WIREs
COGNITIVE SCIENCE

WILEY

Mixing memory and desire: How memory reactivation supports deliberative decision-making.

Journal:	*WIREs Cognitive Science*
Manuscript ID	COGSCI-711.R2
Wiley - Manuscript type:	Advanced Review
Date Submitted by the Author:	24-Aug-2021
Complete List of Authors:	Wang, Shaoming; New York University, Psychology Feng, Samuel; Khalifa University of Science and Technology, Mathematics; Khalifa University of Science and Technology, Khalifa University Centre for Biotechnology Bornstein, Aaron; University of California Irvine, Cognitive Sciences; University of California Irvine, Center for the Neurobiology of Learning and Memory; University of California Irvine, Institute for Mathematical Behavioral Sciences
Keywords:	Episodic memory, Decision making, Reinforcement learning, evidence accumulation
Choose 1-3 topics to categorize your article:	Reasoning and Decision Making (BEAL) < Psychology (BEAA), Cognition (BGAE) < Neuroscience (BGAA), Computation (BGAF) < Neuroscience (BGAA)

Mixing memory and desire: How memory reactivation supports deliberative decision-making.

Shaoming Wang[1*], Samuel F. Feng[2,3], Aaron M. Bornstein[4,5,6*]
[1] *Department of Psychology, New York University*
[2] *Department of Mathematics, Khalifa University of Science and Technology, Abu Dhabi, UAE*
[3] *Khalifa University Centre for Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, UAE*
[4] *Department of Cognitive Sciences, University of California, Irvine*
[5] *Center for the Neurobiology of Learning & Memory, University of California, Irvine*
[6] *Institute for Mathematical Behavioral Sciences, University of California, Irvine*

[*] *Correspondence: shaoming@nyu.edu, aaron.bornstein@uci.edu*

Abstract

Memories affect nearly every aspect of our mental life. They allow us to both resolve uncertainty in the present and to construct plans for the future. Recently, renewed interest in the role memory plays in adaptive behavior has led to new theoretical advances and empirical observations. We review key findings, with particular emphasis on how the retrieval of many kinds of memories affect deliberative action selection. These results are interpreted in a sequential inference framework, in which reinstatements from memory serve as "samples" of potential action outcomes. The resulting model suggests a central role for the dynamics of memory reactivation in determining the influence of different kinds of memory in decisions. We propose that representation-specific dynamics can implement a bottom-up "product of experts" rule that integrates multiple sets of action-outcome predictions weighted on the basis of their uncertainty. We close by reviewing related findings and identifying areas for further research.

Introduction

Most decisions involve some form of memory. Decades of research has focused on understanding how one kind of memory, about the summary statistics of a task or environment, are employed in the service of evaluating choice options, either through incremental learning of stimulus-outcome associations, or via extracting regularities present in the structure of the environment (Balleine, 2007; Daw et al., 2011; Dayan, 1993; Gläscher et al., 2010; Tolman, 1948). These types of memories are differentiated by their distinct representational properties and divergent neural substrates (Dolan & Dayan, 2013; Poldrack & Packard, 2003; Yin & Knowlton, 2006). Critically, however, they share in common a reliance on extensive experience — often measured within a narrowly controlled, highly repetitive laboratory task — in order to learn usable statistics (Behrens et al., 2007; Daw et al., 2011). This leaves open the question of how decisions are made on the basis of little direct experience (Lengyel & Dayan, 2008), or in complex environments from which it may be intractable to extract sufficiently detailed regularities (Kaelbling et al., 1998; Silver & Veness, 2010) — as in many real-world decisions faced by humans and animals (Lake et al., 2015; Lien & Cheng, 2000; Niv et al., 2015).

Humans and animals constantly draw on memories of the past to inform decisions about the future (Redish, 2016; Schacter et al., 2017). An emerging framework describes this phenomenon as a simulation-driven estimation process, in which decision-makers examine what might result from each available action by consulting memories of similar previous settings. This approach, generally referred to as *memory sampling* (Bordalo et al., 2020; Gershman & Daw, 2017; Kuwabara & Pillemer, 2010; Lengyel & Dayan, 2008; Lieder et al., 2018; Ritter et al., 2018; Shadlen & Shohamy, 2016; Zhao et al., 2019), can approximate the sorts of option value estimates that would be learned across repeated experience by, e.g., temporal-difference reinforcement learning (TDRL; Bornstein et al., 2017; Gershman & Daw, 2017; Lengyel & Dayan, 2008), while retaining the flexibility to diverge from long-run averages when doing so may be adaptive. At one extreme, drawing on individual memories in this way allows one to effectively tackle choice problems even in the low-data limit (e.g., in novel environments), where processes that rely on abstraction over multiple experiences are unreliable (Lengyel & Dayan, 2008).

Examining memory *retrieval* from the perspective of reinforcement learning complements the use of RL to study representation formation -- e.g. of cached values (Barto et al., 1995), motor sequences (Botvinick et al., 2009; Keramati et al., 2016; Miller et al., 2018, 2019), or environmental structure (Dayan, 1993; Gershman, 2018; Wilson et al., 2014). Therefore, we begin this review by describing the RL formulation of the computational problem of optimal action selection among immediately available options. We continue with a review of how known cognitive and neurobiological properties of long- and short-term memory retrieval in humans and animals suggest an implementation of one form of approximate solution to this problem, the stochastic sampling of past experiences. Then, we briefly introduce the mathematical framework that describes the optimal solution to two-alternative forced choice on the basis of unreliable evidence — the drift-diffusion model (DDM) — with emphasis on what is known about how organisms approach the special case of evidence in the form of internally-generated signals.

We next review theoretical frameworks and key empirical studies that describe how various kinds of memory, ranging from action sequences to "cognitive maps" to long-term autobiographical memories, can provide these internally-generated signals for action selection. We focus especially on a representative selection of studies that have shown that episodic features[1] mediate the selection of which memories are retrieved during decision deliberation; these constitute an informative limiting case of the memory sampling framework.

Next, we examine how these properties of memory retrieval during action selection constrain the process of accumulating evidence from memory. We focus on areas in which the properties of memory sampling contrast with those of sensory evidence accumulation, such as

[1] We use the term "memories with episodic features" to refer to representations of past experience that exhibit dense, multi-sensory associations, formed during a single experience, which potentially include attributes incidental to goals at the time of that experience (Allen & Fortin, 2013; Bornstein & Pickard, 2020 Box 1). Though "episodic memory" has variously been defined by its relationship to conscious, declarative recall, these properties may not be functionally necessary to an influence on choices, and so we sidestep the question of awareness in the present review.

the relationship between representational properties and retrieval dynamics, and the sequential structure of retrieval.

We close with a synthesis of the reviewed findings, and suggest that action selection based on memory retrieval can be best described by a time-varying evidence accumulation process, in which the momentary rate of accumulation is determined by several cognitive and neural factors. The resulting model approximates a "product of experts" rule for integrating action tendencies from multiple control processes — in this case, memory representations with different associative content, relational structure, and history-dependence. It follows directly that the involvement of different forms of memory in action selection depends on the temporal dynamics of these factors, *via* their influence on the effective rate of production of evidence samples, which can implement the principle of uncertainty-weighted arbitration between different decision systems (Daw et al., 2005; Keramati et al., 2011). We close with a brief review of existing empirical evidence in support of this model, and suggest potential directions for further research.

I. The view from Reinforcement Learning

We begin by detailing key aspects of the predominant framework for value-based decisions, Reinforcement Learning (RL; Sutton & Barto, 2018). We begin here because memory sampling shares with RL the use of primitives such as *states*, *actions,* and *rewards* -- but, crucially, it operates on these elements with a different computational form that provides a distinct set of guarantees about efficiency and optimality. Understanding these provides the basis for understanding why each approach makes different empirical predictions in certain settings. Importantly, RL provides a formal understanding of the value estimation problem, and thus for evaluating different kinds of estimates. This framework will be crucial for understanding our later description of how and why multiple memory systems can contribute to decisions.

RL examines the problem of learning how to best navigate an uncertain environment guided primarily by feedback, in the form of reward or punishment, obtained after taking actions within that environment. While the framework allows for a wide range of possible approaches, its primary applications in neuroscience research to date have followed a particular form involving incremental learning of a *value function*[2] relating *states* and *actions* to the long-term, *discounted* rewards that can be expected to result (Eqn. 1). When fitting human behavior, a common practice (Daw et al., 2011) is to specify an action selection function that translates these values into a likelihood of taking each available action (Eqn. 2). We next describe particular instances of these equations and the key features relevant to the current review:

$$Q(a,s) \leftarrow Q(a,s) + \alpha[R + \gamma \max_{a'} Q(a',s') - Q(a,s)] \quad (1)$$

$$P(a^* == A) \propto \frac{\exp[\beta Q(A,s)]}{\Sigma_{a'} \exp[\beta Q(a',s)]} \quad (2)$$

[2] Multiple variants of each equation achieve similar goals under different settings. For more in-depth treatment, see Sutton & Barto (2018); for a review of the neural instantiation of these variables, see (Glimcher, 2011).

The first equation describes the incremental, experience-driven learning of value expectations (the value function, *Q*). The quantity specified by the value function is an estimate of the total *future* reward expected after taking action *a* in state *s* (and continuing to act optimally thereafter). This future reward is the sum of the reward directly obtained by taking the action (*R*), plus the total future reward to be obtained by taking the *best* action in the ensuing state *s'*. (Future rewards are, throughout, treated as less important to momentary action selection than immediate rewards, so they are discounted according to a constant 0 < □ ≤ 1.) The expectation is updated by the difference between this sum and the previous value of the expectation, after scaling by a learning rate ($0 < \alpha \leq 1$) in order to regularize the estimate. The second equation specifies the probability of choosing a given action (*A*) as the relative profitability of that action, versus all candidate actions. The sensitivity of this likelihood to the value difference is specified by the *temperature* parameter, □.

Importantly, the first equation is an approximation to the full value computation (Eqn. 3), which incorporates knowledge about the *transition structure* of the world — the likelihood that taking a given action *a* in state *s* is going to lead to a particular state *s'*. The true discounted future reward thus integrates over transition probabilities to all possible successor states. An agent with knowledge of this transition structure may be able to make better decisions than one who just learns reward values, but representing and working with this structure can be quite costly.

$$Q(a,s) = \sum_{s'} T(s,a,s')V(s') \qquad (3)$$

Note that the future return of the target states, V(s'), is recursively defined:

$$V(s') = R(s') + \gamma V(s'') \qquad (4)$$

Unrolling the recursion gives a converging sum of (discounted) rewards:

$$V(s') = R(s') + \gamma R(s'') + \gamma^2 R(s''') + \ldots \qquad (5)$$

where future states after s' are denoted by s'', s''', and so on. Computing this (recursive) expectation is difficult in practice, especially with limited experience of the transition structure. Therefore, approximate computations may be employed, either the incremental approach of Equation 1 above, which marginalizes over transitions, or via methods that directly estimate the transition structure (Daw et al., 2005). More broadly, however, the computational goal — choosing on the basis of total discounted future reward — can be achieved in multiple ways.

One approach, called *memory sampling*, avoids the dependence on extensive experience by simply consulting the values obtained directly, "remembering" *individual* experiences with the current (and potential future) state(s). Formally, rather than computing this estimate by updating a cached value function with each experience (Eqn. 1), the alternative computes it dynamically, possibly even on-demand (Eldar et al., 2020), by *sampling* past encounters with the states of interest (and, potentially, *generalizing* from similar states) and averaging the resulting values. This approach can be used to estimate both the reward to be received from the current action (Bornstein et al., 2017), and also that of states that follow from each action (Bornstein & Norman, 2017; Gershman & Daw, 2017; Vikbladh et al., 2017). When multiple relevant experiences exist, they can be selected from according to a sample-selection function (Fig. 1; Equation 6a, function *S*), that specifies some probability distribution over rewards for each action given by the distance between current state *s* and given sample state *s'* in a probability space defined over their shared features (Eqn. 6b). While in practice this distance incorporates any set of features relevant to the current comparison (Fig. 1), in laboratory experiments task states are usually distinguishable along only a small number of well-controlled dimensions. For example, samples could be weighted by their proximity in time to the current moment (Eqn. 6c) — capturing the intuition that the remembered states most like the state I am currently in are the states I have most recently visited. In this formulation, samples at time *t* are most likely to be drawn from the most recent trial (i=t-1), and exponentially less likely to be drawn from preceding trials *i* (i.e. where i=t-2, t-3, t-4, …), with decay specified by the parameter α. Because the value of α is between 0 and 1, exponentiating this value by t-i will result in progressively smaller probabilities for trials further in the past (greater *i*). Values estimated by this approach have the same form of dependence on recent experience as do those learned by TDRL (Bornstein et al., 2017).

$$(s', r') \leftarrow S(s, a) \tag{6a}$$

$$P(Q(a,s) == R(s')) \propto ||s - s'|| \tag{6b}$$

$$P(Q(a,s) == R_i) = \alpha(1 - \alpha)^{t - i} \tag{6c}$$

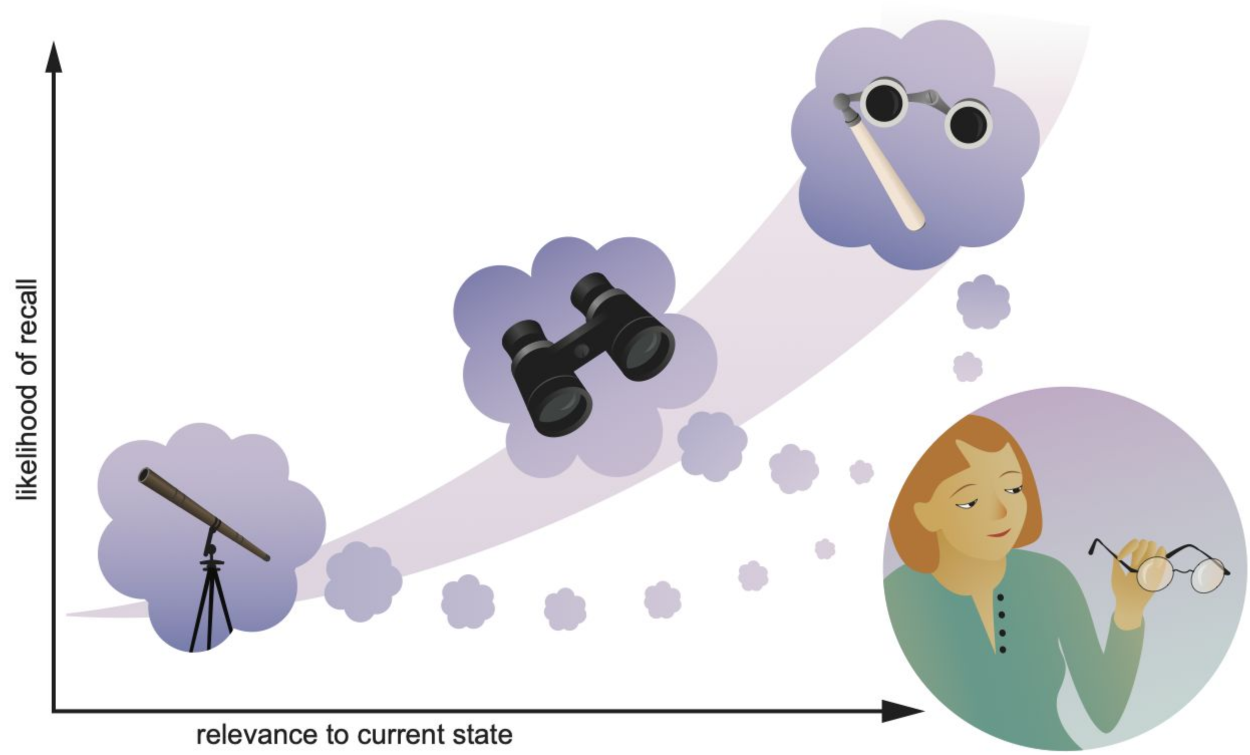


Figure 1. Relevance-based retrieval of memories. The memory sampling framework (Eqn. 6) involves the probabilistic retrieval of memories according to their *relevance* to the current state. This relevance may be assessed across any number of dimensions or attributes, depending on the task at hand. In the illustrated example, the decision-maker is examining a pair of eyeglasses and deciding whether they are useful for her current goal (e.g. watching a play). In doing so, she retrieves memories of past experiences with similar items. The most likely item to remember is the one most relevant to the current state. Other items, of decreasing relevance, may also be retrieved, though are progressively less likely according to their usefulness for viewing events at various distances.

Sampling from past experiences can also in principle approximate the extended sum of Equation 5, by leveraging the sequential structure of memory retrieval (Weidemann et al., 2019) to serially sample experiences from successive states (rather than a single state, as presented in Equation 6) and integrate them[3] (Bornstein & Norman, 2017; Gershman & Daw, 2017). Though this process is less resource-efficient than TDRL, it is more flexible: Specifically, it can generate reliable estimates even after just a few experiences in an environment (Lengyel & Dayan, 2008), can dynamically adjust to momentary goals (Bornstein & Daw, 2013), and can smoothly incorporate newly available information about transition or value functions (Vikbladh et al., 2017). These features arise when the sample selection process admits many possible Monte Carlo approximations to Q(a,s) — in other words, by sampling from multiple memory stores that represent experiences in different forms (Bornstein & Daw, 2013). Depending on which representation is being sampled from, these approximations can be wholly nonparametric, in the limit of individual samples with episodic features that also carry direct reward signals (Bornstein et al., 2017), or it can include sequences of actions (Smith & Graybiel, 2013) or states (Fortin et al., 2002; Pezzulo et al., 2014) bound together across repeated experience and terminating in a given outcome (Keramati et al., 2016). Sequences sampled in this way can be probabilistic in nature, for instance in "map-like" representations of the history of transition experience that have abstracted away reward, allowing them to be combined with local reward information (Dayan, 1993; Gershman, 2018). Evidence supports the existence of multiple such maps, connecting states at different levels of resolution reflecting different histories of integration (Bornstein & Daw, 2012; Brunec et al., 2018; Collin et al., 2015; Jiang et al., 2015; Madarasz & Behrens, 2019; Samejima & Doya, 2007).

Finally, it is important to note that although the above formulation is written in terms of reward values, the end result of the process is to select actions[4]. If we assume that action probabilities are proportional to (relative) action values (Eqn. 2), and because we are describing the two-alternative case[5], then each memory sample, by contributing to the estimate of P(choose A), also updates the (relative) likelihood of a given action being preferred (i.e. $\log\frac{P}{1-P}$). Understanding memory sampling as sequential inference of the "best" action to take connects it to the Sequential Probability Ratio Test (SPRT: Laming, 1968; the correspondence between online planning and sequential inference was also noted by Solway & Botvinick, 2012) and, by

[3] The full equation describing sample-averaging is an expansion of Equation 6, and is omitted here for space reasons. See the supplemental materials of (Bornstein et al., 2017; Bornstein & Norman, 2017) for the expanded form.

[4] Indeed, several frameworks propose that memory retrieval plays a direct role in action selection, rather than being mediated by value estimation (Henson & Gagnepain, 2010; Pezzulo et al., 2019; Wang et al., 2015). Recent evidence supports the general idea that decisions for reward are actually deliberated in action space, rather than with values intermediating (Koechlin, 2019), and that the effect of memory on subsequent preferences is only present when the memory evokes a choice, rather than an item presented in the absence of choice (DuBrow et al., 2019). The distinction between deliberating in terms of values and deliberating in terms of actions is important, with consequences both in the shape of behavioral variability and the understanding of the substance of neural representations; though outside the scope of this review, we refer the reader to (Hayden & Niv, 2020) for an excellent discussion of the implications.

[5] Though a similar procedure can apply to the multialternative scenario (Baum & Veeravalli, 1994).

extension, to the canonical evidence accumulation algorithm, the drift-diffusion model (DDM; Bogacz et al., 2006; Busemeyer & Diederich, 2010; Ratcliff & Smith, 2004).

II. Memory in action

Extensive recent findings support the idea that action selection is influenced by memories -- even of individual experiences -- retrieved at the point of decision. One example is found in a series of studies by Ludvig, Madan, and Spetch (Ludvig et al., 2015; Madan et al., 2014, 2015) who showed that individual choices between risky lotteries are influenced by reminders of past choices (and their outcomes), guiding individuals towards riskier options when they were reminded of choices on which they had been "lucky" in the past. These effects were observed within a single lab session, but Wimmer & Poldrack (2018) demonstrated that the sense of "luckiness" associated with reward-associated memoranda was detectable in explicit elicitation at least three weeks later.

A different study examined participants as they learned the values of trial-unique lotteries and performed a decision-making task between learned and novel lotteries (Murty et al., 2016). They found that participants were more likely to re-engage with learned lotteries that had previously resulted in higher rewards, but only for lotteries whose values were correctly identified in a subsequent recognition memory test. These results suggest that memories about *specific* rewarding events are successfully encoded and then subsequently reactivated upon a second encounter, consistent with the idea of evidence arising from discrete packets, and with an evaluation function that is predicated on the value experienced in that previous episode, rather than one computed anew. However, these data could also be consistent with separate effects of positive reward prediction errors on choice and memory (Jang et al., 2019; Rouhani & Niv, 2021). The question of whether memory sampling *requires* explicit recollection at the time of choice remains an area of active interest.

In another study, participants learned the value of repeated options through choice and feedback, which was presented alongside trial-unique images of everyday objects (referred to as "tickets"; Bornstein et al., 2017). Choice trials were interspersed with recognition memory probes that implicitly reminded participants of selected past choices. Tracking the average value of each option via incremental learning is a profitable approach to performing the choice task. However, when choices were preceded by memory probes, participants' decisions were biased by the action taken and the value received on the trial where the images were first introduced. This result was captured by a memory sampling model which treated the probed experiences as more recent than they would otherwise have been (Eqn. 6c). This matched previous work suggesting that decisions which appeared to be a running average of recent rewards could instead be better captured by an algorithm that relies on single samples of past trials (Biele et al., 2009), and extended the idea by linking the samples to episodic memories.

Bornstein and colleagues (2017) also used the same model to reanalyze previously collected data from a four-choice decision task (Daw et al., 2006), which further revealed that in addition to participants' choices, neural decision variables measured in fMRI were better explained by a memory sampling model than by TDRL. Although forming and retrieving individual memories is thought to be more cognitively demanding than maintaining summary statistics of a task (or a semanticized model; Daw et al., 2005), these results indicate that individual memories of past rewards influence choice even under the circumstances where they may not be locally relevant to task performance.

The idea that sampling draws on episodic representations implies that the sampling process should reactivate richly associative information, which could also guide both action selection, and also the content of successive samples. A critical feature of episodic memory, as originally defined (Donaldson & Tulving, 1972), is that it is situated within time and place, bound up with other events that occurred in a contiguous associative mental context. Critically, this context need not be explicitly temporal: the associative nature of mental context is not identical to the sequence of experiences, but may be instead or also sculpted by latent or semantic associations, a point we return to below. Supporting the idea that sample selection changes as a result of memory reactivation, recent computational, behavioral, and neural work has shown that encoding context affects the sequential structure of memory retrieval: when we recall an event from a context, the next memory to be recalled is likely to be one from the same context (Folkerts et al., 2018; Howard & Kahana, 2002; Socher et al., 2009). In terms of Equation 6a, recent memory reactivations are a component of s. Crucially for the process of action selection, sequential memory retrieval can proceed along dimensions that may be informative about a range of option values (e.g. multiple flavors of ice cream tried at the same shop). This means that, rather than simply serving as repeated samples of the same reward, successively recalled events may have different, even opposing, action and reward implications.

This sort of *context-guided memory sampling* was demonstrated using a variant of the "ticket" bandit task previously discussed, altered such that memories with shared associative content ("context", indicated by photographs of scenes) sharply differed in which action was most likely to be rewarded (Bornstein & Norman, 2017). This allowed a dissociation of the influence on choice of individual event reinstatement from that of ensuing reinstatement of events sharing that context. When probed with a cue reminding them of a particular choice event, participants' subsequent choices were influenced by the properties of other decisions made in the same context as the reminded one; critically, this effect was mediated by neuroimaging markers of whether -- and *which* -- visual context was retrieved at the time of the decision, even if that retrieved context was not the one actually experienced, supporting the hypothesis that the value estimate is constructed at retrieval time, rather than being imbued in the reminder cue. The correlation between this behavioral effect and the specific, momentary content of memory retrieval suggests that factors that modulate memory reactivation also influence choice, and thus that these reinstatements are used to estimate values at the time of decision. The memory modulation effect has also been widely observed in other studies, where results indicate that decisions made in familiar contexts are more likely to be influenced by past events than decisions in novel contexts (Duncan & Shohamy, 2016), consistent with the notion that context is part of the input to the selection function; that remembered options are more likely to be chosen as compared to forgotten ones despite the fact that the chosen options are comparatively unattractive (Gluth et al., 2015; Mechera-Ostrovsky & Gluth, 2018); that the opposite pattern holds when both options are in the loss domain (Weilbächer et al., 2020), consistent with the idea that memory samples reduce uncertainty in the value estimate; and that inducing imagination of episodically rich future scenarios alter impulsivity and risk-taking behavior, suggesting that reactivating episodic memory may be a shared mechanism during both decisions from experience and those that involve simulating potential future events on the basis of past experience (Peters & Büchel, 2010; St-Amand et al., 2018).

In addition to decisions that involve re-engaging with previously experienced options, pattern completion (see Section IV, below) allows memory reactivation to also support decisions about never before seen options. For example, Barron and colleagues (2013) asked participants to choose between novel food items that are combinations of two familiar food types that had not been previously tested together (Barron et al., 2013). They found that the prospective values of the novel items are constructed at choice time through simultaneously re-activating memories of its constitutive parts in the hippocampus and medial prefrontal cortex. This finding resonates with proposals that representations in these regions are predictive in nature (Bornstein & Daw, 2012, 2013; Gershman, 2018; Hamm & Mattfeld, 2019; Morton et al., 2017, 2020; Schacter et al., 2012; Shohamy & Wagner, 2008; Stachenfeld et al., 2017; Zeithamova et al., 2012). A key property of these representations is that they can be formed in the absence of explicit goals. For instance, a seminal study by Wimmer and Shohamy (2012) found that, in the absence of conscious awareness, value learning through repetition also recruited hippocampus, and that this hippocampal activity supports the transfer, or "spread", of value between paired stimuli. This idea has been extended to networks of rewards and stimuli related via complex, latent associative structures (Wu et al., 2018).

Supporting the idea that these learned regularities support sensory and motor predictions, studies using sequential stimulus identification tasks have shown that hippocampal activity increases with the uncertainty over possible successor stimuli (Bornstein & Daw, 2012; Harrison et al., 2006; Hindy et al., 2016; Kok & Turk-Browne, 2018; Strange et al., 2005). Taking into account the spatial and temporal resolution of fMRI (Mayes et al., 2019), these findings are consistent with observations in rodent electrophysiology studies that hippocampus is continually "prefetching" possible next-step stimuli in order to inform action preparation, and that more prefetching occurs in times of higher uncertainty about the next element in the sequence (Johnson & Redish, 2007; Redish, 2016). Indeed, this appears to be true even in simple sequential responding, of the sort traditionally linked to striatal representations. For example, Bornstein & Daw (2012, 2013) demonstrated that forward-looking activity in both hippocampus and striatum contribute to such learning, with distinct quantitative signatures of the timescale across which they integrate stimulus history to generate predictions. Maintaining multiple representations with different history dependence may be adaptive in environments of unknown or changing volatility (Iigaya et al., 2019; Yu, 2007), and concords with extensive empirical work supporting a diversity of integration timescales across brain regions (Brunec et al., 2018; Gläscher & Büchel, 2005; Meder et al., 2017; Murray et al., 2014; Onoda et al., 2011) and expressed in behavior (Corrado et al., 2005; Staddon & Davis, 1990).

Taken together, the above findings outline a clear role for mnemonic and relational reactivation during decisions about the past and future. This reactivation process is stochastic, is influenced by multiple aspects of the memory representation, supports both novel and repeated decisions, and adaptively selects memories on the basis of their predictive value to the decision at hand. We now turn to the question of how this information is transformed into action.

III. Evidence from memory

We briefly review the standard model of single-trial action selection, sequential evidence accumulation (Bogacz et al., 2006; Ratcliff, 1978). Though questions remain about its exact instantiation in neural circuits (Brody & Hanks, 2016; Gold & Shadlen, 2007), there is widespread support for the idea that a sequence of neural structures are involved in successively signaling momentary sensory evidence in favor of candidate actions, integrating this evidence across time and heterogeneous neural populations, and transforming the resulting timeseries into motor responses, and that the evolution of this time-integrated signal is strikingly well-matched by a biased random walk, approximated in the continuum limit as Brownian motion along a gradient (Ratcliff & McKoon, 2008).

Experiments using this framework are generally constrained such that action-relevant evidence is available only in a single sensory modality (e.g. visual or auditory input). These unimodal evidence signals have multiple downstream effects: neural firing patterns in several successive regions reflect the accumulation of sensory input. These structures carry out distinct transformations of the input, or combine it with other signals (Akrami et al., 2018; Erlich et al., 2015; Hanks et al., 2015; Scott et al., 2017; Yartsev et al., 2018). It remains an open question what is the precise contribution of each of these multiple components. Importantly, even in these tightly controlled settings, neural firing has been shown to reflect changing internal representations of the inferred, latent structure of the environment (Hanks et al., 2011; Yang & Shadlen, 2007). This is likely a special case of a more general property. Namely, when all of the information necessary to make a decision is not actively present in the sensorium or the current mental context — which is arguably the case for nearly every decision made outside of laboratories, as well as many inside of them — the brain must, by definition, rely on reactivation of representations formed during past experiences. Despite this, and despite the fact that early applications of the canonical form of the model were to recognition memory (Ratcliff, 1978), the lion's share of experimental applications over the past four decades have focused on other kinds of decisions. However, findings about the neural architecture of evidence integration in these other modalities are likely to apply to the study of memory-guided decisions, especially when studies employ stimuli whose predictiveness is estimated via associations that emerge across experience (Yang & Shadlen, 2007). As reactivations of those previous experiences echo both previous sensory inputs and also latent, non-sensory information, such as the inferred contingency structure of the environment and the value of rewards available at the time, all of these lead to the subsequent reactivation of the same sorts of action-tendency or value associations as does sensory input. In other words, stimuli may trigger action-related evidence directly as well as via associations with other stimuli which themselves may trigger action-related evidence (Bornstein & Norman, 2017) (though the latter signals may be integrated into the decision calculation at a later time, a point we return to below). A potential synthesis of this necessary corollary with the existing data is that accumulation-reflecting activity downstream from early sensory regions actually represents the integration of multiple inputs, including memories (Bakkour et al., 2019; Mainen & Pouget, 2019).

Mathematical models of sequential inference: Gaussian and "Jump" diffusion.

We now turn to the model itself, which has been a rich area of investigation for over four decades. Here we will only cover a few key points relevant to the review, and refer the reader to several excellent treatments for further details (Bogacz et al., 2006; Gold & Shadlen, 2001; Ratcliff, 1978; Ratcliff & Rouder, 1998).

In canonical form, the DDM is specified as a one-dimensional biased random walk in continuous time, where a decision variable (*x*) is incremented at each time point by a step of average size *Adt*, corrupted by some zero-centered gaussian white noise with standard deviation *c* (*cdW*), as in Equation 7.

$$dx = Adt + cdW \tag{7}$$

Integrating these steps over time, the walk continues until it arrives at one of two absorbing thresholds. At this point, the walk terminates and the action is selected according to which threshold was reached. Thus, the model specifies both the choice made and the time needed to make the decision. This procedure is the continuous-time limit of the Sequential Probability Ratio Test (SPRT), a simple arithmetic procedure for determining which of two hypotheses are supported by a stream of noisy evidence. This equivalence is important because Wald & Wolfowitz (1948) proved that, given a fixed error rate, the SPRT determines the solution after the fewest number of samples. Thus, the DDM describes the optimal procedure for weighing evidence in two alternative forced choice, under reasonably broad assumptions[6].

The SPRT operates by examining whether the likelihood ratio (Eqn. 8a), the conditional probability of each hypothesized stimulus (s_1 and s_2) given the evidence (*e*) observed, reaches a predetermined threshold that corresponds to the desired level of accuracy. When multiple samples ($e_1 \ldots e_n$) are observed, the gross likelihood ratio is simply the product of these individual terms (Eqn. 8b). Gold and Shadlen (2001) proposed that neural circuits could implement evidence accumulation by computing this product in log space. Representing this quantity in logarithmic form allows it to be implemented as a successive summation (Eqn. 8c), which can naturally be implemented by neurons (up to normalization constraints, see (Keung et al., 2020).

$$LR_{1,2|e} = \frac{P(e|s_1)}{P(e|s_2)} \tag{8a}$$

$$LR_{1,2|e_{1..n}} = \frac{P(e_1|s_1)}{P(e_1|s_2)} \times \frac{P(e_2|s_1)}{P(e_2|s_2)} \times \frac{P(e_3|s_1)}{P(e_3|s_2)} \times \frac{P(e_4|s_1)}{P(e_4|s_2)} \times \frac{P(e_5|s_1)}{P(e_5|s_2)} \times \ldots \tag{8b}$$

$$\log LR_{1,2|e_{1..n}} = \log\frac{P(e_1|s_1)}{P(e_1|s_2)} + \log\frac{P(e_2|s_1)}{P(e_2|s_2)} + \log\frac{P(e_3|s_1)}{P(e_3|s_2)} + \log\frac{P(e_4|s_1)}{P(e_4|s_2)} + \log\frac{P(e_5|s_1)}{P(e_5|s_2)} + \ldots \tag{8c}$$

---

[6] Again, a similar form, though with important differences, results when solving for the optimal policy in the multialternative case (Tajima et al., 2019; Baum & Veeravalli, 1994).

Bogacz et al (2006) rearranged these terms to denote the logLR as integrated evidence (I_t) and show that the summation is a recursion which takes the form of a discrete random walk (with stochasticity inherent in the densities given by the evidences e_t):

$$I_t = I_{t-1} + \log\frac{P(e_t|s_1)}{P(e_t|s_2)} \quad (8d)$$

Gold and Shadlen further noted that one benefit of forming decisions in this way is that it provides a "common currency" in which to represent multiple kinds of evidence besides just sensory input, such as prior probabilities. However, in the DDM the drift rate term specifies the average net instantaneous direction of the evidence summation series. That is, it averages out any ephemeral fluctuations in the relative weighting. This is a valuable approximation for tasks with stationary evidence consistency, but breaks down in cases where the properties of arriving evidence fluctuate over time (Wong et al., 2007). Outside of tightly controlled perceptual experiments, evidence may be more like these latter cases. For instance, consumption decisions implicitly aggregate multiple sources of evidence, including sensory input, internal state (e.g. cravings for a particular flavor), and history-dependent representations of the stimulus, each of which may have different properties that could, when those options are examined, alter the momentary drift rate. As a result, the static vector specified by the drift rate may obscure underlying heterogeneity in net direction of evidence.

Along these lines, a variety of alternatives to the "pure" DDM have been proposed. These include time-dependent drift rates, time-dependent thresholds, and non-Gaussian noise (Ratcliff & McKoon, 2008; Srivastava et al., 2017; Voss et al., 2019; Wieschen et al., 2020). These alternatives sacrifice the analytical tractability and theoretical connection to the optimal SPRT in favor of better modeling the underlying stochastic dynamics that give rise to response times. One especially promising approach for modeling the arrival of evidence samples from different distributions, called Lévy Flight models (Fig. 2), considers a variety of intermittent "jumps" that augment and alter the Brownian motion of Equation 7. Recent work on these "jump-diffusion" models suggest that they provide a superior fit to two alternative forced choice data in situations where evidence sources are of varying reliability, are mixed with prior probabilities, and/or differ in the distribution of their arrival times (Voss et al., 2019; Wieschen et al., 2020). In the next section, we review features of memory representations that suggest that these conditions are likely to hold in general when sampling from memory.

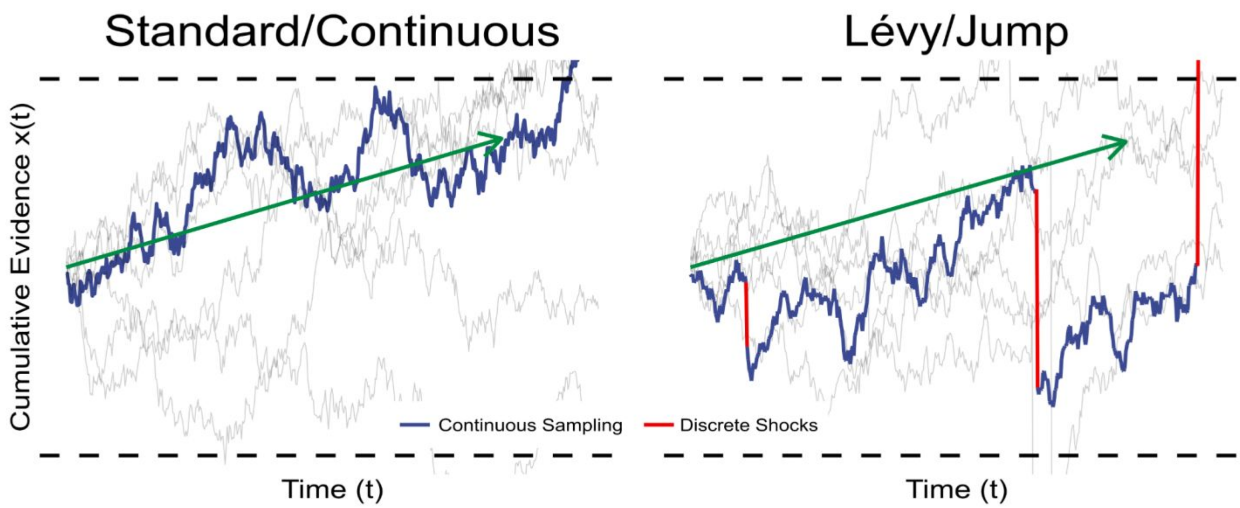


Figure 2. Lévy Flight models add discontinuous jumps to standard diffusion models. Evidence accumulation models describe the integration of evidence samples across time by their average net direction and magnitude of accumulation (green arrows), which dictate the rate at which evidence tends to reach a fixed threshold (dashed black lines). This average obscures considerably heterogeneity both across time within a single decision and also across multiple trials examining related decisions (gray lines). Recent work examines the proper distribution model for describing the variability of these accumulation timeseries. Standard/Continuous models of evidence sampling (Left) and Lévy or Jump models (Right) both have a noisy, continuous component for infinitesimal sampling (blue lines), however, Jump models add the option for sampling discrete shocks from an alternative evidence distribution (red lines). It has been shown that response times in a general class of deliberative decision tasks are better fit when these jumps are added to the standard evidence accumulation timeseries (Voss et al., 2019; Wieschen et al., 2020). An open question is what mechanisms produce these jumps. Here, we propose that one mechanism by which such jumps arise is via parallel sampling from multiple internal evidence sources which produce evidence at different latencies and frequencies.

IV. Mechanisms of memory encoding and retrieval

In this section, we outline the features of *content* and *process* (Zhao et al., 2019) that mediate the impacts of memories on decisions. Specifically, we describe multiple kinds of memory representations, how they differently represent aspects of past experience, and how they lend themselves to different retrieval and transformation dynamics that later affect decision-making.

Content

Significant ongoing work addresses the question of what representations are supported by the hippocampal memory system, and how these representations adapt over the course of experience and rest (Kumaran et al., 2016; Schapiro et al., 2017; Stachenfeld et al., 2017; Yonelinas et al., 2019). A consensus is emerging that multiple representations in the hippocampal formation and adjoining cortical regions are progressively tuned to support adaptive reward-seeking behavior, and that these representations restructure experiences to create “maps” that organize even abstract concepts according to spatial-like codes (Behrens et al., 2018; Bellmund et al., 2018; Vikbladh et al., 2019). Such representations are computationally desirable because they allow complex planning behaviors to be quickly approximated via operations akin to vector products (Gershman, 2018). However, biological agents are likely never truly certain of their current "state", and so some degree of uncertainty carries forward through all operations (Courville et al., 2006; Dayan et al., 2000; Geerts et al., 2019; Soltani & Izquierdo, 2019). With its ability to extract sparse codes from sensory inputs, hippocampus is implicated in the learning of uncertain states by representing the latent contexts that give rise to observations (Gershman et al., 2010; Sanders et al., 2020). Such representations may enable inference about which memory samples should be drawn with partial information about the structure of the environment (Gershman et al., 2015). We now review in detail what is known about the content of representations supported by the hippocampus (*relational* or *latent*).

Stimulus-stimulus relational representations

The influential cognitive map theory proposed that animals encode a mental representation of the environment that reflects the relative locations of objects within it (Tolman, 1948). The theory has particularly influenced the study of spatial navigation, which shows that neurons in the hippocampus are tuned to encode the relations between different locations (O'Keefe & Nadel, 1978). Subsequent work demonstrates that different routes coded in the animal's hippocampus are reactivated and evaluated before an animal enters the same environment (Johnson & Redish, 2007), and can sometimes reflect novel routes that have not actually yet been experienced (Gupta et al., 2010). Recent evidence suggests that similar neural representations, both in the hippocampus and in adjoining medial temporal cortical regions, could also be involved in encoding the relationships between non-spatial objects. Across several recording modalities and model organisms, such flexible yet structured relational codes have been observed in domains as varied as temporal relations (Garvert et al., 2017; MacDonald et al., 2011), sound frequencies (Aronov et al., 2017), conceptual features (Constantinescu et al., 2016; Theves et al., 2019), social relations (Park et al., 2020; Tavares et al., 2015) and sequential planning (Bornstein & Daw, 2013; Doll et al., 2015; Vikbladh et al., 2019). While these codes are observed in distinct (though adjoining) regions and reflect different types of relational coordinate systems, it is widely thought that they serve complementary roles in a general relational network centered on the hippocampus that together reflect the associative structure between events (Eichenbaum & Cohen, 2014; Preston et al., 2004; Shohamy & Wagner, 2008; Zeithamova et al., 2012). Such representations support inferences that necessitate integrating over multiple distant episodes. For instance, one study asked participants to make novel decisions that require integration across episodes with overlapping elements, and found that the activation patterns in the hippocampus during learning predict how well experiences were integrated in support of novel decisions (Shohamy & Wagner, 2008). These studies point to a role of hippocampus in coding relational representations between observations, be it spatial locations or discrete events (Schlichting & Preston, 2017).

Recent advances in the field of reinforcement learning provide a theoretical account of these various relational representations (Gershman, 2018; Stachenfeld et al., 2017), which can potentially unify the above-described theoretical frameworks and empirical findings. Specifically, it is suggested that the place cells in the hippocampus encode the expected occupancy of future states (or locations) following the current state, generally termed as encoding a "successor representation" (Dayan, 1993). The key insight of the theory is that rather than encoding place in an absolute sense, the place cells encode a predictive representation of future states that reflects the relational structure between them (Stachenfeld et al., 2017). As a result, two states that predict similar future states will have similar representations, regardless of their physical adjacency. This idea allows the theory to account for not only a wide range of neurophysiological phenomena in rodent spatial tasks, but also findings that are built on discrete, abstract relational knowledge.

Finally, it has been shown that the relational representations coded by the hippocampus can be used to drive adaptive behavior when combined with reward information, whether learned by experience or instructed (Bornstein & Daw, 2013; Doll et al., 2015; Wimmer & Shohamy, 2012). For example, in the Wimmer & Shohamy (2012) study mentioned above, participants first learned a series of arbitrary associations between stimulus sets A and B, and then learned that some of the stimuli in B led to monetary reward (B+) while others did not (B-). When asked to choose between two A stimuli, participants showed preferences for the A stimuli that had been paired with B+ over the other stimuli that had been paired with B-, though neither stimulus had been directly paired with reward. This decision bias was predicted by greater reactivation of prior related experience (A->B) in the hippocampus during the encoding of new reward information (B->+), suggesting that hippocampal memory representations support the spread of monetary value across related experiences. Other studies show that rewards newly introduced at the time of decision can be combined with state representations to influence choice (Bornstein & Daw, 2013). Taken together, these findings are consistent with the idea that the hippocampus supports adaptive behavior by coding relational representations that connect distinct states (e.g., spatial locations and discrete events).

Stimulus-context latent representation

Although much of the work in memory-guided decisions focuses on how relational representations are constructed during encoding, or "retrospective integration", recent research has begun to understand how individual memories are integrated at the time of decision through retrieval mechanisms, a form of "prospective integration" (Doll et al., 2015; Koster et al., 2018). For example, in one study Doll and colleagues designed a multi-step reward learning task assessing the extent to which participants integrated information about rewards received during other interleaved trials (Doll et al., 2015). Using category-specific images at different decision stages, Doll and colleagues decoded the neural representations that simulate the prospective paths in the hippocampus. The activity patterns were correlated with the degree to which choices reflected successful integration, indicating that the hippocampus supports prospective value computation by supplying information about the sequential relations between actions.

Several key factors that mediate prospective integration have been identified, with context information being the most important one. For example, it has been shown that items are more likely to be retrieved together if they are experienced closer in time (Howard & Kahana, 2002; Sederberg et al., 2008, 2011). The link between stimuli and their *context* is distinguished from links between stimuli within a context in that the context serves as a mediating, *latent*, representation among many events, and represents another scale at which relational associations may be formed -- and, critically, navigated (Shin & DuBrow, 2021). This phenomenon was exemplified by the Temporal Context Model (TCM), which posits that during encoding individual items are bound to a slowly drifting "context vector" in memory. At test, retrieval of an item leads to the reinstatement of the context that the item was bound to, which biases subsequent retrieval towards items that were bound to a similar temporal context as the item that was just retrieved. Several studies have since shown that when individual memories are bound to the (temporal) context in which they are encoded, decisions are influenced by information indirectly related to the present problem through these contextual links (Bornstein & Norman, 2017; Hoskin et al., 2019; Morton et al., 2020).

In sum, experience creates multiple forms of memory representations that variously encode predictive statistics about both observed, stimulus-stimulus associations, as well as inferred links between abstract states. These representations serve a common purpose of allowing humans and animals to more quickly act on regularities in the environment. We next examine the process by which this information is used to enact decisions.

Process: Within-trial dynamics of pattern completion

This section reviews what is known about the ways in which these multiple representations are accessed in the service of behavior; in other words, whereas the previous section examined how representations reflect the dynamics of memory-guided decision-making across experiences, this section illustrates the dynamics of memory-guided decisions within a single choice.

The core idea of memory sampling is that memory retrieval is a form of Monte Carlo estimation, leveraging these representations to estimate possible future states and rewards, given the current state and a candidate action (Eqn. 6). This sort of memory-based simulator has been shown to be useful for effective planning in large, partially observable environments (Silver & Veness, 2010), such as are likely predominant in naturalistic settings. However, it is unknown to what degree these properties correspond to biological organisms. Here, we discuss what is known about the ability of the hippocampal memory system to reinstate past experience on the basis of partial inputs, a process known as *pattern completion* (Marr, 1971).

Pattern completion during episodic recall is known to depend on the hippocampus (Horner et al., 2015). The CA3 region of hippocampus is thought to be instrumental to pattern completion (Guzman et al., 2016; Neunuebel & Knierim, 2014; van Dijk & Fenton, 2018). This area has the multiply-recurrent circuitry and convergent direct external inputs necessary to perform autoassociative computations that can resuscitate stored patterns on the basis of partial input (Koster et al., 2018; Marr, 1971; McNaughton & Morris, 1987; Schapiro et al., 2017). These critical architectonic features may allow CA3 to integrate coincident inputs across both time and sensory modality, supporting a form of fuzzy coincidence detection that can apply to sequences as well as sets (Lisman & Grace, 2005). It is known that pattern completion is ongoing throughout behavior, during awake rest, and even during sleep (Antony et al., 2012). The frequency of pattern completion may be reduced during periods of repeated novel experience (Duncan et al., 2012;

Hasselmo, 2006), or quieted by cholinergic release (Prince et al., 2017) that encourages the formation of new context representations (Gold, 2003).

By definition, pattern completion reinstates many of the same neural ensembles that were co-active during experience, or which have been attached via offline processing. These reinstated patterns can influence processing downstream of the regions where patterns are being reinstated, just as does the original external sensory input (Hoskin et al., 2019). It thus follows that ongoing decision processes should be influenced by this reactivation, suggesting an avenue for goal-directed deployment of this function. Indeed, pattern completion has been shown to be deployed when needed to inform uncertain inference (Hindy et al., 2016). The interaction between internally-generated sequences and the properties of external input is a critical feature of computational work on *state inference*, a necessary function for online planning in environments with uncertain latent contingency structure (Kaelbling et al., 1998; Rao, 2010).

Pattern completion may be especially useful to decision-making because it allows past choices and outcomes to come to mind in situations that are similar to, but not exactly the same as, past encounters. This supports a form of *generalization*, permitting biological agents to navigate new environments or take on new tasks with little previous direct experience (Leutgeb & Leutgeb, 2007). An open question is whether, or in which situations, do completed patterns serve as a rigid template for subsequent action (Lengyel & Dayan, 2008) or something more akin to a proposal for action, to be evaluated in the context of other information available at the time of the current choice (Vikbladh et al., 2017).

Where does the time go?

The dynamics of memory retrieval may play an important role in decisions in biological organisms. If decisions were based on the reactivation of single episodes, they might be expected to execute more or less instantly; unlike sensory decisions, which rely on fundamentally incremental input, memory-guided decisions could in theory have immediate access to the internal representations that serve as evidence. But elongated decision times are not only widely observed, they closely track characteristics of the decision variable (Yang & Shadlen, 2007), and so models that take account of response time can improve the out-of-sample prediction of choices (Clithero, 2018). In an insightful evaluation of this question, Shohamy and Shadlen (2016) propose that one reason memory-guided decisions take time, rather than acting instantly on internally-available information, is because a limited-bandwidth thalamocortical pipeline enforces serial processing. They then assert that retrieval time itself does not play a role in the sequential nature of memory sampling, because sharp-wave ripples (SWR: one, though not the only, putative substrate of memory retrieval; Joo & Frank, 2018), operate in short, high-frequency bursts, much faster than the variability observed in decision times, and so, they argue, couldn't possibly be a rate-limiting factor in decision-making.

However, several features of memory reactivation (encompassing both SWRs and also theta sequences, which are lower frequency and more regular) suggest that retrieval dynamics may play a part in the availability of information. First, though SWRs do indeed unfold over very short timescales, their onset time is highly irregular (Buzsáki & Tingley, 2018), perhaps reflecting other rate-limiting processes that precede any decision-relevant SWR events (e.g. memory search). Memory search has often been fruitfully modeled as a biased random walk along a graph constructed from experience (Collins & Quillian, 1969; Jun et al., 2015). Distinct — even conflicting — action tendency signals may be generated at different steps along the walk. Supporting the idea that memory retrievals' influence on decision unfolds over time is the observation that longer delays before choice lead to greater memory influence on decisions (Foerde & Shohamy, 2011) -- and, in particular, greater influence of extended retrievals from memory (Bakkour et al., 2019; Eldar et al., 2020; Gordon et al., 2014). Second, the behaviorally-relevant features of SWRs are highly variable, both across instances and the population of cells participating, and depend on contextual factors such as cognitive states and vigilance, consistent with the idea that these events provide information in service of current behavioral and cognitive demands (Hussin et al., 2020). As a result, there may not be a simple relationship between individual ripple events and subsequent decisions. Third, the content of memory retrieval that serves as the "common currency" relevant to decisions — whether value representations or action tendencies — is likely not encoded directly in hippocampus, but instead by populations one or more synaptic connections downstream. Suprathreshold activation of these representations may require converging input or preceding innervation from other areas, such as vmPFC (Gluth et al., 2015; Schmidt et al., 2019; Spalding et al., 2018; Weilbächer & Gluth, 2016), or be mediated by intermediate abstract representations, for instance in retrosplenial (Chrastil et al., 2015; Mao et al., 2017, 2018) or inferior temporal cortex (Bornstein & Norman, 2017; Hoskin et al., 2019; Mack & Preston, 2016). Fourth, the influence of value from past decisions may depend on a more elaborative retrieval ("source"; Murty et al., 2015), which computational models posit requires additional activation that may stretch across multiple cycles of hippocampal retrieval (Kerrén et al., 2018). These elaborated representations may develop relatively slowly during retrieval in part because they depend, especially early on in experience, on "big-loop" recurrence, multisynaptic bridges between medial temporal lobe structures and other areas of cortex (Koster et al., 2018; Kumaran & McClelland, 2012). Finally, a recent study examined serial decisions that were initiated by a single composite stimulus, and found that sensory evidence is accumulated in parallel before an integration bottleneck occurs somewhere downstream; evidence that applies to later decisions is "buffered", apparently losslessly (Kang et al., 2020). This finding supports the idea that the time it takes to act on information retrieved from memory can vary greatly across decisions, and that this information can be sampled near-simultaneously from multiple sources. This last point is relevant because we don't fundamentally know how many compound decisions are contained within a single experimental trial response in standard lab tasks -- this is likely at least as true in rodents, in whom most work on these neural substrates has been performed, as it is in humans (for instance, a rodent's decision to enter an arm of a maze may be preceded by several intermediate decisions e.g. to change head direction or to serially *not* enter other arms of the maze). Some of these decisions may not be deliberated for enough time to depend on memory retrieval, especially after extensive practice on the task, as is common in rodent experiments. Additional work is necessary to understand what is the *effective* time required to transmit decision-relevant information from memory retrievals downstream, and how it depends on attributes of the current decision problem.

Such investigations will need to pay special attention to retrieval during early learning, which may be dramatically different in dynamics and content from the kind of online reactivation that occurs after many experiences with a task or learning set (Redish, 2016), and especially when divorced from spatial navigation, the pace of which can confound investigations of the frequency of retrieval of related place field representations. Along these lines, one important recent study examined these dynamics in a non-spatial setting, examining “lookahead” during sequences of odors in well-trained rodents (Shahbaba et al., 2019). Using a novel combination of decoding methods to identify odor identity representations in dorsal CA1, the authors found that they were able to decode anticipatory sequence reactivations on the scale of a few 100s of milliseconds, consistent with the theta-band rhythms observed in spatial navigation studies. Critically, however, they also observed faster sequence reactivations *within* an individual theta cycle, with power that varied with distance from the current odor, suggestive of either simultaneous reactivation at multiple temporal scales or an underlying substrate for the sequences decoded at lower frequencies. Further investigation is necessary to understand whether sub-theta sequence reactivation is alongside, or constituent of, the more well-known theta sequences.

More broadly, however, the dynamics of pattern completion are still poorly understood (Knierim & Neunuebel, 2016). The decoded content of these sequences can shift categorically between individual periods of the theta cycle. This shifting may reflect reactivation on the basis of uncertain sensory or latent inputs, but “flickering” or “fast remapping” has been observed even in the case of spatial representations, in which it is difficult to induce fundamental uncertainty (Jezek et al., 2011). A separate line of research has identified “chunking” of theta sequences; these imply that only partial trajectories may be reactivated in a single theta cycle. Elongated trajectories may therefore take multiple theta cycles to reactivate (Gupta et al., 2012; Tang et al., 2020). Consistent with this idea, and supporting the proposal that these sequences drive behavior, rather than reflect it, disrupting mPFC during deliberation impairs both lookahead theta sequences and associated “vicarious trial and error” behavior (Schmidt et al., 2019).

Finally, though the decoding approach to investigating properties of reactivated place cell sequences has revealed profoundly important structure, trajectory dynamics are not necessarily ballistic. It has recently been observed that population-wide activity, much of which is likely obscured by modal decoding, more closely matches Brownian diffusion along a gradient (Stella et al., 2019). This is consistent with the idea that each reactivated trajectory provides only partial information about the overall content of lookahead, necessitating integration across multiple reactivations, and suggests that behavior may be sensitive to dynamics obscured by extant decoding approaches. Intriguingly, the same study showed that behavior is “superdiffusive”, reflecting occasional “jumps” in diffusion, as would result from Brownian motion convolved with stochastic perturbations in the direction of the gradient. Such jumps may have adaptive value in navigating ecologically normative environments (Viswanathan et al., 2011), but the ultimate source of their neural instantiation remains unclear.

Taken together, the above findings support the idea that multiple memory representations are created during experience, that each is tuned towards different aspects of experience, including history-dependence, and that the dynamics of reactivation are variable and linked to the associative structure of memories and memory sequences. The next section synthesizes these representation-dependent properties of memory reactivation with the accumulation framework and reinforcement learning problem described above.

V. Random walks together

In the previous sections we reviewed evidence that experience produces multiple associative representations (sequences) that vary in the length of history they incorporate, the dimensions or features of experience that they represent (e.g. motor sequences, sensory features, latent states), the scale at which their constituent parts are recorded (coarse to fine), and the degree of determinism in their connection (high or low entropy). Each of these representations has, separately, been empirically shown to be reactivated in response to internal or external stimulus - and, when reinstated, to serve as predictions of future outcomes that guide ongoing action selection.

This proliferation of predictions presents its own puzzle: Which one should be used to guide behavior in any given situation? In other words: How do we decide how to decide? A seminal proposal in this area is that each representation constitutes a “controller”, whose predictions are arbitrated among on the basis of their uncertainty (Daw et al., 2005; Keramati et al., 2011; Simon & Daw, 2011). This principle, originally proposed to explain the apparent trade off between pairs of flexible and inflexible representations (e.g. as encoded in dorsomedial and dorsolateral striatal circuits (Yin et al., 2004, 2005), has been extended to encompass episodic memory as well (Lengyel & Dayan, 2008), with each system predominant after different degrees of experience in a given environment. However, it is as yet unclear how this principle is instantiated in neural circuits. One candidate, that representations “compete” for modal control (Poldrack et al., 2001), is a reasonable explanation of data in tasks with stationary probabilistic structure, but seems not to anticipate the ongoing contribution of multiple systems that is observed when examining non-stationary tasks (Bornstein & Daw, 2012). Related work explores the idea that top-down or other control mechanisms guide this process (Lee et al., 2014), however it is unclear exactly how these signals propagate across such a multitude of representations.

Our review of the relationship between the representational properties listed above and the dynamics of reactivation, viewed through the framework of sequential sampling, points to a potential unifying mechanism that is consistent with each of these proposals, without requiring top-down arbitration. Specifically, if we write out the log odds summation from Equation 8 with multiple sources of evidence, such as arriving from multiple internal memory representations (Fig. 3 and Eqn. 9a - here, c for context and i for item), each arriving at different *latencies* (time to arrival of first sample) and continuing at different *frequencies* (rate at which subsequent samples arrive), we see that the resulting mixture of evidence implements a time-varying weighting across the different source representations (Eqn. 9b).

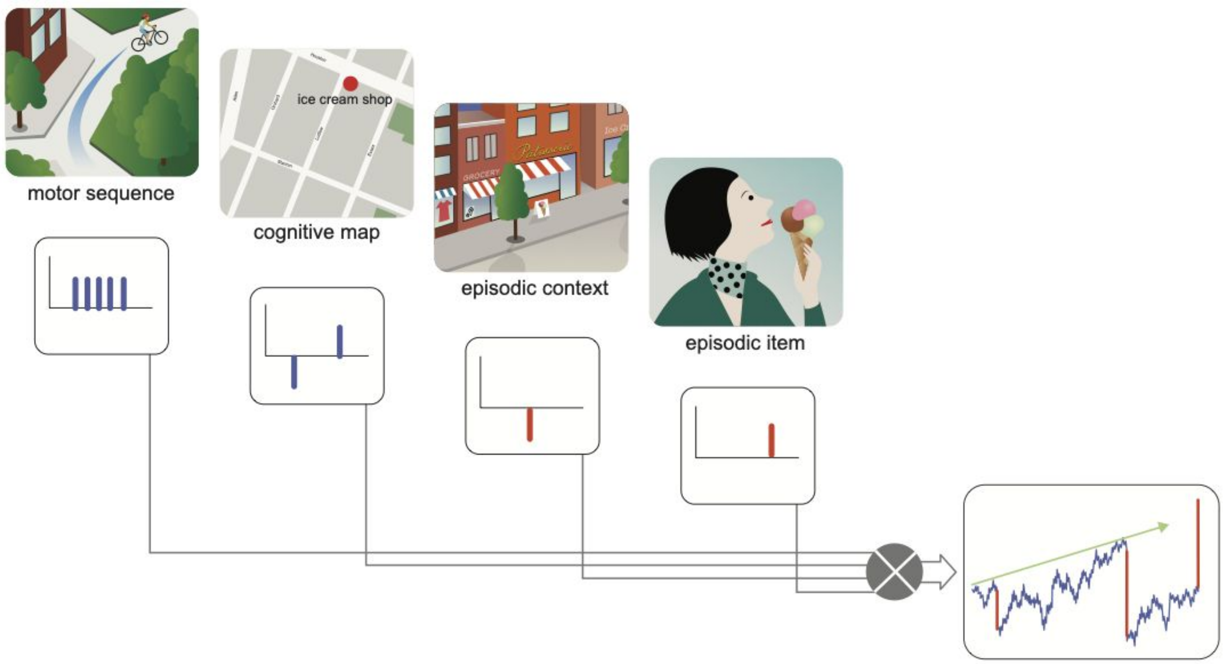


Figure 3. Simultaneous sampling from multiple internal representations implements a "product of experts" via a jump-diffusion process. In this example, a person may draw on multiple forms of internal representation when deciding which ice cream shop to visit. For instance, she may have a well-traveled route from her apartment to an often-visited shop (motor sequence), while also drawing on an allocentric representation of the location of each shop (cognitive map). These can be combined with memories of her more recent experiences with different shops, including the day and surroundings of a previous experience (episodic context) as well as a particular individual experience (episodic item). Each of these is sampled at different latencies, and with different frequencies, and their product results in a "jump-diffusion" timeseries of accumulated evidence. The resulting decision - which boundary is crossed, and at what time - is thus a weighted mixture of the contributing factors.

$$\log LR_{1,2} \approx \log\frac{P(e_{c,1}|s_1)}{P(e_{c,1}|s_2)} + \log\frac{P(e_{c,2}|s_1)}{P(e_{c,2}|s_2)} + \log\frac{P(e_{i,1}|s_1)}{P(e_{i,1}|s_2)} + \log\frac{P(e_{c,3}|s_1)}{P(e_{c,3}|s_2)} + \log\frac{P(e_{c,4}|s_1)}{P(e_{c,4}|s_2)} + \log\frac{P(e_{i,2}|s_1)}{P(e_{i,2}|s_2)} + \ldots \quad (9a)$$

$$\approx \sum_{j=1}^{N} \log\frac{P(e_{c,j}|s_1)}{P(e_{c,j}|s_2)} + \sum_{k=1}^{\frac{N}{2}} \log\frac{P(e_{i,k}|s_1)}{P(e_{i,k}|s_2)} \quad (9b)$$

Note that the form of the weighting may not be monotonic in time, as different representations may take longer to generate their first sample (e.g. memory sequences), or may appear to "pause" in generating samples (e.g. at boundaries identified between adjacent memories whose reward statistics differ — and which thus imply distinct action tendencies; Rouhani et al., 2020). The resulting continuous-time form would be that of the "jump-diffusion" model previously discussed.

No matter the form that the sample arrival dynamics take, the instantaneous weighting implied by Equation 9 implements an organizing principle akin to the "value of information" (Bera et al., 2020; Callaway et al., 2018) in which representations with less-precise predictions or less-immediately available evidence are slower to influence choice, which can allow information that tends to be more precise or immediate to dominate the accumulated evidence calculation. Critically, though this time-varying weighting requires no "top-down" or other bias signal, it can naturally incorporate them. For instance, eye gaze has been shown to modulate the accumulation rate of the attended option in simple choice tasks (Krajbich & Rangel, 2011); in the current framework that modulation may be implemented by the arrival of stimulus-triggered evidence samples from memory (Constantino & Daw, 2010), or by a gain modulation of signals arriving from ongoing reactivations (Aston-Jones & Cohen, 2005).

Whether or not additional signals enter into the calculation, a relationship between the informational characteristics of the representation and its sample dynamics in the form of Equation 9 is also equivalent to a suite of tools from machine learning for online mixing of classifiers with varying "expertise" (reliability) across data domains, known as "product of experts" (Hinton, 1999), one instance of "ensemble learning" (Polikar, 2012). One approach involves multiplying the action tendencies (summing the log likelihoods) produced by each component — exactly the procedure given by the series above. While the field currently lacks analytical results on general optimality guarantees for this method, simulations support its efficacy in navigating partially observable environments ("Boltzmann Multiplication"; Wiering & van Hasselt, 2008). More sophisticated "ensemble fusion" approaches learn adaptive weighting for each component — predictive Hebbian learning mechanisms may be sufficient to develop these with use by altering sequence-specific dynamics (see *Future Directions*, below). Further research is necessary to understand how learning is tuned to support adaptive fusion.

This computational approach could guide further research in the neurobiology of the differential dynamics of memory reactivation across representations. One question raised by this framework is whether the temporal dynamics of memory reactivation are fundamental, adapt to the time available, or are modulated by the content of representation or computations being performed. Intrinsic differences in reactivation dynamics for different representations could be one form of rational "inductive bias" (Griffiths et al., 2010) for fast, flexible decision-making using multiple sources of evidence - memory, sensory, motor - allowing decision weights to adaptively adjust to the expected temporal trajectory of the current decision, conditional on it not having yet completed — e.g. fast motor sequences should guide short decisions, but memory sequences may play a more dominant role if the action remains unresolved[7]. Several recent empirical observations are consistent with this proposal (Hardwick et al., 2019; McDougle & Taylor, 2019); further research is needed to understand how the time-varying mixture of learned representations in memory retrieval reflects its adaptive use in decisions.

[7] Importantly, this is not to say that memory reactivation *only* affects decisions that are not fully resolved by motor sequences. Empirical findings support the idea of continuous flow of information to the effectors, that "late-arriving" evidence samples can play a decisive role not only in choice, but can even *change* decisions for which motor execution has already begun (Resulaj et al., 2009). The same principle may explain how sequential samples implement the discount factors in the unrolled value computation of Equation 5: the discount factor here describes the *average* influence of later evidence samples *across* choices, which have a monotonically increasing probability of terminating before the arrival of the *n*th sample — they are unlikely to affect decisions in the aggregate, but have profound influence when reactivated. This suggestion is consistent with observations that memory accessibility, including as modified by pre-trial "cues", can affect temporal discount rates (Gabaix & Laibson, 2017; Palombo et al., 2015; Peters & Büchel, 2010; Weber et al., 2007), and parallels the way in which memory cues can overcome effective "discounting" of probabilistic transitions in sequential decisions (Bornstein et al., 2017; Vikbladh et al., 2017). Further work is needed to understand how within-trial dynamics affect the integration of information about potential future states.

Consistent with the proposal that sample rate tracks the history of experience embedded in the sample, evidence supports the idea that semantic memories are accessed at a faster rate than are episodes, following classical spreading activation theories of neural processing (Collins & Loftus, 1975; Corbett & Wickelgren, 1978). Supporting the idea that such information is accessed simultaneously, despite different delays to peak efficacy, responses are further speeded when semantic information is congruent with episodic (McKoon et al., 1985); conversely, the availability of congruent semantic information influences the content of ongoing episodic retrieval (Manning et al., 2012). Taken together, neurobiological dynamics, process-rational cognitive models, and dynamical systems considerations support the notion that memory-inflected evidence accumulation is both continuous and irregular.

We have seen that multiple memory representations are learned and transformed on an ongoing basis, reflecting experience integrated across multiple scales, and that these representations are accessed by a pattern completion process whose effective dynamics depend on neural circuit properties and coherence of the representations in question. Taken together, it follows that choices under time pressure will be biased towards options for which this combination of factors results in a faster sample onset and lower latency between successive samples, and that response times will be shaped by the difference between options on these factors (in addition to, for instance, desirability (Fine et al., 2020). In other words, the influence of associative distance on decisions should be mediated via its influence on evidence dynamics. Further investigation is necessary to understand how the temporal dynamics of associative memory retrieval dictate the type of information that guides decisions.

Future directions

A primary direction of future research is understanding how various factors influence the temporal dynamics of memory retrieval. Evidence suggests the influence of at least the following terms: 1. semantic distance (e.g. as estimated using word embeddings; Chadwick et al., 2016), 2. episodic distance (Polyn et al., 2009), and 3. the spread of probability mass across associations at each kind of distance (Socher et al., 2009). Dimov and Link (2017) examined how decisions were made on the basis of cues that varied in each of these factors (operationalized as retrieval fluency and cue validity). They found that, for most participants, retrieval fluency was predominant over cue validity. However, the range of inferred cue validities in the experiment was narrow, which may have limited its usefulness in decisions. Importantly, they observed that subjects' response times varied strongly with the number of cues retrieved for each decision, regardless of what was the dominant factor (fluency or validity) for that subject. The proposal that multiple forms of decisions depend on retrieval dynamics that vary as a function of associative distance may explain why choices and response times appear to covary between tasks that examine how subjects weigh options across many kinds of such distances, for instance in intertemporal choice, patch foraging, and model-based planning (Kane et al., 2019; Shenhav et al., 2014), each of which have been independently shown to depend on long-term memory representations (Palombo et al., 2015; Peters & Büchel, 2010; Schmidt et al., 2019; Vikbladh et al., 2019).

Finally, though we have focused here on memory sampling's involvement in two-alternative forced choice, the mechanism we describe has been observed or shown to be useful in a wide array of functions. Specifically, some form of time-dependent successive sampling from rich, autobiographical memories with episodic features has been proposed in the following domains: as a mechanism for equilibrium strategy discovery in repeated multiplayer economic games (Gonçalves, 2020); to augment the learning trajectories of artificial agents via a form of 'memoization' of partial inferences about environmental contingencies (Ritter et al., 2018); to explain the trajectory of symptom development in anxiety disorders, via biased sampling of threatening stimuli (Sharp et al., 2020); to explain the decision to use substances of abuse after years of abstinence (Bornstein & Pickard, 2020); and to support working memory maintenance (Hoskin et al., 2019). This ubiquity of functional impacts aligns with observations of widespread hippocampal involvement in cognition and perception (Shohamy & Turk-Browne, 2013), and more broadly concords with the centrality of this form of memory in everyday experience (Bergson, 1913). Much work remains to understand how these persistent records of past experience -- and their near-constant reactivation -- shape our thoughts and actions.

Box 1: Open questions

- To what extent does memory sampling require conscious awareness of recollection, at the time of decision, or even explicit recall of the same memoranda, as measured at a later time?

- What are the neural substrates of memory samples? Is it the case that sharp-wave ripples (SWRs) indicate "offline" samples, and theta sequences support decision-time sampling, or is there a more complex interplay?

- What factors - at encoding, retrieval, and during intervening memory transformations - determine how samples are prioritized during decision-making?

- Is memory organized in such a way as to match the retrieval time of information to its use in deliberative decisions? For instance, are more temporally or associatively remote memories more slowly sampled?

Acknowledgements
The authors wish to thank Julia Kuhl (https://somedonkey.work/) for the illustrations in Figures 1 and 3.

Data Availability Statement
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Funding Information

The authors gratefully acknowledge support from a NARSAD Young Investigator Award by the Brain and Behavior Research Foundation (to AMB). This publication is based in part upon work supported by the Khalifa University of Science and Technology under Award No. CIRA-2019-050 (to SFF).

References

Akrami, A., Kopec, C. D., Diamond, M. E., & Brody, C. D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, *554*(7692), 368–372.

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl 2*, 10379–10386.

Antony, J. W., Gobel, E. W., O'Hare, J. K., Reber, P. J., & Paller, K. A. (2012). Cued memory reactivation during sleep influences skill learning. In *Nature Neuroscience* (Vol. 15, Issue 8, pp. 1114–1116). https://doi.org/10.1038/nn.3152

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, *543*(7647), 719–722.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.

Bakkour, A., Palombo, D. J., Zylberberg, A., Kang, Y. H., Reid, A., Verfaellie, M., Shadlen, M. N., & Shohamy, D. (2019). The hippocampus supports deliberation during value-based decisions. *eLife*, *8*. https://doi.org/10.7554/eLife.46080

Balleine, B. W. (2007). The Neural Basis of Choice and Decision Making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(31), 8159–8160.

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

Barto, A. G., Houk, J. C., Davis, J. L., & Beiser, D. G. (1995). *Models of information processing in the basal ganglia*.

Baum, C. W., & Veeravalli, V. V. (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory / Professional Technical Group on Information Theory*, *40*(6), 1994–2007.

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. In *Nature Neuroscience* (Vol. 10, Issue 9, pp. 1214–1221). https://doi.org/10.1038/nn1954

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415). https://doi.org/10.1126/science.aat6766

Bera, K., Mandilwar, Y., & Raju, B. (2020). Value-of-Information based Arbitration between Model-based and Model-free Control. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.*

Bergson, H. (1913). *Matter and memory*. https://doi.org/10.1037/13803-000

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology*, *53*(3), 155–167.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-

choice tasks. *Psychological Review*, *113*(4), 700–765.
Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Memory, attention, and choice. *The Quarterly Journal of Economics*, *135*(3), 1399–1442.
Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *The European Journal of Neuroscience*, *35*(7), 1011–1023.
Bornstein, A. M., & Daw, N. D. (2013). Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS Computational Biology*, *9*(12), e1003387.
Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*, 15958.
Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*(7), 997–1003.
Bornstein, A. M., & Pickard, H. (2020). “Chasing the first high”: memory sampling in drug choice. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *45*(6), 907–915.
Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition*, *113*(3), 262–280.
Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator. In *Current Opinion in Neurobiology* (Vol. 37, pp. 149–157). https://doi.org/10.1016/j.conb.2016.01.003
Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z.-X., Grady, C., Rosenbaum, R. S., Winocur, G., Barense, M. D., & Moscovitch, M. (2018). Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Current Biology: CB*, *28*(13), 2129–2135.e6.
Busemeyer, J. R., & Diederich, A. (2010). *Cognitive Modeling*. SAGE.
Buzsáki, G., & Tingley, D. (2018). Space and Time: The Hippocampus as a Sequence Generator. *Trends in Cognitive Sciences*, *22*(10), 853–869.
Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., & Griffiths, T. (2018). A resource-rational analysis of human planning. *Proceedings of the 40th annual conference of the Cognitive Science Society.*
Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, *113*(36), 10180–10185.
Chrastil, E. R., Sherrill, K. R., Hasselmo, M. E., & Stern, C. E. (2015). There and Back Again: Hippocampus and Retrosplenial Cortex Track Homing Distance during Human Path Integration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(46), 15442–15452.
Clithero, J. A. (2018). Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization*, *148*, 344–375.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.
Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.
Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*(11), 1562–1564.
Constantinescu, A. O., O’Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468.
Constantino, S. M., & Daw, N. D. (2010). A closer look at choice. In *Nature Neuroscience* (Vol. 13, Issue 10, pp. 1153–1154). https://doi.org/10.1038/nn1010-1153
Corbett, A. T., & Wickelgren, W. A. (1978). Semantic memory retrieval: analysis by speed accuracy tradeoff functions. *The Quarterly Journal of Experimental Psychology*, *30*(1), 1–

15.
Corrado, G. S., Sugrue, L. P., Seung, H. S., & Newsome, W. T. (2005). Linear-nonlinear-Poisson models of primate choice dynamics. *Journal of the Experimental Analysis of Behavior*, *84*(3), 581–617.
Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300.
Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.
Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711.
Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.
Daw, N. D., & Others. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, *23*(1).
Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, *5*(4), 613–624.
Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3 Suppl*, 1218–1223.
Dimov, C. M., & Link, D. (2017). Do people order cues by retrieval fluency when making probabilistic inferences? *Journal of Behavioral Decision Making*, *30*(4), 843–854.
Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. In *Neuron* (Vol. 80, Issue 2, pp. 312–325). https://doi.org/10.1016/j.neuron.2013.09.007
Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*(5), 767–772.
Donaldson, W., & Tulving, E. (1972). *Organization of memory*. Academic Press.
DuBrow, S., Eberts, E. A., & Murty, V. P. (2019). A common mechanism underlying choice's influence on preference and memory. *Psychonomic Bulletin & Review*, *26*(6), 1958–1966.
Duncan, K. D., & Shohamy, D. (2016). Memory states influence value-based decisions. *Journal of Experimental Psychology. General*, *145*(11), 1420–1426.
Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's penumbra: episodic memory decisions induce lingering mnemonic biases. *Science*, *337*(6093), 485–487.
Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, *83*(4), 764–770.
Eldar, E., Lièvre, G., Dayan, P., & Dolan, R. J. (2020). The roles of online and offline replay in planning. *eLife*, *2020*(9), e56911.
Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife*, *4*. https://doi.org/10.7554/eLife.05457
Fine, J. M., Zarr, N., & Brown, J. W. (2020). Computational Neural Mechanisms of Goal-Directed Planning and Problem Solving. *Computational Brain & Behavior*, *3*(4), 472–493.
Foerde, K., & Shohamy, D. (2011). Feedback Timing Modulates Brain Systems for Learning in Humans. In *Journal of Neuroscience* (Vol. 31, Issue 37, pp. 13157–13167). https://doi.org/10.1523/jneurosci.2701-11.2011
Folkerts, S., Rutishauser, U., & Howard, M. W. (2018). Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(17), 4200–4211.
Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*(5), 458–462.
Gabaix, X., & Laibson, D. (2017). *Myopia and Discounting* (No. 23254). National Bureau of Economic Research. https://doi.org/10.3386/w23254

Garvert, M. M., Dolan, R. J., & Behrens, T. E. J. (2017). A map of abstract relational knowledge in the human hippocampal--entorhinal cortex. *eLife*, *6*, e17086.
Geerts, J. P., Stachenfeld, K. L., & Burgess, N. (2019). Probabilistic Successor Representations with Kalman Temporal Differences. In Proceedings of the 2019 Conference on Cognitive Computational Neuroscience; doi:10.32470/CCN.2019.1323-0
Gershman, S. J. (2018). The Successor Representation: Its Computational Logic and Neural Substrates. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(33), 7193–7200.
Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197–209.
Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*, 101–128.
Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, *5*, 43–50.
Gläscher, J., & Büchel, C. (2005). Formal learning theory dissociates brain regions with different temporal integration. *Neuron*, *47*(2), 295–306.
Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.
Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. In *Proceedings of the National Academy of Sciences* (Vol. 108, Issue Supplement_3, pp. 15647–15654). https://doi.org/10.1073/pnas.1014269108
Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective Connectivity between Hippocampus and Ventromedial Prefrontal Cortex Controls Preferential Choices from Memory. *Neuron*, *86*(4), 1078–1090.
Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*(1), 10–16.
Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.
Gold, P. E. (2003). Acetylcholine modulation of neural systems involved in learning and memory. *Neurobiology of Learning and Memory*, *80*(3), 194–210.
Gonçalves, D. (2020). *Sequential sampling and equilibrium*. Working Paper. https://duartegoncalves.com/papers/Goncalves%202020%2C%20Sequential%20Sampling%20and%20Equilibrium.pdf
Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cerebral Cortex* , *24*(12), 3350–3364.
Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364.
Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & David Redish, A. (2010). Hippocampal Replay Is Not a Simple Function of Experience. In *Neuron* (Vol. 65, Issue 5, pp. 695–705). https://doi.org/10.1016/j.neuron.2010.01.034
Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2012). Segmentation of spatial experience by hippocampal θ sequences. *Nature Neuroscience*, *15*(7), 1032–1039.
Guzman, S. J., Schlögl, A., Frotscher, M., & Jonas, P. (2016). Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science*, *353*(6304), 1117–1123.
Hamm, A. G., & Mattfeld, A. T. (2019). Distinct neural circuits underlie prospective and concurrent memory-guided behavior. Cell reports, 28(10), 2541-2553.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546), 220–223.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed Decision Time Affects the Weighting of Prior Probability in a Perceptual Decision Task. In *Journal of Neuroscience* (Vol. 31, Issue 17, pp. 6339–6352). https://doi.org/10.1523/jneurosci.5613-10.2011

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. In *Nature Human Behaviour* (Vol. 3, Issue 12, pp. 1252–1262). https://doi.org/10.1038/s41562-019-0725-0

Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks: The Official Journal of the International Neural Network Society*, *19*(5), 535–546.

Hasselmo, M. E. (2006). The role of acetylcholine in learning and memory. *Current Opinion in Neurobiology*, *16*(6), 710–715.

Hayden, B., & Niv, Y. (2020). *The case against economic values in the brain*. https://doi.org/10.31234/osf.io/7hgup

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. In *Nature Neuroscience* (Vol. 19, Issue 5, pp. 665–667). https://doi.org/10.1038/nn.4284

Hinton, G. E. (1999). *Products of experts*. 1–6.

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*, 7462.

Hoskin, A. N., Bornstein, A. M., Norman, K. A., & Cohen, J. D. (2019). Refresh my memory: Episodic memory reinstatements intrude on working memory maintenance. *Cognitive, Affective & Behavioral Neuroscience*, *19*(2), 338–354.

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Hussin, A. T., Leonard, T. K., & Hoffman, K. L. (2020). Sharp-wave ripple features in macaques depend on behavioral state and cell-type specific firing. *Hippocampus*, *30*(1), 50–59.

Iigaya, K., Ahmadian, Y., Sugrue, L. P., Corrado, G. S., Loewenstein, Y., Newsome, W. T., & Fusi, S. (2019). Deviation from the matching law reflects an optimal strategy involving learning over multiple timescales. *Nature Communications*, *10*(1), 1466.

Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, *3*(7), 719–732.

Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M.-B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, *478*(7368), 246–249.

Jiang, N., Kulesza, A., Singh, S., & Lewis, R. (2015). The Dependence of Effective Planning Horizon on Model Accuracy. *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 1181–1189.

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(45), 12176–12189.

Joo, H. R., & Frank, L. M. (2018). The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. In *Nature Reviews Neuroscience* (Vol. 19, Issue 12, pp. 744–757). https://doi.org/10.1038/s41583-018-0077-1

Jun, K.-S., Zhu, J., Rogers, T. T., Yang, Z., & Yuan, M. (2015). Human Memory Search as Initial-Visit Emitting Random Walk. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R.

Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28, pp. 1072–1080). Curran Associates, Inc.
Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1), 99–134.
Kane, G. A., Bornstein, A. M., Shenhav, A., Wilson, R. C., Daw, N. D., & Cohen, J. D. (2019). Rats exhibit similar biases in foraging and intertemporal choice tasks. *eLife*, *8*. https://doi.org/10.7554/eLife.48429
Kang, Y. H. R., Löffler, A., Jeurissen, D., Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2020). Multiple decisions about one object involve parallel sensory acquisition but time-multiplexed evidence incorporation. In *Cold Spring Harbor Laboratory* (p. 2020.10.15.341008). https://doi.org/10.1101/2020.10.15.341008
Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology*, *7*(5), e1002055.
Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(45), 12868–12873.
Kerrén, C., Linde-Domingo, J., Hanslmayr, S., & Wimber, M. (2018). An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Current Biology: CB*, *28*(21), 3383–3392.e6.
Keung, W., Hagen, T. A., & Wilson, R. C. (2020). A divisive model of evidence accumulation explains uneven weighting of evidence over time. *Nature Communications*, *11*(1), 2160.
Knierim, J. J., & Neunuebel, J. P. (2016). Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of Learning and Memory*, *129*, 38–49.
Koechlin, E. (2019). Human Decision-Making beyond the Rational Decision Theory. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.11.001
Kok, P., & Turk-Browne, N. B. (2018). Associative Prediction of Visual Shape in the Hippocampus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(31), 6888–6899.
Koster, R., Chadwick, M. J., Chen, Y., Berron, D., Banino, A., Düzel, E., Hassabis, D., & Kumaran, D. (2018). Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes. *Neuron*, *99*(6), 1342–1354.e6.
Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857.
Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, *20*(7), 512–534.
Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological Review*, *119*(3), 573–616.
Kuwabara, K. J., & Pillemer, D. B. (2010). Memories of past episodes shape current intentions and decisions. In *Memory* (Vol. 18, Issue 4, pp. 365–374). https://doi.org/10.1080/09658211003670857
Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.
Laming, D. R. J. (1968). *Information theory of choice-reaction times*. https://psycnet.apa.org/record/1968-35026-000?utm_source=sootoday.com&utm_campaign=sootoday.com&utm_medium=referral
Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, *81*(3), 687–699.

Lengyel, M., & Dayan, P. (2008). Hippocampal Contributions to Control: The Third Way. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), *Advances in Neural Information Processing Systems 20* (pp. 889–896). Curran Associates, Inc.
Leutgeb, S., & Leutgeb, J. K. (2007). Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. *Learning & Memory* , *14*(11), 745–757.
Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. In *Psychological Review* (Vol. 125, Issue 1, pp. 1–32). https://doi.org/10.1037/rev0000074
Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: a coherence hypothesis. *Cognitive Psychology*, *40*(2), 87–137.
Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, *46*(5), 703–713.
Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2015). Priming memories of past wins induces risk seeking. *Journal of Experimental Psychology. General*, *144*(1), 24–29.
MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal “Time Cells” Bridge the Gap in Memory for Discontiguous Events. *Neuron*, *71*(4), 737–749.
Mack, M. L., & Preston, A. R. (2016). Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. In *NeuroImage* (Vol. 127, pp. 144–157). https://doi.org/10.1016/j.neuroimage.2015.12.015
Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, *21*(3), 629–636.
Madan, C. R., Spetch, M. L., & Ludvig, E. A. (2015). Rapid makes risky: Time pressure increases risk seeking in decisions from experience. *Journal of Cognitive Psychology* , *27*(8), 921–928.
Madarasz, T., & Behrens, T. (2019). Better Transfer Learning with Inferred Successor Maps. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 9026–9037). Curran Associates, Inc.
Mainen, Z. F., & Pouget, A. (2019). Learning optimal decisions with confidence. *Proceedings of the*. https://www.pnas.org/content/early/2019/11/14/1906787116.short?casa_token=VRUPvlIQpIgAAAAA:QnJ3qDUPI_1BmQUP9FYUOuH73IDDsib2_eyo1gYxtKcXJMu70QWwYbLZYeD-0ghks2KiKmqsMo1eRdg
Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., & Kahana, M. J. (2012). Spontaneously Reactivated Patterns in Frontal and Temporal Lobe Predict Semantic Clustering during Memory Search. In *Journal of Neuroscience* (Vol. 32, Issue 26, pp. 8871–8878). https://doi.org/10.1523/jneurosci.5321-11.2012
Mao, D., Kandler, S., McNaughton, B. L., & Bonin, V. (2017). Sparse orthogonal population representation of spatial context in the retrosplenial cortex. *Nature Communications*, *8*(1), 243.
Mao, D., Neumann, A. R., Sun, J., Bonin, V., Mohajerani, M. H., & McNaughton, B. L. (2018). Hippocampus-dependent emergence of spatial sequence coding in retrosplenial cortex. *Proceedings of the National Academy of Sciences*, *115*(31), 8015–8018.
Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *262*(841), 23–81.
Mayes, A. R., Montaldi, D., Roper, A., Migo, E. M., Gholipour, T., & Kafkas, A. (2019). Amount, not strength of recollection, drives hippocampal activity: A problem for apparent word familiarity-related hippocampal activation. *Hippocampus*, *29*(1), 46–59.
McDougle, S. D., & Taylor, J. A. (2019). Dissociable cognitive strategies for sensorimotor learning. *Nature Communications*, *10*(1), 40.

McKoon, G., Ratcliff, R., & Dell, G. S. (1985). The role of semantic information in episodic retrieval. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 11, Issue 4, pp. 742–751). https://doi.org/10.1037/0278-7393.11.1-4.742

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. In *Trends in Neurosciences* (Vol. 10, Issue 10, pp. 408–415). https://doi.org/10.1016/0166-2236(87)90011-7

Mechera-Ostrovsky, T., & Gluth, S. (2018). Memory Beliefs Drive the Memory Bias on Value-based Decisions. *Scientific Reports*, *8*(1), 10592.

Meder, D., Kolling, N., Verhagen, L., Wittmann, M. K., Scholl, J., Madsen, K. H., Hulme, O. J., Behrens, T. E. J., & Rushworth, M. F. S. (2017). Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. *Nature Communications*, *8*(1), 1942.

Miller, K. J., Ludvig, E. A., Pezzulo, G., & Shenhav, A. (2018). Chapter 18 - Realigning Models of Habitual and Goal-Directed Decision-Making. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making* (pp. 407–428). Academic Press.

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological Review*, *126*(2), 292–311.

Morton, N. W., Schlichting, M. L., & Preston, A. R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proceedings of the National Academy of Sciences*, *117*(47), 29338–29345.

Morton, N. W., Sherrill, K. R., & Preston, A. R. (2017). Memory integration constructs maps of space, time, and concepts. *Current Opinion in Behavioral Sciences*, *17*, 161–168.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., & Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, *17*(12), 1661–1663.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(16), 6255–6264.

Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. In *Journal of Experimental Psychology: General* (Vol. 145, Issue 5, pp. 548–558). https://doi.org/10.1037/xge0000158

Neunuebel, J. P., & Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*, *81*(2), 416–427.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(21), 8145–8157.

O'keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

Onoda, K., Okamoto, Y., Kunisato, Y., Aoyama, S., Shishida, K., Okada, G., Tanaka, S. C., Schweighofer, N., Yamaguchi, S., Doya, K., & Yamawaki, S. (2011). Inter-individual discount factor differences in reward prediction are topographically associated with caudate activation. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *212*(4), 593–601.

Palombo, D. J., Keane, M. M., & Verfaellie, M. (2015). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking. In *Hippocampus* (Vol. 25, Issue 3, pp. 345–353). https://doi.org/10.1002/hipo.22376

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*, *107*(6), 1226–1238.e8.

Peters, J., & Büchel, C. (2010). Episodic Future Thinking Reduces Reward Delay Discounting

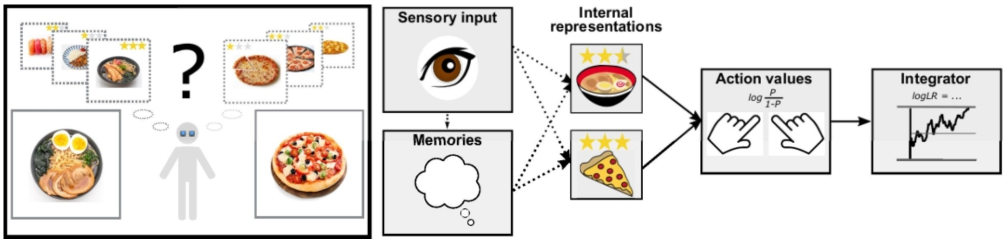
through an Enhancement of Prefrontal-Mediotemporal Interactions. In *Neuron* (Vol. 66, Issue 1, pp. 138–148). https://doi.org/10.1016/j.neuron.2010.03.026
Pezzulo, G., Donnarumma, F., Maisto, D., & Stoianov, I. (2019). Planning at decision time and in the background during spatial navigation. *Current Opinion in Behavioral Sciences*, *29*, 69–76.
Pezzulo, G., van der Meer, M. A. A., Lansink, C. S., & Pennartz, C. M. A. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, *18*(12), 647–657.
Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546–550.
Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251.
Polikar, R. (2012). Ensemble Learning. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning: Methods and Applications* (pp. 1–34). Springer US.
Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.
Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*(2), 148–152.
Prince, L. Y., Tsaneva-Atanasova, K., Clopath, C., & Mellor, J. R. (2017). Acetylcholine reconfigures hippocampal circuits to enable rapid formation of overlapping memory ensembles. In *Cold Spring Harbor Laboratory* (p. 201699). https://doi.org/10.1101/201699
Rao, R. P. N. (2010). Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Frontiers in Computational Neuroscience*, *4*, 146.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*. https://psycnet.apa.org/record/1978-30970-001
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.
Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356.
Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367.
Redish, A. D. (2016). Vicarious trial and error. *Nature Reviews. Neuroscience*, *17*(3), 147–159.
Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266.
Ritter, S., Wang, J. X., Kurth-Nelson, Z., Jayakumar, S. M., Blundell, C., Pascanu, R., & Botvinick, M. (2018). Been There, Done That: Meta-Learning with Episodic Recall. *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80:4354-4363
Rouhani, N., & Niv, Y. (2021). Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife*, *10*. https://doi.org/10.7554/eLife.61077
Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269.
Samejima, K., & Doya, K. (2007). Multiple representations of belief states and action values in corticobasal ganglia loops. *Annals of the New York Academy of Sciences*, *1104*, 213–228.
Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *eLife*, *9*. https://doi.org/10.7554/eLife.51140
Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677–

694.
Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic Future Thinking: Mechanisms and Functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50.
Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0049
Schlichting, M. L., & Preston, A. R. (2017). The Hippocampus and Memory Integration: Building Knowledge to Navigate Future Decisions. In D. E. Hannula & M. C. Duff (Eds.), *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* (pp. 405–437). Springer International Publishing.
Schmidt, B., Duin, A. A., & Redish, A. D. (2019). Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision making. *Journal of Neurophysiology*, *121*(6), 1981–2000.
Scott, B. B., Constantinople, C. M., Akrami, A., Hanks, T. D., Brody, C. D., & Tank, D. W. (2017). Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron*, *95*(2), 385–398.e5.
Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, *18*(3), 455–468.
Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. In *Psychological Review* (Vol. 115, Issue 4, pp. 893–912). https://doi.org/10.1037/a0013396
Shadlen, M. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. *Neuron*, *90*(5), 927–939.
Shahbaba, B., Li, L., Agostinelli, F., Saraf, M., Elias, G. A., Baldi, P., & Fortin, N. J. (2019). Hippocampal ensembles represent sequential relationships among discrete nonspatial events. In *Cold Spring Harbor Laboratory* (p. 840199). https://doi.org/10.1101/840199
Sharp, P. B., Miller, G. A., Dolan, R. J., & Eldar, E. (2020). Towards formal models of psychopathological traits that explain symptom trajectories. *BMC Medicine*, *18*(1), 264.
Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, *17*, 1249.
Shin, Y. S., & DuBrow, S. (2021). Structuring memory through inference-based event segmentation. *Topics in Cognitive Science*, *13*(1), 106–127.
Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology. General*, *142*(4), 1159–1170.
Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378–389.
Silver, D., & Veness, J. (2010). Monte-Carlo Planning in Large POMDPs. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems* (Vol. 23, pp. 2164–2172). Curran Associates, Inc.
Simon, D. A., & Daw, N. D. (2011). Environmental statistics and the trade-off between model-based and TD learning in humans. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 24* (pp. 127–135). Curran Associates, Inc.
Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, *79*(2), 361–374.
Socher, R., Gershman, S., Sederberg, P., Norman, K., Perotte, A., & Blei, D. (2009). A Bayesian analysis of dynamics in free recall. *Advances in Neural Information Processing*

Systems, *22*, 1714–1722.
Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews. Neuroscience*, *20*(10), 635–644.
Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological Review*, *119*(1), 120–154.
Spalding, K. N., Schlichting, M. L., Zeithamova, D., Preston, A. R., Tranel, D., Duff, M. C., & Warren, D. E. (2018). Ventromedial prefrontal cortex is necessary for normal associative inference and memory integration. *Journal of Neuroscience*, *38*(15), 3767–3775.
Srivastava, V., Feng, S. F., Cohen, J. D., Leonard, N. E., & Shenhav, A. (2017). A martingale analysis of first passage times of time-dependent Wiener diffusion models. *Journal of Mathematical Psychology*, *77*, 94–110.
Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653.
Staddon, J. E. R., & Davis, D. G. S. (1990). Memory for Reward in Probabilistic Choice: Markovian and Non-Markovian Properties. *Behaviour*, *114*(1-4), 37–64.
St-Amand, D., Sheldon, S., & Otto, A. R. (2018). Modulating Episodic Memory Alters Risk Preference during Decision-making. *Journal of Cognitive Neuroscience*, *30*(10), 1433–1441.
Stella, F., Baracskay, P., O'Neill, J., & Csicsvari, J. (2019). Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion. *Neuron*, *102*(2), 450–461.e7.
Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Networks: The Official Journal of the International Neural Network Society*, *18*(3), 225–230.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (Vol. 2). MIT Press.
Tajima, S., Drugowitsch, J., Patel, N., & Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, *22*(9), 1503–1511.
Tang, W., Shin, J. D., & Jadhav, S. P. (2020). Multiple time-scales of decision making in the hippocampus and prefrontal cortex. In *Cold Spring Harbor Laboratory* (p. 2020.10.17.343699). https://doi.org/10.1101/2020.10.17.343699
Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A Map for Social Navigation in the Human Brain. *Neuron*, *87*(1), 231–243.
Theves, S., Fernandez, G., & Doeller, C. F. (2019). The Hippocampus Encodes Distances in Multidimensional Feature Space. *Current Biology: CB*, *29*(7), 1226–1231.e3.
Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.
van Dijk, M. T., & Fenton, A. A. (2018). On How the Dentate Gyrus Contributes to Memory Discrimination. In *Neuron* (Vol. 98, Issue 4, pp. 832–845.e5). https://doi.org/10.1016/j.neuron.2018.04.018
Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., & Daw, N. D. (2019). Hippocampal Contributions to Model-Based Planning and Spatial Memory. *Neuron*, *102*(3), 683–693.e4.
Vikbladh, O., Shohamy, D., & Daw, N. (2017). Episodic contributions to model-based reinforcement learning. *Annual Conference on Cognitive Computational Neuroscience, CCN*. https://www2.securecms.com/CCNeuro/docs-0/5928307668ed3fec338a2570.pdf
Viswanathan, G. M., da Luz, M. G. E., Raposo, E. P., & Eugene Stanley, H. (2011). *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters*. Cambridge University Press.
Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin &*

Review, *26*(3), 813–832.
Wald, A., & Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test. *Annals of Mathematical Statistics*, *19*(3), 326–339.
Wang, J. X., Cohen, N. J., & Voss, J. L. (2015). Covert rapid action-memory simulation (CRAMS): a hypothesis of hippocampal-prefrontal interactions for adaptive behavior. *Neurobiology of Learning and Memory*, *117*, 22–33.
Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: a query-theory account. *Psychological Science*, *18*(6), 516–523.
Weidemann, C. T., Kragel, J. E., Lega, B. C., Worrell, G. A., Sperling, M. R., Sharan, A. D., Jobst, B. C., Khadjevand, F., Davis, K. A., Wanda, P. A., Kadel, A., Rizzuto, D. S., & Kahana, M. J. (2019). Neural activity reveals interactions between episodic and semantic memory systems during retrieval. *Journal of Experimental Psychology. General*, *148*(1), 1–12.
Weilbächer, R. A., & Gluth, S. (2016). The Interplay of Hippocampus and Ventromedial Prefrontal Cortex in Memory-Based Decision Making. *Brain Sciences*, *7*(1). https://doi.org/10.3390/brainsci7010004
Weilbächer, R. A., Kraemer, P. M., & Gluth, S. (2020). The reflection effect in memory-based decisions. *Psychological Science*, *31*(11), 1439–1451.
Wiering, M. A., & van Hasselt, H. (2008). Ensemble algorithms in reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society*, *38*(4), 930–936.
Wieschen, E. M., Voss, A., & Radev, S. (2020). Jumping to conclusion? a lévy flight model of decision making. *TQMP*, *16*(2), 120–132.
Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279.
Wimmer, G. E., Elliott Wimmer, G., Li, J. K., Gorgolewski, K. J., & Poldrack, R. A. (2018). Reward Learning over Weeks Versus Minutes Increases the Neural Representation of Value in the Human Brain. In *The Journal of Neuroscience* (Vol. 38, Issue 35, pp. 7649–7666). https://doi.org/10.1523/jneurosci.0075-18.2018
Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273.
Wong, K.-F., Huk, A. C., Shadlen, M. N., & Wang, X.-J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Frontiers in Computational Neuroscience*, *1*, 6.
Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. In *Nature Human Behaviour* (Vol. 2, Issue 12, pp. 915–924). https://doi.org/10.1038/s41562-018-0467-4
Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080.
Yartsev, M. M., Hanks, T. D., Yoon, A. M., & Brody, C. D. (2018). Causal contribution and dynamical encoding in the striatum during evidence accumulation. *eLife*, *7*. https://doi.org/10.7554/eLife.34929
Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews. Neuroscience*, *7*(6), 464–476.
Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *The European Journal of Neuroscience*, *19*(1), 181–189.
Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. In *European Journal of Neuroscience* (Vol. 22, Issue 2, pp. 513–523). https://doi.org/10.1111/j.1460-9568.2005.04218.x

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews. Neuroscience*, *20*(6), 364–375.
Yu, A. J. (2007). Adaptive Behavior: Humans Act as Bayesian Learners. *Current Biology: CB*, *17*(22), R977–R980.
Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: building memories to navigate future decisions. *Frontiers in Human Neuroscience*, *6*, 70.
Zhao, W. J., Richie, R., & Bhatia, S. (2020). Process and Content in Decisions from Memory. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.*



Decisions can be informed by multiple sources of evidence, including memories and multiple scales of associations. These are transformed internally to action tendencies, which are then integrated to make decisions.

103x24mm (600 x 600 DPI)