Decoding cognition from spontaneous neural activity

Yunzhe Liu,[1,2,3]* Matthew M Nour,[3,4] Nicolas W Schuck,[3,5] Timothy E J Behrens,[4,6] Raymond J Dolan [1,3,4,7]

1. State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
2. Chinese Institute for Brain Research, Beijing, China
3. Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London/Berlin, UK/Germany
4. Wellcome Centre for Human Neuroimaging, University College London, London, UK
5. Max Planck Research Group Neurocode, Max Planck Institute for Human Development, Berlin, Germany
6. Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK
7. Department of Psychiatry, Universitätsmedizin Berlin (Campus Charité Mitte), Berlin, Germany

∗ Corresponding author. E-mail: yunzhe.liu@bnu.edu.cn

Abstract

In human neuroscience, studies of cognition are rarely grounded in non-task evoked, "spontaneous", neural activity. Indeed, studies of "spontaneous" activity tend to focus predominantly on intrinsic neural patterns, for example, resting-state networks. Taking a "representation rich" approach bridges an apparent gap between cognition and resting-state communities: this approach relies on decoding task-related representations from spontaneous neural activity, allowing for quantification of the representational content and rich dynamics of such activity. For example, if we know the neural representation of an episodic memory, we can decode its subsequent replay during rest. We argue that such approach advances cognitive research beyond a focus on immediate task demand and provide insight into the functional relevance of intrinsic neural pattern (e.g., default mode network). This in turn enables a greater integration between human and animal neuroscience, facilitating experimental testing of theoretical accounts of intrinsic activity, and opening new avenues of research in psychiatry.

Introduction

An overarching goal of neuroscience is to understand the relationship between cognition and underlying neural activity. In humans, much progress in this direction has been driven by **task-based** neuroimaging studies. However, it is salutary to note that a large proportion of variance in neural activity is **off-task** [1,2]. We deploy the term "spontaneous" to refer to intrinsic activities that are not mere response to external events. We review past work and an emerging new paradigm for studying cognition in the spontaneous neural activity.

A dichotomy in human neuroscience

In cognitive neuroscience, a standard approach is to temporally align neural activity to specific task-events (e.g., presentation of a visual stimulus, like "house"), and localise the corresponding neural response in the brain (**Figure 1a**). In this way, neuroscientists have provided a rich characterisation of cognition in the context of task-evoked processing.

A major difficulty in studying spontaneous neural activity, as compared to task-evoked, is that researchers do not have direct access to either the identity or timing of putative states hypothesized to drive these neural activity (hence the term "spontaneous"). This makes attribution to a causal cognitive process less straightforward. Broadly speaking we can conceptualise two broad approaches, spanning two ends of a continuum for studying spontaneous neural activity (**Figure 1b**).

On one hand, neuroscientists adopt a "data-driven" approach (**Figure 1b, right**). Here, instead of a concern with changes in neural activity "triggered" by external events, the focus is on intrinsic physiological features, such as **functional connectivity**, power or phase-coupling of frequency-specific oscillations (e.g., alpha rhythm [3]). The putative cognitive role of these features are sometimes inferred based on their relationship to behavioural or psychological measures [3,4], or by evaluating an anatomical overlap with task evoked activity patterns reported in other studies [5,6]. This approach aims to find

a correspondence between intrinsic and task evoked neural patterns (sometimes termed the brain's 'functional architecture') [7]. Nevertheless, such "data driven" approaches are not as well positioned to provide a cognitive grounding because a direct relationship to task events cannot be assumed (unlike in task evoked studies, where neural activity and task events can be temporally aligned).

At the other end of the spectrum, researchers have studied spontaneous activity using approaches that bear similarity to the analysis of task-evoked activity, which we term "intervention-based". This approach introduces an external intervention known to evoke an associated event within an otherwise task-free session (**Figure 1b, left**). Examples here include using **targeted memory reactivation (TMR)**[8,9] in sleep studies [10,11]; tracking spontaneous neural activity during pre-stimulus time [12], or presenting stimuli at the peak (strongest time) or trough (weakest time) in spontaneous fluctuation of neural activity of interest (e.g., dopaminergic midbrain) [13]. In this way, spontaneous neural activity can be analysed by aligning to the onset of such events, with precise timing akin to task evoked studies. Unlike a "data-driven" approach, this "intervention-based" approach enables a direct inference with respect to underlying cognitive process. This benefit, however, comes at the expense of interrupting internal computations that are "spontaneous" in nature, such as imagination or mind wandering - processes that are by definition not tied to immediate task demand [14]. Thus, cognitive neuroscientists were faced with a choice: to study unperturbed spontaneous neural activity, albeit with a restricted window on to its functional relevance; or instead to have more explicit control of cognitive process, but at the expense of disrupting spontaneity.

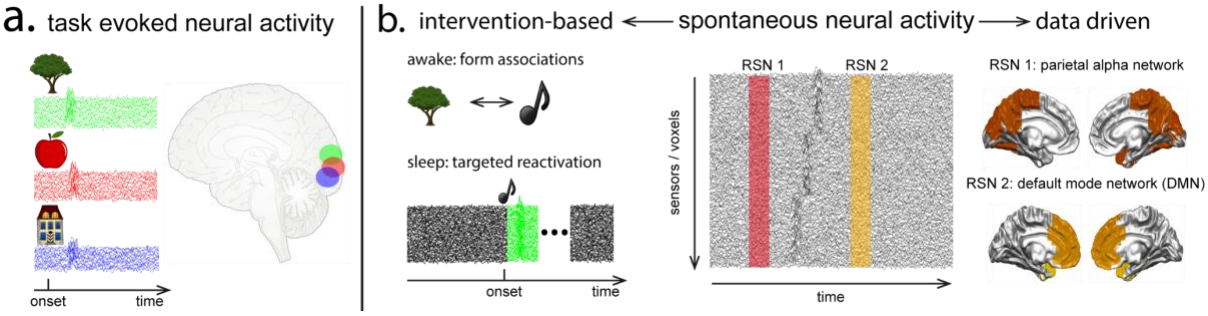

Fig. 1. A (relative) dichotomy in human neuroscience. a. Cognitive processes are typically studied by aligning neural activity to the onset of perturbing stimuli, with the aim of finding underlying neural correlates. For example, using functional magnetic resonance imaging (fMRI), evoked neural activity in response to different stimuli (or task demands) are then mapped in the brain, here illustrated by responses to a house (blue), apple (red) and tree (green) respectively. **b.** Studies of spontaneous neural activity can be rendered more akin to task evoked studies by introducing an external intervention. For example, in a TMR approach, a tone associated with a specific stimulus is used to evoke task-related processing during sleep (left panel, green colour indicates tree-related processing, elicited by its paired tone). Alternatively, a data driven approach can be used to characterise the canonical functional connectivity patterns during rest (far right panel). Two examples of resting state networks (RSNs (parietal alpha network (red colour) and default mode network (DMN, yellow colour)). Although there may be task-related reactivations during rest (e.g., transient synchronous activity bumps in between RSN1 and RSN2 epochs, shown in darker patch), their functional relevance is not accessible to methodologies employed within standard resting state studies.

Panel b adapted with permission from Higgins, et al. [15]

In recent years, an emerging approach has endeavoured to combine the best of both techniques outlined above. This approach relies on exploiting the representational content of neural activity and is predicated on an assumption that the same neural representations of task events (e.g., “house” or “apple”) are active both on-task and off-task. Such representations can first be derived from task evoked neural activity and their reactivations subsequently obtained through **decoding**. In essence, this approach probes the task-relevant content of intrinsic neural activity, going beyond a characterisation from the “data-driven” approach. In so doing it provides information regarding both when and what representation has been activated, absent from external intervention. We refer to this line of research as “**representation rich**”.

Task related representations can be obtained in multiple ways. One approach is to rely on the neural pattern similarity of task events between on- and off-task[16] or different brain regions[17]. For example, to look for features of memory consolidation, Tambini and Davachi [18] compared the pairwise multivoxel correlation structure between stimuli at encoding and post-encoding rest, versus that at pre-encoding rest, and found increased hippocampal pattern similarity attributable to learning. This correlation-based approach has similarities to representational similarity analysis (RSA) widely used to study task evoked neural activity [19,20], and most often for localising where (in the brain) a pattern emerges.

A “representation rich” approach to spontaneous neural activity aims to uncover the temporal structure of task-related representations, e.g., how their temporal dynamics unfold [21-25]. This is typically implemented using a decoding-based method, transforming spontaneous neural activity into a time series of task related reactivations [23-25]. This line of research in human neuroimaging has an interesting parallel in animal work. For example, in rodent hippocampus, researchers have identified pyramidal cells that encode spatial locations during active navigation, known as place cells [26], and also observed these same cells fire spontaneously in an organised sequence during rest. This firing recapitulates past or potential future trajectories and is referred to as “hippocampal replay” [17,27-40]. The ability to read-out reactivation of specific locations during rest allows researchers to go beyond a mere characterisation of neurophysiological features, e.g., **sharp wave ripples (SWRs)** [41], enabling a probing of the representational content of neural activity, particularly with respect to task variables. This feature has allowed studies of hippocampal replay to forge a link between cognition and physiology [42], and in so doing, shed light on a range of cognitive functions subserved by spontaneous neural activity, including memory, learning and decision-making [36,43,44] (**Box 1**).

Recent technical advances for characterising task-related reactivations in human neuroimaging has inspired a series of studies investigating “human replay” (**Figure 2a**). These address complex forms of non-spatial cognition [23,25,45], especially those informed by reinforcement learning (RL) models[46-48] (**Box 2**). They also provide a unique opportunity to link resting dynamics of whole-brain connectivity (e.g., DMN) to spontaneous task-related reactivation (e.g., replay) [15].

In this review, we first discuss these technical advances and then go on to consider studies that have exploited this approach. Our aim is to demonstrate the exciting prospects afforded by “representation-rich” approach in bridging task-based and spontaneous brain activity.

Measuring spontaneous reactivations

Studies of neural dynamics in human brain non-invasively typically involve electroencephalographic (EEG) or magnetoencephalography (MEG) monitoring, performed over a period of several minutes or up to hours. A "decoding-based" approach to characterising its task related information involves two stages (**Fig. 2b**). The first identifies a mapping between neural activity and a task variable of interest. In the simplest example, these variables are discrete sensory stimuli (e.g., apple, house). The mapping can be obtained by training either a **discriminative model** or a **generative model** linking object labels to their associated neural responses (e.g., multi-voxel patterns in the case of fMRI, or multi-channel patterns in the case of MEG or EEG [49]). Typically, the training data is obtained from an incidental 'localizer' task, in which multiple examples of task events and associated evoked neural activity are collected. This stage is similar to multi-voxel pattern analysis (MVPA) [50-52] (**Fig. 2b, left**).

The second stage applies these trained models to neural activity obtained at other time periods of interest, which might constitute off-task rest or on-task epochs, and yields a time course of spontaneous task related reactivations [45,53] (**Fig. 2b, right**). In this way, researchers can access the representational content of such activity, thereby allowing a comparison between the observed temporal structure of neural reactivations (e.g. whether reactivation of 'task state A' reliably precedes that of 'task state B') and patterns derived from a formal hypothesis (e.g., transition matrix in a RL-based model [25]).

Note, the primary difference between an analysis of task evoked neural responses (e.g., RSA), and an approach that focuses on spontaneous neural activity, is that the former tracks neural representations by reference to the explicit timing of task events. By contrast, the temporal characterisation of task related reactivations is itself a primary research question in the analysis of spontaneous neural activity, made tractable by an assumed overlap between evoked and reactivated task representations. This assumption raises several methodological considerations.

Methodological considerations

The first issue relates to the fact that spontaneous neural reactivation of a task event is likely to be less pronounced compared to its evoked response (not least because the very neural decoders used to identify reactivations in spontaneous activity are trained using evoked neural responses). This can be considered as providing an increased risk of false negatives when quantifying spontaneous reactivations in off-task neural activity (Type II error). A second consideration is that without '**ground truth**' information about the identity and timing of task related reactivations, the statistical inference procedure must also protect against false positives (that is, Type I error), for example those arising due to non-specific neural dynamics, such as **autocorrelations** [53,54]. Recent methodological work suggests that linear modelling with careful control of confounding regressors (for example, reactivations of other states, see details in Liu, et al. [53]) and appropriate permutation-based statistical inference procedures are sufficient to deal with these concerns in human MEG or EEG, as well as rodent electrophysiology data [53] (**Fig. 2c**).

Other considerations relate to data modality specific features. For example, in fMRI, we rely on the blood oxygenation level dependent (BOLD) signal, which has a

temporal resolution in the order of seconds, one that is almost certainly too slow to capture complete neural reactivation patterns associated with neural replay (on the order of milliseconds). Nevertheless, recent work demonstrates that such fast sequences can be detected using an fMRI decoding approach that assumes reactivation patterns cause systematic neural patterns of overlap in the delayed BOLD responses [45] (**Fig. 2d**).

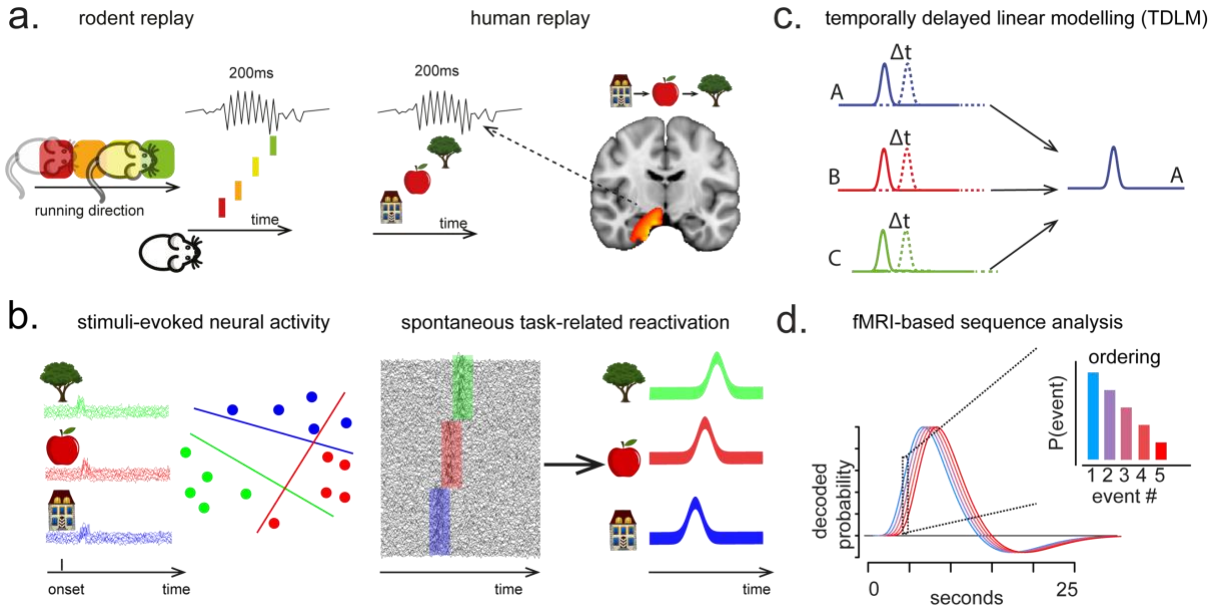

Fig. 2. "Representation rich" paradigm on spontaneous neural activity. a. Neural replay in rodents and humans. Left panel: rodent replay. Different colours indicate firing fields of place cells in rodent hippocampus. During rest, those place cells are reactivated in a consistent order, termed "replay". Rodent replays are time compressed, and typically associated with short-wave ripples (SWRs) - the fast oscillations depicted in the figure. Right panel: human replay. Human replays (of visual stimuli) are also time compressed and associated with fast oscillations. The brain activation figure shows an initialization of replay events that arises in hippocampus. **b**. Decoding-based approach, comprised of two stages. Left panel: stage 1 - indexing neural representations of different task objects. This can be achieved by training a decoding model and finding a multivariate decision boundary in the data pertaining to each object. Examples of 3 task objects: tree (green), apple (red) and house (blue) are shown here, with dots indicate samples and lines denote decision boundaries. Right panel: stage 2 - applying these trained decoding models to spontaneous neural activity of interest, enables us to ascertain what has been reactivated and when. Transient bumps in the spontaneous neural activity can be identified as task-related reactivations for tree (green), apple (red) and house (blue). **c.** Illustration of temporally delayed linear modelling (TDLM) approach. It asks whether averaged statistical likelihood of some transitions (e.g., B→A) happening more than others (e.g., C→A). Dashed lines indicate the time shifted (by Δt) copies. **d.** Illustration of a fMRI-based sequence analysis approach. Fast sequences of events will cause systematic patterns of overlap in delayed responses that can be inferred [45]. Different coloured lines indicate time course of different decoded events. The inset shows an ordering of their reactivation strength at a given time slice.

Panel a is adopted with permission from Ref 23.

Panel c is adapted with permission from Ref 53.

Panel d is adapted from Ref 45.

Spontaneous task-related reactivations

Investigation of spontaneous task-related reactivations in humans can be thought of as falling into two broad categories: 1) reactivation during rest (e.g. off-task reactivation) [15,23,25]; 2) reactivation during task performance (e.g. during intervals between sequential trials within a task) [21,22,55-61]. Note, spontaneous neural activity measured during off-task periods is not necessarily free from task-related influences. It is likely that reactivations in both time periods are subject to task-related modulation [62]. While "on-task" reactivations (for example those occur in inter-trial intervals) relate more directly to immediate task demands, we include these phenomena under the rubric of 'spontaneous' reactivations as the representational content in question relates to states that are not immediately determined by current sensory input. For example, mental simulations of future experiences or the recall of episodic memories may all be detected 'on task' (i.e., during a task inter-trial interval or after receipt of an outcome), yet 'go beyond' immediate task evoked processing. Both categories of spontaneous activity can be studied in a similar manner under "representation rich" paradigm. We note that studies of human reactivations (both off-task and on-task) are sometime referred to as "replay" in the literature[58,63]. For consistency, in this paper, we use "reactivation" to refer to task representations obtained during learning and encoding time that are later reinstated in spontaneous neural activity [64]. We use the term "replay" to describe a *sequential* reactivation of these task representations.

Off-task spontaneous reactivation

Outside of a "representation rich" approach, studies probing the functional relevance of resting-state activity typically link physiological features (e.g., functional connectivity) to behavioural measures of task performance collected before, or after, a resting session. For example, using fMRI, Tambini, et al. [65] reported enhanced functional connectivity between hippocampus and lateral occipital cortex during rest following an associative memory encoding task, which was related to later memory performance. This approach is analogous to linking electrophysiological signatures of reactivation (e.g., SWRs) to memory consolidation in rodents [42].

Implementing a "representation rich" paradigm, other studies have endeavoured to probe cognitive process within resting state activity [14]. For example, comparing stimulus encoding patterns in hippocampus for pre- vs. post-encoding rest period, has reveal a stronger memory reactivation during post-encoding rest that relate to enhanced memory performance [18]. In a decoding analysis of EEG signals, previous learnt information was found to be reactivated during sleep at category-level [66]. Using fMRI acquired during sleep, reactivation of past memory was found at both category-level [67], as well as the level of individual stimuli [68]. More recently, Schapiro, et al. [63] showed that prioritized reactivation of weakly encoded memories in hippocampus during awake rest benefits later memory performance. Together, these studies provide evidence that offline hippocampal reactivation plays an important role in human memory consolidation [69]. This research trend towards exploitation of a representation rich approach enables a fine-grained tracking of the representational content and dynamics of reactivations, mirroring the tracking of spontaneous place cell activity in rodent hippocampus during rest.

Sequential reactivation (or replay) in humans concerns the ordering of reactivated task states at the representation level. The sequential quality of such reactivation renders it a suitable testbed to assess the neural representation of a **cognitive map**, where only a limited number of transitions (e.g., A→B, B→C are valid sequences, but C→A is not) are allowed under a given relational structure (e.g., A→B→C). Schuck and Niv [25], took fMRI-based decoding a step further to provide evidence for sequential reactivation of task states in human hippocampus during rest. Here a focus on representational content enabled the authors to make claims about the regularity and consistency of the same experiences between rest and task, enabling them to connect hippocampal replay to the neural representation of a mental model in orbitofrontal cortex [70]. Specifically, the authors suggested that human replay might participate in building or maintaining a mental representation of the task structure during rest. Relatedly, using MEG at millisecond temporal resolution, Liu, et al. [23] demonstrated that organized experiences are sequentially replayed during rest at a fast time scale (40 ms state-to-state transition). Such replays were not mere "echoes" of past experience [35], but instead were ordered in a manner consistent with a learnt task structure, again, suggestive of replay building or maintaining cognitive map off task.

Of particular note is that the spontaneous sequences of cortical events, detected in a non-spatial context in humans using MEG [23], show striking parallels to the characteristics of hippocampal replay in rodents during sharp-wave ripple epochs in spatial tasks [27,31,36]. Like rodents, human replays (i) appear spontaneously during rest, (ii) compress time from seconds to tens of milliseconds [28], (iii) reverse in direction following receipt of reward [37], (iv) involve a coordination between hippocampus and sensory cortex [32], and (v) are associated with a power increase in ripple frequency (120 Hz -150 Hz), that can be source localised to hippocampus.

On-task spontaneous reactivation

A rich literature has characterised representational content of spontaneous neural activity on-task, ranging from perception [12] to action [71]. An interesting finding is that reactivation of objects in mind can bias subsequent perception, and in extreme cases, leads to hallucinations. For example, Pajani, et al. [12] showed that the representational content of pre-stimulus activity in early visual cortex is linked to subsequent perception: if there is a bias toward the expected grating stimuli, it could predispose to perceptual hallucination. Similarly, Hahamy, et al. [72] found spontaneous fluctuations in early visual cortex might activate the visual hierarchy, and drive hallucination in participants with **Charles Bonnet syndrome**.

Over the past 5 years, there has been a upsurge in representation-rich research on memory and decision-making in the context of RL [73], especially model-based RL [74]. In RL terms, a "model" details the relationships between current and future states. This has a similar meaning to the notion of a "cognitive map". A commonality among these is that a model allows us to infer things we have not experienced directly or explicitly [75], enabling for instance multi-step planning [22] or inferential learning [24]. If 'off-task' reactivation relates to building or maintaining a mental model of the world, then 'on-task' reactivation might be a means to study how such a model is utilised for adaptive behaviour. The study of model-based reasoning de facto concerns probing internal processes that are not tied to current sensation, a line of investigation fits well with "representation rich" approach.

One important field here is that of memory-based research [73], which typically focuses on the cued retrieval of associative memories [76]. Studies of associative memory entail an encoding phase (e.g., a house → tree association, house is the cue, tree is the associated event), followed by a cued retrieval phase (e.g., house →?). A pairwise association can be conceived as a simple relational structure, with cued retrieval framed as memory-based decision making [73]. For example, using intracranial EEG, Norman, et al. [77] found successful memory recall was preceded by an increased probability of hippocampal SWRs, during which there was also a transient re-emergence of activation patterns in higher visual cortical areas. Relatedly, Vaz, et al. [76] found bursts of spikes in the human temporal lobe that were organized into sequences during memory encoding, and these same sequences were 'replayed' during successful memory retrieval. Using fMRI, Wimmer and Shohamy [60] found monetary rewards led to automatic reactivation of a past associative memory within the hippocampus, with these reactivations biasing later value-based decision-making. This effect was subsequently replicated in a MEG-based version of the same task, using a decoding based technique [57]. More recently, Wimmer, et al. [59] studied memory retrieval of more extended episodes consisting of multiple elements, and found a differentiation among retrieval patterns involving clustered representations compared with sequential reactivation of individual episode elements, with more strongly encoded memories retrieved via a clustered reactivation and weaker memories by sequential reactivation [59].

Other 'on-task' research has focused on activity during decision time or following outcome feedback in value-based decision-making tasks [21,22,55,56,58,61]. In rodents, this is typically studied in a spatial navigation setting [37,78]. In humans, we are able to probe more flexible cognitive processes in an abstract task space using the approach outlined. Here a common experimental design involves model-based RL tasks where participants update the value of each action (in each state) based on experienced or inferred (model-based) rewards, and then make a choice based on these values. For example, when performing a two-step like task in fMRI (**Box 2**), prospective reactivation of task goal was found to support model-based choice [55], on-task reactivation of counterfactual value signals reflected the non-chosen option [79], and spontaneous reactivation following reward receipt was modulated by prediction error [58]. In the model-based inference, an important role of hippocampal-orbitofrontal interactions is highlighted [80]. Other decision-making studies report spontaneous reorganisation of task strategies reflected in neural representations, even before a strategy change is evident in behaviour [81].

Using a sequentially structured RL task in combination with MEG, Liu, et al. [24] identified reverse sequential reactivation (reverse replay) of non-local (i.e. inferred) experiences following reward receipt, with a 160 msec state-to-state lag, akin to a putative neural mechanism for model-based RL. This replay was more pronounced for experiences of greater utility for future behaviour, consistent with RL theory [48]. Although this study focused on outcome time, other studies have probed mental planning process at choice time. In a non-spatial sequential planning task, Kurth-Nelson, et al. [22] found fast reverse sequences during planning, although these sequences did not reflect the path to be taken, but instead represented equally all valid transitions. This is suggestive of a process involving mental rehearsal of sequences, possibly to maintain a representation of task structure. Subsequently, Eldar, et al. [56] varied both problem complexity and temporal constraints in a decision task, showing that people differ substantially in terms of their decision strategies. Using a different

task, Eldar, et al. [21] further demonstrated that on-task (during planning) and off-task (during rest) replay supported planning in distinctive ways, dependent on actual decision strategies. Such model-based planning has also been recently found in the aversive domain [61], akin to findings in the rodent literature [40].

Bridging cognition and physiology

Having considered how a "representation rich" approach might advance our understanding of human cognition, we now discuss potential insights this approach can bring to the studies of resting states, and in particular to psychiatry research where there has long been a strong emphasis on spontaneous neural activity [2,82,83].

A predominant focus of resting-state studies is the intrinsic physiological features of brain activity (e.g., the covariance of activity between brain regions, termed 'functional connectivity'). One common approach is to characterise the functional connectivity patterns relating different regions in terms of whole-brain resting state networks (RSNs) [5,84,85]. Among such RSNs, the DMN [85,86] is of particular interest [87].

The DMN is a characteristic pattern of inter-connected brain regions (that includes medial prefrontal cortex, posterior cingulate cortex, and medial temporal lobes) that show high co-activation off-task [87-89]. Initially, the DMN was thought to be 'task negative', as it is typically deactivated during task execution, e.g., working memory [90,91]. Later, brain regions that comprise the DMN were found to be related to internally oriented cognitive states [16], such as imagination [92], mind wandering [93], memory recall [94], planning [95], or consolidating social information [96]. In recent studies focusing on the neural codes underpinning mental models (e.g. conceptual spaces [97], social spaces [98], or narrative schemas [99]), the distribution of such neural profiles bear a remarkable overlap with DMN regions during rest, leading to a suggestion that this functional network might be encoding the cognitive map of task space [75]. These interpretations, however, derive in large part from an anatomical overlap with brain regions reported in task-based cognitive studies.

Ideally, we want to link functional connectivity patterns (e.g., DMN) to concurrent spontaneous cognition within the *same* dataset. In a "representation rich" approach, this can be achieved by studying the relationship between DMN activations and task-state related reactivations. Combining two recent methodological advances in MEG analysis - measurement of sequential replay during rest [23,53] and tracking of DMN activation dynamics with millisecond temporal resolution [100] - Higgins, et al. [15] established a connection between spontaneous human replay and DMN activation in the same resting-state session. More specifically, the authors showed that human replay exhibits a highly organised temporal structure, where replay events did not occur randomly but were instead packaged into transient bursts. The latter coincided with a concentration within epochs of DMN activation (**Figure 3a**) characterised by large synchrony in the delta and/or theta band. Moreover, DMN was unique among all RSNs in its association with transient increases in higher frequency power (including the frequency band associated with ripple), which source localised to temporal lobe. This work suggests that a coupling between temporal-lobe SWR and DMN might provide a physiological basis for how human replay supports memory consolidation.

A coupling between temporal-lobe SWR and cortical DMN during rest [15] is reminiscent of prior in-vivo work in nonhuman primates. For example, using simultaneous whole-

brain fMRI as well as hippocampal electrophysiology recordings, Kaplan, et al. [101] found a selective increase in the DMN following hippocampal ripples, but not other RSNs or hippocampal electrophysiological events (**Figure 3b**). Similarly, using simultaneous recordings of hippocampal electrophysiology during wide-field calcium imaging of cortical activity in the mouse brain, Liu, et al. [102] found cortical–hippocampal coordination involving hippocampal SWR and medial parietal cortex (part of DMN) activation (**Figure 3c**). This set of results suggest a plausible cross-species function for DMN during rest in supporting off-task memory consolidation (or map building), potentially through replay in coordination with SWR.

Contrary to findings in relation to off-task replay, studies of human on-task replay have not shown an association with high frequency power increases [59]. One intriguing hypothesis is that on-task replay in humans (e.g., the slower replay with 160 ms state to state time lag [24]) might relate to theta sequence seen in rodents [103-105] (**Box 1**), and reflect a more 'conscious' on-task computation, a rich topic for future work.

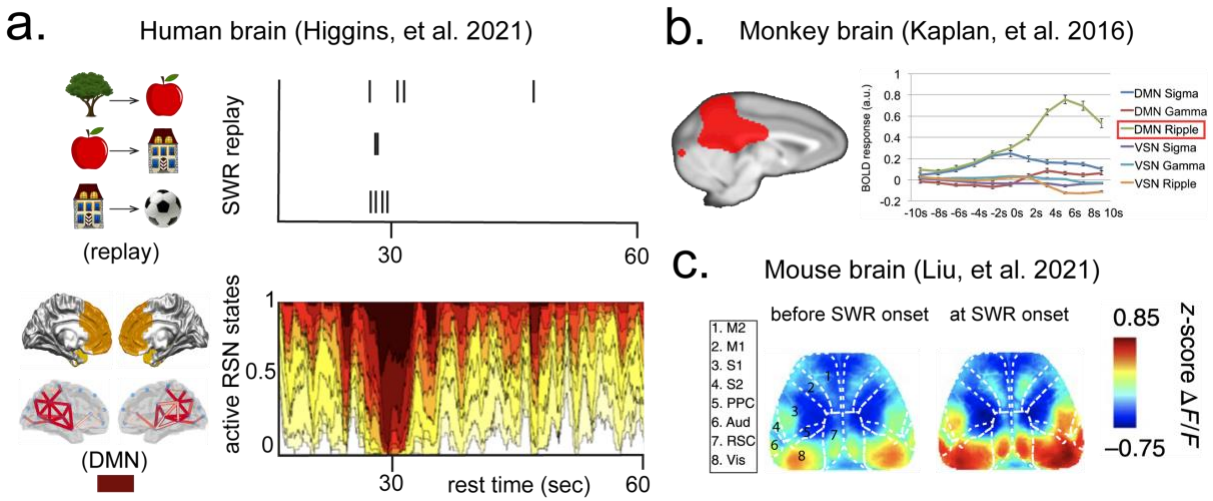

Fig. 3. Coordination between hippocampal sharp wave ripples (SWR) and large-scale cortical activity across species. a. In humans, the top part shows an example (60 sec in a 5 min resting state) of raster plot (up right) of three replay sequences (up left). The lower part shows time course of different resting state network reactivation (bottom right, RSNs - from 1-12, brown to yellow), and the power (yellow) and phase locking (red) profile of default mode network (bottom left), Y axis indicates reactivation probability of RSNs, higher value means stronger activation. The default mode network DMN (brown colour) shows the strongest activation at the time of SWR replay bursts. **b.** In monkey, hippocampal SWR has also been shown to be coupled with DMN (left panel, red area, measured using fMRI). This coupling is specific to DMN (among other RSNs) and ripple frequency (among other frequency bands). **c.** In mice, widespread cortical activation is associated with onset of hippocampal SWRs. Dashed lines indicate identified cortical regions based on Allen Brain Atlas. M2, secondary motor cortex; M1, primary motor cortex; S1, primary somatosensory cortex; S2, secondary somatosensory cortex; PPC, posterior parietal cortex; Aud, auditory cortex; RSC, retrosplenial cortex; Vis, visual cortex. This cortical activity (including retrosplenial cortex, a part of the DMN) rise occurs right before onset of hippocampal SWRs. ΔF/F measure the change in fluorescence intensity relative to its resting level, higher value indicates higher activation.

Panel a is adapted with permission from Ref 15.

Panel b is adapted with permission from Ref 101.

Panel c is adapted with permission from Ref 102.

Representation rich paradigm in psychiatry research

Finally, there are compelling reasons to believe a representation rich approach can yield novel insights into the neurobiology of psychiatric disorders. Over the last two decades, extensive fMRI, MEG and EEG investigations have reported widespread alterations in resting state network characteristics in neuropsychiatric patients compared to healthy volunteers [82,83]. However, in these studies, linkages to clinical variables and cognitive functions are generally inferred indirectly, e.g., by relating brain activations to clinical questionnaires [106]. In view of the relationship between spontaneous neural activity and model-based cognition, we consider "representation rich" perspective is uniquely placed for bridging a conceptual gap between brain activity, cognition, and clinical symptoms in psychiatric disorders.

With this in mind, a condition of particular interest is schizophrenia, a neuropsychiatric disease characterised by symptoms such as delusions, hallucinations and 'thought disorder' [107]. Genetic mouse models of schizophrenia have identified signatures of abnormal hippocampal reactivations during rest (e.g., augmented SWR power and temporal disorganization of place cell reactivations), suggesting that abnormal replay plays a central role in this debilitating condition [108-110]. Patients with a diagnosis of schizophrenia (PScz) exhibit impairments in inferring indirect associations (e.g., if A>B, B>C, what is relationship between A&C) [111,112], which can be interpreted as a deficit in leveraging cognitive map. Recently, utilizing an MEG decoding approach, Nour, et al. [113] examined neural replay in PScz (vs. carefully matched control participants) during rest following a similar relational inference task as Liu, et al. [23]. The authors found that spontaneous neural replays of learned task structure, as evident in control participants akin to findings of Liu et al [23], were reduced in PScz. Intriguingly, PScz were characterised by augmented SWR power during replay, and a distorted neural representation of cognitive map, consistent with the genetic mouse model [108]. The sequential replay deficit in PScz related to behavioural impairment in inferring correct sequential relationships between task states, a process likely to require abstracted representation of task structure. These results raise a tentative suggestion that previous reports of DMN deficits in schizophrenia [114] might relate to a compromised maintenance of correct cognitive map during rest.

This early study of spontaneous neural replay in a clinical population motivates future studies across a range of psychiatric disorders, for example testing computational hypotheses pertaining to sequential planning deficits (e.g., maladaptive pruning) and recurrent intrusive thoughts (e.g., obsessions and ruminations) in disorders such as anxiety and depression [115-117].

Future directions

A central goal of neuroscience is to understand how neural activity supports cognition and thereby adaptive behaviour [74,75]. We suggest a relative neglect of cognition in the context of spontaneous neural activity (e.g., resting state) can now be redressed by reference to a "representation rich" approach, and where emerging data indicates it can also inform the study of model-based cognition. We envisage three broad directions where a representation rich paradigm will advance understanding of cognition in humans and animals alike.

Sleep and cognition

An exciting direction now open to investigation is decoding the representational content of sleep. This has broad relevance for understanding both the functional relevance of sleep and its translational implications [118], given its deficits is shown in most of psychiatric disorders.

It has long been hypothesised that memory consolidation, or new memory formation, unfolds during SWR events in slow wave sleep [119-121]. In fact, early studies of rodent sleep replay showed that a time compression feature[27,122] of neural reactivation supports Hebbian learning by reactivating memory traces within a time window that is amenable to spike timing dependent plasticity [36,123]. Most recently, evidence of theta sequence expression during rapid eye movement (REM) sleep has been reported in rats after they completed a spatial learning task, suggesting a role for REM associated theta sequences in memory function [124]. In addition, there are interesting differences in replay dynamics during awake rest vs. sleep [125,126]. For example, replays in rodents have been found to represent Brownian diffusive spatial trajectories during sleep [127], while resembling more super-diffusive dynamics during awake rest [126], hypothesized to serve different computational goals [125].

In humans, although sleep is a topic of intense investigation, sleep replay has rarely been studied (however, cf.[66-68] for studies of memory reactivation during sleep). A significant barrier here is the considerable difference in neural signals between sleep and awake [128] states, such that decoding models trained during awake may generalize poorly to sleep time [129]. A promising direction is to explore a family of generative models. Thus, instead of treating the mapping between a task variable and multivariate neural activity as a black box (as is the case in discriminative models), this class of model specifies the generative process of neural activity in relation to task variables, a priori. These class of model can, in principle, generalize better from awake to sleep if the underlying assumptions are reasonably met [49,130]. These approaches may also allow an exploration of links between sleep replay and dreaming in humans.

Integrating human and animal neuroscience

A unique advantage of "representation rich" paradigm is its ability to integrate findings across species. This is because neural signals (either electrophysiology in rodents or neuroimaging in humans) can be transformed into task-related representations. This focus on the representational level renders possible a comparison of human and animal neuroscience findings.

Previous work on the spatial organization of neural codes for visual objects in the brain, for example, have found inferior temporal cortex supports a common neural representation profile for animate vs. inanimate objects across humans and monkeys [131]. Recent work on the temporal profile of spontaneous neural reactivations (e.g., replay), suggests that human replay bears a strong resemblance to that seen in rodents [23] . This leads to an expectation of greater crosstalk and assimilation of findings across species under a "representation rich" paradigm [132], especially with utilization of domain general methods (e.g., TDLM [53]).

In a recent example of this integrative approach, Barron, et al. [133] probed inferential decision making in both humans (with fMRI) and mice (with electrophysiology) with a similar associative inference task, and revealing hippocampal involvement in both

species. In this study, hippocampal replay in rodents represented inferred relationships during rest, whereas human hippocampus was found to use a prospective code to forecast learned associations. It is possible that when solving the same task, distinct species employ different cognitive maps with implications for performance efficiency [134,135].

More complex forms of cognition

Humans possess a remarkable mental ability that extends well beyond spatial cognition, including an ability to reason, to flexibly deploy language and to generalise experience to novel contexts [136,137]. Understanding the neural code of these highly flexible forms of human cognition is of great interest in many related fields, including both neuroscience and artificial intelligence [75]. Under a "representation rich" approach, it is now possible to probe the internal computations of those complex cognitive process. For example, using MEG-based decoding, Al Roumi, et al. [138] studied how sequences, and operations on sequences, are represented in the brain, finding evidence for an abstract, language-like code (or primitives) for flexible sequence representation. Similarly, Liu, et al. [23] demonstrated a "factorised representation" (with independent representation of abstract structural knowledge and concrete sensory information) in human replay, likely to be useful for inference and generalization in novel contexts [75,139,140]. Undoubtedly there are other organization principles of neural information for supporting flexible behaviour [75,141]. For example, the dynamics of semantic representations during visual understanding task might reflect a unique human ability to reason about arbitrary novel problems [142].

Finally, when studying abstract psychological processes, it may turn out that there is no apparent decodable content. While decoding mental states, such as emotional states, is possible in principle [143,144], that other, more tractable, approaches including experience sampling[145,146] have been used. To gain a complete understanding of the rich dynamics of spontaneous neural activity, it is important that we develop new methods that are suited to probe these abstract cognitive processes in the future.

Conclusion

Recent advances in decoding cognition from spontaneous neural activity provides a basis for grounding human cognitive studies that are beyond immediate task demand. We suggest a "representation rich" approach, that relies on a cognitive map of task space, can advance our understanding of a wide range of cognitive processes extending beyond task-evoked response. These processes include memory retrieval, planning and inference, which lie at the heart of sophisticated model-based reasoning. In considering both off-task and on-task neural reactivation we outline how this approach can help bridge a divide between studies of resting states and those that focus on task-evoked activity. Finally, we consider that linking physiological features of neural activity to its representational content will have profound implications for future research in psychiatry, particularly in light of recent findings in schizophrenia.

Glossary

Task - experiments that designed to manipulate assumed cognitive process.

Off-task - period without explicit task demand, e.g., during rest.

functional connectivity: temporal dependency of neuronal activation (e.g., correlation) between anatomically separated brain regions.

Targeted memory reactivation (TMR): cue-event associations were first learnt during task state, then the cue (typically a sound) is introduced during resting state, e.g., sleep, to elicit reactivation of the "target", i.e., the associated event.

Decoding: read out task-related information from neural activity

Representation rich: grounding cognitive research in spontaneous neural activity at the representation level, by making explicit reference to task related event without intervention.

Sharp-wave ripple: a short (around 200 ms), bursty high frequency (approx. 200 Hz in rodents) oscillatory event, i.e., ripple, which is typically associated with synchronous spiking, reflected as a strong deflection, i.e., sharp wave

Discriminative model: Formally, a model of the conditional probability of the target Y, given an observation X, i.e., P (Y | X). It makes minimal assumptions about the underlying distribution of the neural data itself.

Generative model: In contrast to discriminative models, generative models rely on an explicit modelling of the underlying distribution of the neural data, conditional on the task label, i.e., P(X|Y).

Ground truth: information that is known to be true.

Autocorrelation: correlation of a signal (e.g., X, a time series of decoded state) with a delayed copy of itself, e.g., $X(\Delta t)$.

Cognitive map: a mental representation of space that describes the relationship among events. This space can be either physical or non-physical.

Charles Bonnet syndrome: a condition where visual hallucinations occur as a result of vision loss.

Box 1: Neural sequences in rodents

Broadly speaking there are two types of neural sequences that are the subject of intense investigations. One relates to sharp-wave ripples (SWR) sequences, and the other relates to theta sequences [104,147].

SWR sequences refer to sequences of hippocampal cell firing embedded within sharp wave ripple epochs (~200 Hz LFP oscillatory bursts). These events typically happen during rest [27,31,33], but also during pauses in a behavioural task [148]. These are commonly referred to as "replay". The direction of replay is normally defined with respect to actual experience, with forward replay, the order of the pattern activity can occur in the same order as in the actual experience and backward replay the reverse order of experience. Both forward and backward replays are reported in rodent literature, and have been shown to be modulated by task demands[148]. Although the exact function of replay direction is still unknown, forward replay has been more associated with planning [78] (but cf. Gillespie, et al. [149]), and backward replay associated with learning, e.g., propagating prediction error from reward site [30,37]. Accumulating evidence suggests that SWR sequences (or replay) are important for mental functions as diverse as memory, learning and decision-making [43,44].

Theta sequences are sequences of hippocampal place cells firing within a single theta cycle (~100 – 170 ms wavelength), generally proceeding from the location of the animal forward towards potential goals. Key here is the observation that during movement (e.g., running through a maze) and pausing (e.g., at decision point), place cell firing is organized within an oscillatory process reflected in a hippocampal theta rhythm (6–10 Hz). Theta sequences are dominantly forward although reverse theta sequences are also reported [103]. Neurophysiologically, theta sequences are typically associated with theta phase precession during spatial navigation (cf. Chadwick, et al. [150] for independent theta phase coding), in which the firings of a particular place cell are embedded within progressively earlier portions of the theta cycle (phase precession) as the animal traverses an associated place field [151]. Functionally, theta sequences may reflect planning process [105,152,153], memory formation [104,147], or possibly represent multiple prospective futures in alteration [153,154].

Box 2: Reinforcement learning in human neuroscience

Reinforcement Learning (RL) is concerned with a specific family of questions: how to make decisions to maximise an expected future (discounted) cumulative reward (or avoidance of punishment); how to update or adjust behavior on the basis of a discrepancy between expectation and experienced outcome (i.e., prediction error), etc [46]. In neuroscience, reinforcement learning is widely linked to specific neural mechanisms, particularly phasic dopamine signalling in mesolimbic circuits reflecting reward prediction errors [155].

Based on whether RL relies on a mental representation of task space (that is, relational structures among task states), RL is conventionally divided into model-free and model-based processes[74,156,157] (See ref 157 for a more complete view). Model-free RL proceeds via trial-and-error learning and relies on consolidating stimulus-response mappings. The best known model is Rescorla–Wagner (RW) [158], developed in the context of classical conditioning [159]. Although RW explains many psychological phenomena, and continues to provide remarkable insights into human learning and decision-making [160], it does not readily address more flexible forms of cognition, such as those concerned with sequential decision making or computations that go beyond directly-experienced stimulus→response associations (for example, planning detours or considering counterfactuals), both of which necessities reliance on an internal 'model' of the task (that is, model-based).

A classical paradigm developed to study model-based RL is "two-step" task [156,161-163]. In its original formulation by Daw, et al. [162], this involved 2-stage sequential decision-making steps, where only the second stage choice results in a monetary outcome. The state transition structure between a first and second stage is designed to yield different patterns of choice behaviour in model-free vs. model-based agents (where the former has no internal 'model' of the transition structure and makes choices based on cached stimulus–action values). Such tasks may be used to characterise the extent of model-based computation in decision-making, according to the degree to which they make use of an internal model of the environment. When combined with the 'representation-rich' approach, tasks of this nature yield new insights into the intrinsic neural mechanisms supporting model-based cognition [24,55]. Note, in these neuroimaging studies, transition structure in the two-step like task can also be deterministic [55], or even simplified to a one-step decision [24], to ease the use of decoding techniques.

Acknowledgments

The authors thank Elliott Wimmer, Cameron Higgins and Philipp Schwartenbeck for helpful discussions, and fruitful collaborations on the work and ideas presented in this paper. This work is supported by the Fundamental Research Funds for the Central Universities to Y.L.; a Wellcome Trust Investigator Award (098362/Z/12/Z) to R.J.D; a UCL Welcome PhD Fellowship for Clinicians (102186/B/13/Z) to M.M.N; a Wellcome Trust Senior Research Fellowship (104765/Z/14/Z), and a Principal Research Fellowship (219525/Z/19/Z), together with a James S. McDonnell Foundation Award (JSMF220020372) to T.B; an Independent Research Group Grant from the Max

Planck Society (M.TN.A.BILD0004) and a Starting Grant from the European Union (ERC-2019-StG REPLAY-852669) to N.W.S.

M.M.N. is a pre-doctoral fellow of the International Max Planck Research School on Computational Methods in Psychiatry and Ageing Research (https://www.mps-ucl-centre.mpg.de/en/comp2psych. Participating institutions: Max Planck Institute for Human Development, Berlin & UCL). The Max Planck UCL Centre is supported by UCL and the Max Planck Society. The Wellcome Centre for Human Neuroimaging (WCHN) is supported by core funding from the Wellcome Trust (203147/Z/16/Z). The Wellcome Centre for Integrative Neuroimaging (WIN) is supported by core funding from the Wellcome Trust (203139/Z/16/Z).

Competing interests statement

There is no compete of interests.

References

1 Uddin, L. Q. Bring the noise: reconceptualizing spontaneous neural activity. *Trends in Cognitive Sciences* (2020).

2 Zhang, D. & Raichle, M. E. Disease and the brain's dark energy. *Nature Reviews Neurology* **6**, 15-28 (2010).

3 Becker, R., Van De Ville, D. & Kleinschmidt, A. Alpha oscillations reduce temporal long-range dependence in spontaneous human brain activity. *Journal of Neuroscience* **38**, 755-764 (2018).

4 Allaman, L., Mottaz, A., Kleinschmidt, A. & Guggisberg, A. G. Spontaneous network coupling enables efficient task performance without local task-induced activations. *Journal of Neuroscience* **40**, 9663-9675 (2020).

5 Smith, S. M. *et al.* Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences* **17**, 666-682 (2013).

6 Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences* **106**, 13040-13045 (2009).

7 Tavor, I. *et al.* Task-free MRI predicts individual differences in brain activity during task performance. *Science* **352**, 216-220 (2016).

8 Rudoy, J. D., Voss, J. L., Westerberg, C. E. & Paller, K. A. Strengthening individual memories by reactivating them during sleep. *Science* **326**, 1079-1079 (2009).

9 Rasch, B., Büchel, C., Gais, S. & Born, J. Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science* **315**, 1426-1429 (2007).

10 Wang, B. *et al.* Targeted memory reactivation during sleep elicits neural signals related to learning content. *Journal of Neuroscience* **39**, 6728-6736 (2019).

11 Cairney, S. A., El Marj, N. & Staresina, B. P. Memory consolidation is linked to spindle-mediated information processing during sleep. *Current Biology* **28**, 948-954. e944 (2018).

12 Pajani, A., Kok, P., Kouider, S. & de Lange, F. P. Spontaneous activity patterns in primary visual cortex predispose to visual hallucinations. *Journal of Neuroscience* **35**, 12947-12953 (2015).
13 Chew, B. *et al.* Endogenous fluctuations in the dopaminergic midbrain drive behavioral choice variability. *Proceedings of the National Academy of Sciences* **116**, 18732-18737 (2019).
14 Tambini, A. & Davachi, L. Awake Reactivation of Prior Experiences Consolidates Memories and Biases Cognition. *Trends in Cognitive Sciences* (2019).
15 Higgins, C. *et al.* Replay bursts in humans coincide with activation of the default mode and parietal alpha networks. *Neuron* **109**, 882-893. e887 (2021).
16 Yeshurun, Y., Nguyen, M. & Hasson, U. The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, 1-12 (2021).
17 Sutherland, G. R. & McNaughton, B. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology* **10**, 180-186 (2000).
18 Tambini, A. & Davachi, L. Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proceedings of the National Academy of Sciences* **110**, 19591-19596 (2013).
19 Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* **2**, 4 (2008).
20 Diedrichsen, J. & Kriegeskorte, N. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS computational biology* **13**, e1005508 (2017).
21 Eldar, E., Lièvre, G., Dayan, P. & Dolan, R. J. The roles of online and offline replay in planning. *eLife* **9**, e56911 (2020).
22 Kurth-Nelson, Z., Economides, M., Dolan, Raymond J. & Dayan, P. Fast Sequences of Non-spatial State Representations in Humans. *Neuron* **91**, 194-204 (2016).
23 Liu, Y., Dolan, R. J., Kurth-Nelson, Z. & Behrens, T. E. J. Human replay spontaneously reorganizes experience. *Cell* **178**, 640-652 (2019).
24 Liu, Y., Mattar, M. G., Behrens, T. E., Daw, N. D. & Dolan, R. J. Experience replay is associated with efficient nonlocal learning. *Science* **372** (2021).
25 Schuck, N. W. & Niv, Y. Sequential replay of nonspatial task states in the human hippocampus. *Science* **364**, eaaw5181 (2019).
26 Burgess, N. & O'Keefe, J. Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* **6**, 749-762 (1996).
27 Skaggs, W. E. & McNaughton, B. L. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271**, 1870-1873 (1996).
28 Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J. & Buzsáki, G. Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience* **19**, 9497-9507 (1999).
29 Louie, K. & Wilson, M. A. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* **29**, 145-156 (2001).
30 Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680 (2006).

31 Diba, K. & Buzsáki, G. Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience* **10**, 1241 (2007).
32 Ji, D. & Wilson, M. A. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience* **10**, 100 (2007).
33 Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63**, 497-507 (2009).
34 Karlsson, M. P. & Frank, L. M. Awake replay of remote experiences in the hippocampus. *Nature Neuroscience* **12**, 913 (2009).
35 Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695-705 (2010).
36 Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience* **14**, 147 (2011).
37 Ambrose, R. E., Pfeiffer, B. E. & Foster, D. J. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron* **91**, 1124-1136 (2016).
38 Ólafsdóttir, H. F., Carpenter, F. & Barry, C. Coordinated grid and place cell replay during rest. *Nature Neuroscience* **19**, 792 (2016).
39 O'Neill, J., Boccara, C. N., Stella, F., Schoenenberger, P. & Csicsvari, J. Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science* **355**, 184-188 (2017).
40 Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nature Neuroscience* **20**, 571-580 (2017).
41 Buzsáki, G. Hippocampal sharp wave‐ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073-1188 (2015).
42 Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454-1458 (2012).
43 Foster, D. J. Replay comes of age. *Annual Review of Neuroscience* **40**, 581-602 (2017).
44 Ólafsdóttir, H. F., Bush, D. & Barry, C. The Role of Hippocampal Replay in Memory and Planning. *Current Biology* **28**, R37-R50 (2018).
45 Wittkuhn, L. & Schuck, N. W. Dynamics of fMRI patterns reflect sub-second activation sequences and reveal replay in human visual cortex. *Nature Communications* **12**, 1-22 (2021).
46 Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. (MIT press, 2018).
47 Sutton, R. S. Dyna, an integrated architecture for learning, planning, and reacting. *ACM Sigart Bulletin* **2**, 160-163 (1991).
48 Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience* **21**, 1609 (2018).
49 Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage* **173**, 434-447 (2018).
50 Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences* **10**, 424-430 (2006).
51 Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425-2430 (2001).
52 Peelen, M. V. & Downing, P. E. Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends in Cognitive Sciences* **11**, 4-4 (2007).

53 Liu, Y. *et al.* Temporally delayed linear modelling (TDLM) measures replay in both animals and humans. *eLife* **10**, e66917 (2021).
54 Harris, K. D. Nonsense correlations in neuroscience. *bioRxiv* (2020).
55 Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D. & Daw, N. D. Model-based choices involve prospective neural activity. *Nature Neuroscience* **18**, 767 (2015).
56 Eldar, E., Bae, G. J., Kurth-Nelson, Z., Dayan, P. & Dolan, R. J. Magnetoencephalography decoding reveals structural differences within integrative decision processes. *Nature Human Behaviour* **2**, 670-681 (2018).
57 Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R. & Dayan, P. Temporal structure in associative retrieval. *eLife* **4**, e04919 (2015).
58 Momennejad, I., Otto, A. R., Daw, N. D. & Norman, K. A. Offline replay supports planning in human reinforcement learning. *eLife* **7**, e32548 (2018).
59 Wimmer, G. E., Liu, Y., Vehar, N., Behrens, T. E. J. & Dolan, R. J. Episodic memory retrieval success is associated with rapid replay of episode content. *Nature Neuroscience* **23**, 1025–1033 (2020).
60 Wimmer, G. E. & Shohamy, D. Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science* **338**, 270-273 (2012).
61 Wise, T., Liu, Y., Chowdhury, F. & Dolan, R. J. Model-based aversive learning in humans is supported by preferential task state reactivation. *Science Advances* **7**, eabf9616 (2021).
62 Wittkuhn, L., Chien, S., Hall-McMaster, S. & Schuck, N. W. Replay in minds and machines. *Neuroscience & Biobehavioral Reviews* **129**, 367-388 (2021).
63 Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C. & Norman, K. A. Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications* **9**, 1-11 (2018).
64 Genzel, L. *et al.* A consensus statement: defining terms for reactivation analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375** (2020).
65 Tambini, A., Ketz, N. & Davachi, L. Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron* **65**, 280-290 (2010).
66 Schönauer, M. *et al.* Decoding material-specific memory reprocessing during sleep in humans. *Nature Communications* **8**, 1-9 (2017).
67 Shanahan, L. K., Gjorgieva, E., Paller, K. A., Kahnt, T. & Gottfried, J. A. Odor-evoked category reactivation in human ventromedial prefrontal cortex during sleep promotes memory consolidation. *eLife* **7**, e39681 (2018).
68 Deuker, L. *et al.* Memory consolidation by replay of stimulus-specific neural activity. *Journal of Neuroscience* **33**, 19373-19383 (2013).
69 Antony, J. W. & Schapiro, A. C. Active and effective replay: systems consolidation reconsidered again. *Nature Reviews Neuroscience* **20**, 506-507 (2019).
70 Schuck, Nicolas W., Cai, Ming B., Wilson, Robert C. & Niv, Y. Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* **91**, 1402-1412 (2016).
71 Kornysheva, K. *et al.* Neural competitive queuing of ordinal structure underlies skilled sequential action. *Neuron* (2019).
72 Hahamy, A., Wilf, M., Rosin, B., Behrmann, M. & Malach, R. How do the blind 'see'? The role of spontaneous brain activity in self-generated perception. *Brain* **144**, 340-353 (2021).
73 Shohamy, D. & Daw, N. D. Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences* **5**, 85-90 (2015).

74 Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312-325 (2013).
75 Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490-509 (2018).
76 Vaz, A. P., Wittig, J. H., Inati, S. K. & Zaghloul, K. A. Replay of cortical spiking sequences during human memory retrieval. *Science* **367**, 1131-1134 (2020).
77 Norman, Y. *et al.* Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science* **365** (2019).
78 Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D. & Spiers, H. J. Hippocampal place cells construct reward related sequences through unexplored space. *eLife* **4**, e06063 (2015).
79 Moneta, N., Garvert, M. M., Heekeren, H. R. & Schuck, N. W. Parallel representation of context and multiple context-dependent values in ventro-medial prefrontal cortex. *bioRxiv* (2021).
80 Wang, F., Schoenbaum, G. & Kahnt, T. Interactions between human orbitofrontal cortex and hippocampus support model-based inference. *PLoS Biology* **18**, e3000578 (2020).
81 Schuck, N. W. *et al.* Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron* **86**, 331-340 (2015).
82 Woodward, N. D. & Cascio, C. J. Resting-state functional connectivity in psychiatric disorders. *JAMA psychiatry* **72**, 743-744 (2015).
83 Snyder, A. Z. & Raichle, M. E. A brief history of the resting state: the Washington University perspective. *Neuroimage* **62**, 902-910 (2012).
84 Smith, S. M. *et al.* Resting-state fMRI in the human connectome project. *Neuroimage* **80**, 144-168 (2013).
85 Raichle, M. E. & Snyder, A. Z. A default mode of brain function: a brief history of an evolving idea. *Neuroimage* **37**, 1083-1090 (2007).
86 Gusnard, D. A. & Raichle, M. E. Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews Neuroscience* **2**, 685-694 (2001).
87 Raichle, M. E. The brain's default mode network. *Annual Review of Neuroscience* **38**, 433-447 (2015).
88 Buckner, R. L. & DiNicola, L. M. The brain's default network: updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience* **20**, 593-608 (2019).
89 Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R. & Buckner, R. L. Functional-anatomic fractionation of the brain's default network. *Neuron* **65**, 550-562 (2010).
90 Raichle, M. E. *et al.* A default mode of brain function. *Proceedings of the National Academy of Sciences* **98**, 676-682 (2001).
91 Buckner, R. L. The serendipitous discovery of the brain's default network. *Neuroimage* **62**, 1137-1145 (2012).
92 Agnati, L. F., Guidolin, D., Battistin, L., Pagnoni, G. & Fuxe, K. The neurobiology of imagination: possible role of interaction-dominant dynamics and default mode network. *Frontiers in Psychology* **4**, 296 (2013).
93 Smallwood, J. & Schooler, J. W. The science of mind wandering: empirically navigating the stream of consciousness. *Annual Review of Psychology* **66**, 487-518 (2015).
94 Schacter, D. L., Addis, D. R. & Buckner, R. L. Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience* **8**, 657-661 (2007).
95 Hassabis, D. & Maguire, E. A. Deconstructing episodic memory with construction. *Trends in Cognitive Sciences* **11**, 299-306 (2007).

96 Meyer, M. L., Davachi, L., Ochsner, K. N. & Lieberman, M. D. Evidence that default network connectivity during rest consolidates social information. *Cerebral Cortex* **29**, 1910-1920 (2019).
97 Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464 (2016).
98 Park, S. A., Miller, D. S. & Boorman, E. D. Novel Inferences in a Multidimensional Social Network Use a Grid-like Code. *bioRxiv* (2020).
99 Baldassano, C., Hasson, U. & Norman, K. A. Representation of real-world event schemas during narrative perception. *Journal of Neuroscience* **38**, 9689-9699 (2018).
100 Vidaurre, D., Myers, N. E., Stokes, M., Nobre, A. C. & Woolrich, M. W. Temporally unconstrained decoding reveals consistent but time-varying stages of stimulus processing. *Cerebral Cortex* **29**, 863-874 (2019).
101 Kaplan, R. *et al.* Hippocampal sharp-wave ripples influence selective activation of the default mode network. *Current Biology* **26**, 686-691 (2016).
102 Liu, X. *et al.* Multimodal neural recordings with Neuro-FITM uncover diverse patterns of cortical–hippocampal interactions. Report No. 1546-1726, (Nature Publishing Group, 2021).
103 Wang, M., Foster, D. J. & Pfeiffer, B. E. Alternating sequences of future and past behavior encoded within hippocampal theta oscillations. *Science* **370**, 247, doi:10.1126/science.abb4151 (2020).
104 Buzsáki, G. & Moser, E. I. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience* **16**, 130 (2013).
105 Redish, A. D. Vicarious trial and error. *Nature Reviews Neuroscience* **17**, 147 (2016).
106 Sylvester, C. M. *et al.* Individual-specific functional connectivity of the amygdala: A substrate for precision psychiatry. *Proceedings of the National Academy of Sciences* **117**, 3808-3818 (2020).
107 McCutcheon, R. A., Marques, T. R. & Howes, O. D. Schizophrenia—an overview. *JAMA psychiatry* **77**, 201-210 (2020).
108 Suh, J., Foster, D. J., Davoudi, H., Wilson, M. A. & Tonegawa, S. Impaired hippocampal ripple-associated replay in a mouse model of schizophrenia. *Neuron* **80**, 484-493 (2013).
109 Altimus, C., Harrold, J., Jaaro-Peled, H., Sawa, A. & Foster, D. J. Disordered ripples are a common feature of genetically distinct mouse models relevant to schizophrenia. *Molecular neuropsychiatry* **1**, 52-59 (2015).
110 Zeidman, P. & Maguire, E. A. Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nature Reviews Neuroscience* **17**, 173-182 (2016).
111 Adams, R. A. *et al.* Impaired theta phase coupling underlies frontotemporal dysconnectivity in schizophrenia. *Brain* **143**, 1261-1277 (2020).
112 Titone, D., Ditman, T., Holzman, P. S., Eichenbaum, H. & Levy, D. L. Transitive inference in schizophrenia: impairments in relational memory organization. *Schizophrenia research* **68**, 235-247 (2004).
113 Nour, M. M., Liu, Y., Arumuham, A., Kurth-Nelson, Z. & Dolan, R. Impaired neural replay of inferred relationships in schizophrenia. *Cell* **184**, 4315-4328 (2021).
114 Whitfield-Gabrieli, S. & Ford, J. M. Default mode network activity and connectivity in psychopathology. *Annual review of clinical psychology* **8**, 49-76 (2012).
115 Huys, Q. J. *et al.* Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences* **112**, 3098-3103 (2015).

116 Huys, Q. J. *et al.* Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLOS Computational Biology* **8**, e1002410 (2012).
117 Heller, A. S. & Bagot, R. C. Is Hippocampal Replay a Mechanism for Anxiety and Depression? *JAMA psychiatry* **77**, 431-432 (2020).
118 Lewis, P. A., Knoblich, G. & Poe, G. How Memory Replay in Sleep Boosts Creative Problem-Solving. *Trends in Cognitive Sciences* **22**, 491-503 (2018).
119 Klinzing, J. G., Niethard, N. & Born, J. Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience* **22**, 1598-1610 (2019).
120 Wei, Y., Krishnan, G. P., Marshall, L., Martinetz, T. & Bazhenov, M. Stimulation Augments Spike Sequence Replay and Memory Consolidation during Slow-Wave Sleep. *Journal of Neuroscience* **40**, 811-824 (2020).
121 Tamminen, J., Lambon Ralph, M. A. & Lewis, P. A. The Role of Sleep Spindles and Slow-Wave Activity in Integrating New Information in Semantic Memory. *Journal of Neuroscience* **33**, 15376-15381 (2013).
122 Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676-679 (1994).
123 Pfeiffer, B. E. The content of hippocampal "replay". *Hippocampus* **30**, 6-18 (2020).
124 Zielinski, M. C., Shin, J. D. & Jadhav, S. P. Hippocampal theta sequences in REM sleep during spatial learning. *bioRxiv* (2021).
125 McNamee, D. C., Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. Flexible modulation of sequence generation in the entorhinal–hippocampal system. *Nature Neuroscience*, 1-12 (2021).
126 Krause, E. L. & Drugowitsch, J. A large majority of awake hippocampal sharp-wave ripples feature spatial trajectories with momentum. *Neuron*, doi:10.1016/j.neuron.2021.11.014.
127 Stella, F., Baracskay, P., O'Neill, J. & Csicsvari, J. Hippocampal reactivation of random trajectories resembling Brownian diffusion. *Neuron* **102**, 450-461. e457 (2019).
128 Pereira, S. I. R. & Lewis, P. A. Sleeping through brain excitation and inhibition. *Nature Neuroscience* **23**, 1037-1039 (2020).
129 Belal, S. *et al.* Identification of memory reactivation during sleep by EEG classification. *NeuroImage* **176**, 203-214 (2018).
130 Higgins, C. *Uncovering temporal structure in neural data with statistical machine learning models*, University of Oxford, (2019).
131 Kriegeskorte, N. *et al.* Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126-1141 (2008).
132 Barron, H. C., Mars, R. B., Dupret, D., Lerch, J. P. & Sampaio-Baptista, C. Cross-species neuroscience: closing the explanatory gap. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20190633 (2021).
133 Barron, H. C. *et al.* Neuronal computation underlying inferential reasoning in humans and mice. *Cell* **183**, 228-243. e221 (2020).
134 Akam, T., Costa, R. & Dayan, P. Simple Plans or Sophisticated Habits? State, Transition and Learning Interactions in the Two-Step Task. *PLOS Computational Biology* **11**, e1004648 (2015).
135 Miranda, B., Malalasekera, W. M. N., Behrens, T. E., Dayan, P. & Kennerley, S. W. Combined model-free and model-sensitive reinforcement learning in non-human primates. *PLOS Computational Biology* **16** (2020).

136 Mack, M. L., Preston, A. R. & Love, B. C. Ventromedial prefrontal cortex compression during concept learning. *Nature Communications* **11**, 1-11 (2020).
137 Morton, N. W., Schlichting, M. L. & Preston, A. R. Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proceedings of the National Academy of Sciences* **117**, 29338-29345 (2020).
138 Al Roumi, F., Marti, S., Wang, L., Amalric, M. & Dehaene, S. Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. *Neuron* **109**, 2627-2639. e2624 (2021).
139 Bernardi, S. *et al.* The geometry of abstraction in hippocampus and prefrontal cortex. *bioRxiv* (2018).
140 Higgins, I. *et al.* Darla: Improving zero-shot transfer in reinforcement learning. *arXiv* (2017).
141 Whittington, J. C. *et al.* The Tolman-Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell* **183**, 1249-1263. e1223 (2020).
142 Schwartenbeck, P. *et al.* Generative replay for compositional visual understanding in the prefrontal-hippocampal circuit. *bioRxiv* (2021).
143 Kragel, P. A., Knodt, A. R., Hariri, A. R. & LaBar, K. S. Decoding spontaneous emotional states in the human brain. *PLoS biology* **14**, e2000106 (2016).
144 Tusche, A., Smallwood, J., Bernhardt, B. C. & Singer, T. Classifying the wandering mind: revealing the affective content of thoughts during task-free rest periods. *Neuroimage* **97**, 107-116 (2014).
145 Van Calster, L., D'Argembeau, A., Salmon, E., Peters, F. & Majerus, S. Fluctuations of attentional networks and default mode network during the resting state reflect variations in cognitive states: evidence from a novel resting-state experience sampling method. *Journal of Cognitive Neuroscience* **29**, 95-113 (2017).
146 Smallwood, J. *et al.* The neural correlates of ongoing conscious thought. *Iscience*, 102132 (2021).
147 Colgin, L. L. Rhythms of the hippocampal network. *Nature Reviews Neuroscience* **17**, 239 (2016).
148 Ólafsdóttir, H. F., Carpenter, F. & Barry, C. Task Demands Predict a Dynamic Switch in the Content of Awake Hippocampal Replay. *Neuron* **96**, 925-935.e926 (2017).
149 Gillespie, A. K. *et al.* Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *bioRxiv* (2021).
150 Chadwick, A., van Rossum, M. C. & Nolan, M. F. Independent theta phase coding accounts for CA1 population sequences and enables flexible remapping. *eLife* **4**, e03542 (2015).
151 Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149-172 (1996).
152 Wikenheiser, A. M. & Redish, A. D. Hippocampal theta sequences reflect current goals. *Nature Neuroscience* **18**, 289-294 (2015).
153 Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience* **27**, 12176-12189 (2007).
154 Kay, K. *et al.* Constant sub-second cycling between representations of possible futures in the hippocampus. *Cell* **180**, 552-567. e525 (2020).

155 Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593-1599 (1997).
156 Doll, B. B., Simon, D. A. & Daw, N. D. The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology* **22**, 1075-1081 (2012).
157 Daw, N. D. & Dayan, P. The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130478 (2014).
158 Siegel, S. & Allan, L. G. The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review* **3**, 314-321 (1996).
159 Sutton, R. S. & Barto, A. G. in *Proceedings of the ninth annual conference of the cognitive science society.* 355-378 (Seattle, WA).
160 Gallistel, C. R., LoLordo, V. M., Rozin, P. & Seligman, M. E. P. Robert A. Rescorla (1940–2020). *American Psychologist* **76**, 391-392 (2021).
161 Vikbladh, O. M. *et al.* Hippocampal contributions to model-based planning and spatial memory. *Neuron* **102**, 683-693. e684 (2019).
162 Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204-1215 (2011).
163 Kool, W., Cushman, F. A. & Gershman, S. J. When Does Model-Based Control Pay Off? *PLOS Computational Biology* **12**, e1005090 (2016).