet unexplored areas, which is particularly suggestive of pre-planning novel routes.

40. Kurth-Nelson Z, Barnes G, Sejdinovic D, Dolan R, Dayan P: **Temporal structure in associative retrieval**. *eLife* 2015, **4**:e04919.

41. Wimmer GE, Shohamy D: **Preference by association: how memory mechanisms in the hippocampus bias decisions**. *Science* 2012, **338**:270-273.

42. Kumaran D, Summerfield JJ, Hassabis D, Maguire EA: **Tracking the emergence of conceptual knowledge during human decision making**. *Neuron* 2009, **63**:889-901.

43. Shohamy D, Wagner AD: **Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events**. *Neuron* 2008, **60**:378-389.

44. Zeithamova D, Dominick AL, Preston AR: **Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference**. *Neuron* 2012, **75**:168-179.

45. Sutton RS: **Dyna an integrated architecture for learning, planning, and reacting**. *ACM SIGART Bull* 1991, **2**:160-163.

46. Johnson A, Redish AD: **Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model**. *Neural Netw* 2005, **18**:1163-1171.

47. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND: **Differential roles of human striatum and amygdala in associative learning**. *Nat Neurosci* 2011, **14**:1250-1252.

48. Deserno L, Huys QJ, Boehme R, Buchert R, Heinze HJ, Grace AA, Dolan RJ, Heinz A, Schlagenhauf 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.

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

50. 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.
A clever behavioral investigation that manipulates rest periods to provide evidence that offline computations, like replay, influence flexible decision making in revaluation tasks.

51. Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP: **Human relational memory requires time and sleep**. *Proc Natl Acad Sci U S A* 2007, **104**:7723-7728.

52. Keramati M, Dezfouli A, Piray P: **Speed/accuracy trade-off between the habitual and the goal-directed processes**. *PLoS Comput Biol* 2011, **7**:e1002055.

53. Pezzulo G, Rigoli F, Chersi F: **The mixed instrumental controller: using value of information to combine habitual choice and mental simulation**. *Front Psychol* 2013:4.

54. Moore AW, Atkeson CG: **Prioritized sweeping — reinforcement learning with less data and less time**. *Mach Learn* 1993, **13**:103-130.

55. Kurzban R, Duckworth A, Kable JW, Myers J: **An opportunity cost model of subjective effort and task performance**. *Behav Brain Sci* 2013, **36**:661-679.

56. Shenhav A, Botvinick MM, Cohen JD: **The expected value of control: an integrative theory of anterior cingulate cortex function**. *Neuron* 2013, **79**:217-240.

57. 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.

58. Biele G, Erev I, Ert E: **Learning, risk attitude and hot stoves in restless bandit problems**. *J Math Psychol* 2009, **53**:155-167.

59. Collins AG, Frank MJ: **How much of reinforcement learning is
•• working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis**. *Eur J Neurosci* 2012, **35**:1024-1035.
A tight combination of computational modeling and experimental investigation that verifies — and formalizes — the long-lurking worry that human performance on reinforcement learning tasks may reflect hitherto poorly understood working-memory based reasoning rather than widely invoked incremental learning mechanisms.

60. Cohen NJ, Eichenbaum H: *Memory A: the hippocampal system*. Cambridge, MA: MIT Press; 1993.

61. Lengyel M, Dayan P: **Hippocampal contributions to control: the third way**. *NIPS*. 2007:889-896.