

## Opinion piece



**Cite this article:** van der Meer MAA, Kemere C, Diba K. 2020 Progress and issues in second-order analysis of hippocampal replay. *Phil. Trans. R. Soc. B* **375**: 20190238. <http://dx.doi.org/10.1098/rstb.2019.0238>

Accepted: 14 January 2020

One contribution of 18 to a Theo Murphy meeting issue 'Memory reactivation: replaying events past, present and future'.

### Subject Areas:

neuroscience

### Keywords:

replay, reactivation, sharp-wave ripple, decoding, sequence analysis

### Authors for correspondence:

Matthijs A. A. van der Meer  
e-mail: mvdm@dartmouth.edu  
Caleb Kemere  
e-mail: caleb.kemere@rice.edu  
Kamran Diba  
e-mail: kdiba@umich.edu

<sup>†</sup>All authors contributed equally.

# Progress and issues in second-order analysis of hippocampal replay

Matthijs A. A. van der Meer<sup>1,†</sup>, Caleb Kemere<sup>2,†</sup> and Kamran Diba<sup>3,4,†</sup>

<sup>1</sup>Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

<sup>2</sup>Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA

<sup>3</sup>Neuroscience Graduate Program, and <sup>4</sup>Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

**id** MAAvdM, 0000-0002-2206-4473; KD, 0000-0001-5128-4478

Patterns of neural activity that occur spontaneously during sharp-wave ripple (SWR) events in the hippocampus are thought to play an important role in memory formation, consolidation and retrieval. Typical studies examining the content of SWRs seek to determine whether the identity and/or temporal order of cell firing is different from chance. Such 'first-order' analyses are focused on a single time point and template (map), and have been used to show, for instance, the existence of preplay. The major methodological challenge in first-order analyses is the construction and interpretation of different chance distributions. By contrast, 'second-order' analyses involve a comparison of SWR content between different time points, and/or between different templates. Typical second-order questions include tests of experience-dependence (replay) that compare SWR content before and after experience, and comparisons or replay between different arms of a maze. Such questions entail additional methodological challenges that can lead to biases in results and associated interpretations. We provide an inventory of analysis challenges for second-order questions about SWR content, and suggest ways of preventing, identifying and addressing possible analysis biases. Given evolving interest in understanding SWR content in more complex experimental scenarios and across different time scales, we expect these issues to become increasingly pervasive.

This article is part of the Theo Murphy meeting issue 'Memory reactivation: replaying events past, present and future'.

## 1. Introduction

The hippocampus spontaneously generates spike sequences whose firing order corresponds to the order observed during behaviour. These ordered spike sequences occur during specific time windows identified by sharp waves in the stratum radiatum of the CA1 region, along with fast ripple oscillations in the CA1 pyramidal layer (hence known as sharp-wave ripples, SWRs; [1]). SWRs and their temporally ordered activity are not only a strikingly beautiful phenomenon, but also provide an important systems-level neural access point into understanding higher-order cognitive and mnemonic processes such as memory encoding, consolidation and planning. Experimental studies have shown impairments in learning and performance of various memory tasks when SWRs are disrupted [2–5]. In parallel, computational models illustrate how SWRs may contribute to the learning and performance of such tasks [6–8], motivating further work that increasingly relies on identifying SWR *content* under different conditions. By SWR *content* we mean not simply aggregate properties such as the number or duration of SWRs, but 'what is being replayed': the structured spiking patterns during SWRs, such as the activation of specific ensembles and temporal orderings. These patterns have been associated with particular events, trajectories and experiences, motivating a growing body of work that seeks to identify and decode SWR content.

Early studies of SWR content were mostly concerned with demonstrating the statistical robustness of particular kinds of non-randomness in SWR activity

(see [9] for an excellent review). Although this seemingly straightforward issue is far from trivial (as we will discuss below; see also [10]), it is now established beyond doubt that SWR activity is structured in ways that deviate robustly from chance. However, the status quo is largely facing different questions: to what extent is SWR activity shaped by experience? Does SWR activity transform or prioritize particular patterns to serve specific cognitive or network benefits, e.g. by preferentially replaying salient experiences, trajectories less or more travelled, and/or conversely forgetting or suppressing other experiences? Does the presence of some specific set of factors dramatically change SWR content? These questions come with additional complexity in the data analysis, in that it is no longer sufficient to simply demonstrate that SWR activity is more structured relative to some notion of chance.

As experimental and theoretical interest in probing SWR content continues to evolve, we see a corresponding need for analysis methods that are appropriate for increasingly complex and subtle questions. The objective of this paper is to facilitate the analysis of SWR content by organizing questions about SWR content into a rudimentary taxonomy. A key feature of this taxonomy is the distinction between *first-order* questions, which rely on determining whether and how SWR content is different from chance, and *second-order* questions, which seek to establish whether SWR content differs between two or more conditions, such as time points or arms of a maze. Then, we will provide an inventory of challenges in data analysis specific to second-order questions, and point to some possible ways to diagnose, prevent and address them.

## 2. First-order versus second-order replay analysis

Figure 1a diagrams the distinction between first-order and second-order SWR content questions. First-order questions (cyan rectangles) are concerned with a single set of SWR events and a single *template* (depending on the question, a template may be a list of cells or ensembles, a set of tuning curves, or a specific ordering of cells); no comparison between multiple templates or between different sets of SWRs is required. First-order questions have sought to establish, for instance, that there is sequential order in SWR content similar to an environment that has yet to be explored ('preplay', [11–13]); that sequential activity can be in forward and reverse order compared to the place cell order experienced during behaviour [14,15]; and that SWR content can be of a remote environment (i.e. one that the animal is not currently in, [16]). These first-order content and order detection questions remain important, requiring careful consideration of assumptions built into the null hypotheses, particularly the shuffle/resampling methods, to determine whether the observed sequence activity is unexpected by chance. Such first-order issues have been recently discussed in several papers [10,13,17] and we will review them briefly in the next section.

Second-order questions include comparisons between different experimental conditions (red rectangles in figure 1a) and/or time points (purple rectangles). For example, when viewed this way, establishing whether SWR content reflects *replay* of prior experience is a second-order question that requires comparing SWR content at two different time points: prior to, and following experience [18]. Other typical second-order questions include asking whether the left or a

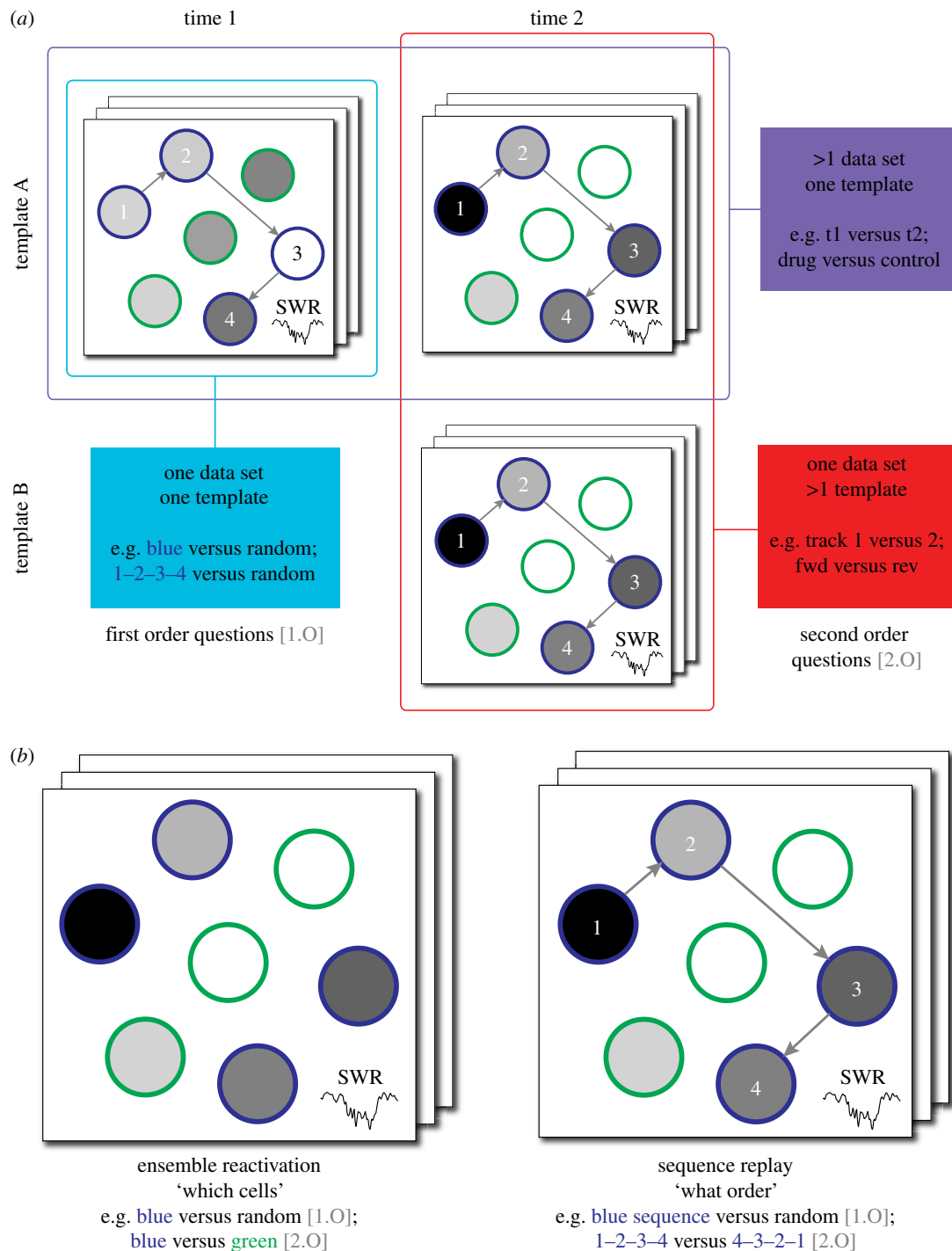
right arm of a T-maze is preferentially expressed during SWRs [19–21], whether there is an increase in replay in the presence or absence of reward [22,23], whether pharmacological and/or genetic manipulations affect SWR content [24,25], and so on. These second-order questions face additional scrutiny that requires accounting for an additional set of potential confounds. A major goal of this paper is to identify challenges in detecting and interpreting second-order patterns in replay and to suggest some practical advice.

## 3. First-order replay analysis

### (a) Which cells and what order?

The main question at stake in a first-order SWR analysis is the determination of whether activity in a given SWR (or set of SWRs, taken together) is different from a chance distribution. Activity may differ from chance in a number of ways, such as in *which cells* are active (for instance, particular pairs, ensembles, etc.), and in whether their activity displays *temporal ordering* (figure 1b). The choice of which chance distribution(s) to compare the observed data to (typically 'shuffles' in a resampling procedure) determines the interpretation and level of specificity of the conclusion that can be drawn from a first-order analysis. Note that in the literature, the term 'replay' does not have a consistent technical definition. Some studies consider the (re)activation of specific cells, pairs or ensembles without any specific temporal ordering as replay, whereas others use the term *reactivation* to distinguish it from temporally ordered (sequential) activity. In this review, we try to be as explicit as possible in distinguishing *ensemble reactivation*, which can occur without any temporal order but does not preclude it, and *sequence replay*, which has a temporal order, but may additionally involve activity in a specific ensemble. In places, however, we will use the term 'replay' nonspecifically to include either or both of these types of SWR structure.

This distinction between ensemble reactivation (which cells) and sequence replay (temporal order) illustrated in figure 1b is important from several interrelated but distinct perspectives: neural, psychological and analytical. Neurally, different kinds of structure in SWR activity will likely be read out differently depending on the downstream circuitry. Several studies have shown relationships between SWR content in the hippocampus and spiking activity in putative populations of readout neurons (ventral striatum: [26,27]; prefrontal cortex: [28–30]; entorhinal cortex: [31], but see [32]; lateral septum: [33]; auditory cortex: [34]; ventral tegmental area: [35]). However, these correspondences generally take the form of a general statistical model that does not reveal what specific features of hippocampal SWR activity are most important for downstream neurons. For instance, it is currently not known how accurately postsynaptic neurons distinguish between forward and reverse SWR sequences (but see [36] for a proposed mechanism for how single neurons may do so). Alternatively, some structures, such as the lateral septum [33], may only care about the strength and size of the SWR-associated population activity. In general, the problem of determining what features of SWRs matter is an important overall challenge in the replay field, and will need to be confronted with complementary lines of inquiry (see Box 1 for some promising directions). When considering different analysis issues and methods, such as those discussed in this review, the choice of analysis method should ultimately be informed by the features



**Figure 1.** Schematic illustration of two conceptual distinctions in replay analysis: first-order versus second-order questions about replay content (a) and ensemble reactivation versus sequence replay (b). (a) *First-order* questions (cyan boxes) test whether a given activity pattern (e.g. which cells, what order) is distinct from chance. First-order questions are concerned with a single set of SWR events and a single template. In contrast, *second-order* questions compare replay content between multiple data sets (e.g. time 1 and time 2, purple boxes), templates (e.g. track A and track B, red boxes) or both. Typical second-order questions include determining whether post-experience SWR content better resembles activity during behaviour compared to pre-experience content, and whether the left arm of a maze is more frequently (re)activated than the right arm. (b) Both first-order and second-order questions about SWR content can be categorized as focusing on *which cells* participate, while ignoring temporal order (*ensemble reactivation*, left), and/or as focusing on *what order* these cells are active (*sequence replay*, right). Typical ensemble reactivation questions include determining if a given pair or ensemble of cells is more co-active than expected by chance (a first-order question) or which of multiple possible ensembles is more active (e.g. ‘blue’ versus ‘green’ cells, a second-order question); typical sequence replay questions include determining if firing order is different from chance (a first-order question), and determining which of two firing orders is more prevalent (a second-order question). (Online version in colour.)

of SWR activity that are neurophysiologically or behaviourally meaningful.

A second reason to care about different kinds of structure in SWR activity is that each may be associated with distinct cognitive and/or mnemonic processes. For example, temporally structured sequence replay may be better suited to signalling multimodal, temporally organized episodic events for which

encoding and initial retrieval are supported by the hippocampus (though we note, that it is currently unclear whether the content of replay is, in fact, episodic; the relationship between SWRs and the subjective experience of retrieval or mental time travel remains to be established). Conversely, ensemble reactivation that is not explicitly ordered in time, but does involve a consistently co-active set of cells, may be

### Box 1. Ground truth for replay?

Spiking activity during sharp-wave ripples (SWRs) exhibits a diverse array of structured patterns, including the systematic co-activation of particular subsets of cells (pairs and ensembles, ‘ensemble reactivation’), and temporal ordering which matches that observed during behaviour (sequence replay). Analyses of SWR activity aim to determine whether such patterns are different from what could be expected by chance; however, they cannot, by themselves, decide which patterns are physiologically or behaviourally relevant. ‘Grounding’ SWR content is an important but difficult problem that may be pursued by, among others, the following approaches:

*Approach 1: determine what SWR content features are read out by downstream single neurons and brain structures.* For extra-hippocampal neurons that show a statistically reliable change in activity following SWRs, a number of analysis strategies may be used to determine that neuron’s preferred SWR pattern, or more generally, its tuning to SWR activity. Ideally, candidate readout neurons would be positively identified as receiving hippocampal input (e.g. with opto-tagging), and show temporal relationships with SWRs that reflect a genuine readout rather than merely correlation. Since SWR activity is potentially dynamic and high-dimensional, a range of appropriate dimensionality reduction techniques may need to be employed to characterize readout. Different neurons and brain structures may be ‘tuned’ to different SWR features, applying different projections or decision boundaries to the input. For instance, lateral septum neurons may respond preferentially to an ‘overall activity’ dimension [33], whereas ventral striatal neurons may respond preferentially to activity associated with reward [37].

*Approach 2: trigger disruption or stimulation on specific SWR content features and observe the neural/behavioural consequences.* Real-time detection of specific SWR features can be used for a variety of interventions that causally test the importance of such features. For instance, selectively disrupting forward or reverse replay may reveal distinct behavioural impairments associated with each. Recording in putative readout areas while artificially triggering SWRs, and inhibiting synaptic terminals in triggered on SWRs occurring, can provide complementary evidence that a given neuron or area, in fact, performs a readout.

appropriate for retrieval of specific cues, spatial contexts and semantic information that lacks a clear temporal component. Sequence replay and ensemble reactivation are in principle not mutually exclusive either: for instance, the phenomenon of remapping suggests that distinct contexts are associated with specific ensembles (maps) in which multiple distinct temporally organized experiences can occur [38–40]. Thus, a given SWR may involve both the reactivation of a specific ensemble (indicating a specific context) and a specific ordering of cells (indicating a specific trajectory or experience). Plausibly, the neural mechanisms underlying both these components are distinct to some degree, and certain experimental manipulations may have effects on only one, but not the other.

Finally, in terms of analysis, the distinction between the reactivation of cells and their temporal order is important because the appropriate analysis method depends on the phenomenon of interest. Cell reactivation (e.g. [22,41]), pairwise co-activation methods (e.g. [18,42–45]) and PCA/ICA-based ensemble reactivation studies (e.g. [46–48]) are agnostic towards temporal order. These methods also typically apply a fixed time window to measure co-activity in cells and ensembles, neglecting temporal compression across different brain states. While asymmetries in pairwise cross-correlograms can capture temporal relationships [49–51], as the template used to evaluate the order of firing is expanded to three [52] or more neurons [14,53,54], the methods become increasingly sensitive to the temporal structure. All of these approaches require careful consideration of the null hypothesis (shuffles) to compare the data to, and may give false positives based on factors such as firing rate differences (discussed in more detail below); but if the analysis is carried out appropriately then results can be interpreted cleanly in terms of either ‘which cells’ or ‘what order’ cause the non-random structure of the data.

How do current analysis methods detect reactivation and/or sequential replay? In this issue, Tingley & Peyrache [9] review the various methods used to detect replay in ensembles and sequences and the statistics associated with each. In the present review, we focus our attention particularly on sequence

replay methods that capture the order of firing across populations of neurons.

### (b) Sequence detection

There are several popular metrics for quantifying sequence replay: the rank-order (Spearman) correlation [14,15,55], the replay score (radon transform) introduced by Davidson *et al.* [56] and the linear weighted correlation [13,57,58]—though other methods continue to undergo development [59–61]. These metrics ostensibly capture sequential activity, but can also be affected by factors unrelated to temporal order. This is because for any of these metrics, the key question becomes: what is the appropriate null distribution against which we should compare the data? Each shuffling method is based on assumptions about what random data ought to look like. If the data deviates from these assumptions then the null hypothesis can be rejected. However, rejection of the null hypothesis may be due to aspects of these assumptions that are not directly related to the sequential firing of cells during replay. We will next highlight several instances of this general issue.

For example, the *spike-id shuffle* randomizes the cell identity for each spike in a candidate SWR event (e.g. in [62]), creating a surrogate dataset in which the overall firing patterns during the event are preserved, but each neuron is equally likely to fire any given spike from the onset to offset of this event. Therefore, this null hypothesis can potentially be rejected if the distribution of inter-spike intervals of neurons deviates from uniformity, as is known to be the case [63].

The *cell-id shuffle* (randomizing the cell identity for each train of spikes in a candidate event; used in [14,15]) further controls for the statistics of neurons’ observed firing patterns (e.g. burstiness, inhibition of return), but otherwise assumes that neurons fire independently of each other during ripple events. By randomizing the order of neurons, which also randomizes which neurons are co-firing in a given time window, this null hypothesis may be rejected by non-uniform co-activity, as well as non-random sequential activity. Thus, it



tests against ensemble reactivation and sequence replay simultaneously, which may or may not be desirable. Moreover, it does not readily lend itself to Bayesian decoding analysis because, similar to the spike-id shuffle, it does not preserve non-uniformity in the firing rates of neurons [64].

The *circular place-field shuffle* (circularly shifting each tuning curve by a different random amount; used by [58]) allows that neurons' spike-train statistics should be preserved but assumes that their preferred firing locations are randomly dispersed. Importantly, this shuffle does not assume continuity in place-fields; surrogate place-fields can represent locations that start at the end of a track and reprise on the other side of the track. Therefore, this null hypothesis can be rejected if there is an uneven distribution of place-fields across the track, or the shape of place-fields observes specific relationships to the maze, as has been reported in several studies [65–67].

Other surrogate methods first separate candidate events into distinct time bins prior to shuffling. After binning, the *circular column cycle shuffle* (used by [17,56]) shifts the decoded (posterior) positions in candidate bins by random amounts, thus controlling for the limited variance of decoded position in each bin, but assumes that all maze positions are equally likely to be reprised in any given time bin. Similar to the place-field shuffle, it also allows that decoded locations can be discontinuous across the end/beginning of tracks in the null distribution, which does not happen in real data. Therefore, this null hypothesis can be rejected if the decoded positions in candidate events are not uniformly distributed along the track and/or have specific skewed relationships to the end/start of tracks.

The *pooled time-swap shuffle* (also introduced by [56]) randomizes the decoded position across all the candidate replay events in the dataset. Thus, it assumes that the decoded position in a given bin is random, but maintains the overall distribution of decoded positions across all observed bins. The *within-event time-swap* (used by [13]) shuffles the decoded position bins in each candidate event separately (rather than across events), thus more conservatively preserving the overall statistics of decoded positions in each event. Yet, both of these time-swap shuffles assume that any decoded position bin is equally likely to follow any other decoded position in the null case—an assumption that is violated when neurons fire continuously in bursts that span multiple time bins [63].

For each of these shuffles, there are further experimenter decisions and imposed criteria that can affect the detected candidates and their null distributions. These include, but are not limited to, the percentage and/or number of active cells, duration of events, the choice of time bin size, the average or maximum jumps between bins allowed and how empty bins are handled (e.g. is the posterior probability uniform or zero across all positions in such bins and whether they are ignored in shuffles if they fall in the first or last bins), a point which is often not explicitly noted in methods sections. Another critical point is that if multiple templates are being independently evaluated (e.g. for forward versus reverse, and for each of several maze arms), there is a greater chance of false positives due to multiple-hypothesis testing (see also [17]). It is, therefore, important that the surrogate data following shuffles is treated similarly to real data and tested against each template to determine null hypothesis replay scores. Additionally, further properties of the data, aside from the factors already mentioned, can yield deviations from null distribution due to unintended reasons. For example, serial-position effects, such as biases for SWRs to be initiated by place-cells encoding the current and/or

rewarded locations (as opposed to random locations along the track), could produce deviations from shuffle distributions that randomize decoded locations. Likewise, stationary events (SWRs with content that is fixed in one location) would qualify as 'replay' under some metrics (e.g. 'replay score') but not others (e.g. 'weighted correlation'). One reasonable solution to manage these issues is to consider combinations of measures (e.g. weighted correlation and mean jump distance) at different thresholds [10,17].

In summary, the main message from the above considerations is that replay analysis methods can be sensitive to *both* which cells are active and to their temporal order. Depending on the choice of shuffle(s), data may deviate from the null distribution due to factors that are not related to the sequential firing patterns during SWRs. In these cases, it is important to avoid interpreting the results as necessarily being due to one particular factor, such as sequential order. In the next section, we consider 'second-order' questions that compare replay content across different conditions, such as 'is there more replay in sleep after experience than in sleep before?' These questions inherit all the above issues related to the choice of shuffle and associated interpretations and additionally entail another layer of analysis issues, which we discuss next.

## 4. Second-order replay analysis

A typical second-order sequence replay question is of the form: is sequence A replayed more frequently and/or more strongly than sequence B? Second-order questions go beyond the existence proof required of first-order questions (does reverse replay/remote replay/etc. exist?) to comparing the relative prevalence or strength of sequential structure across different conditions. Thus, the question 'does reverse replay exist?' is a first-order question, whereas comparing the relative strengths of forward and reverse replay [15,23] is a second-order question. Other common second-order questions include comparing replay of different segments of a maze environment, or replay after versus before an experience.

### (a) Forward versus reverse sequences

Such comparisons began almost immediately after the observation of awake replay [14]. Diba & Buzsáki [15] compared the relative prevalence of forward (positively correlated) and reverse (negatively correlated) sequences both before and after runs across the track. Forward sequences were found to be of upcoming trajectories prior to a run, suggestive of planning, while reverse sequences were observed after consumption of reward at the end of the run, suggestive of reward processing. Extending this work, Ambrose *et al.* [23] directly compared templates for outgoing and incoming trajectories under high-reward, low-reward and no-reward conditions on a linear track. They found that increasing the amount of reward on every trial produced an increase in the relative amount of reverse versus forward sequences, supporting the proposed role of reverse replay in reward processing [14]. Olafsdottir *et al.* [68] examined the relative prevalence of different trajectories during short pauses in the task versus long pauses that likely reflect disengagement from the task, and found that forward sequences were more prevalent in short pauses when the animal remained engaged. On the other hand, Davidson *et al.* [56] found that when the animal stopped in the middle of a track and could resume running in either of two directions,

there was no correlation between sequence replays and the trajectories taken by the animal. Similarly, Shin *et al.* [29] reported a strong correspondence between forward sequences and upcoming trajectories, and reverse sequences and completed paths—except at choice points. In a remarkable study, Xu *et al.* [69] demonstrated that while forward sequences corresponded to the upcoming trajectory of the animal from its current position, reverse sequences frequently originated at remote locations and propagated toward reward sites, highlighting the special roles of reward processing in reverse sequences and planning in forward sequences. Pfeiffer & Foster [70] examined these questions in open field two-dimensional trajectories, where place-fields are omnidirectional (so that forward versus reverse sequences are not distinguishable) and found that SWR sequences more closely matched upcoming trajectories rather than replays of past trajectories, consistent with the proposed role for replay in planning [15]. Interestingly, Stella *et al.* [71] found that in the absence of a reliable goal, SWRs instead reflect random-walk trajectories through the two-dimensional open field.

In summary, while many of these studies differed regarding the relative prevalence of forward versus reverse sequences across all trajectory sequences, there appears to be an emerging consensus that forward sequences benefit immediate planning while reverse sequences are modulated by recent reward. Nevertheless, it is important to keep in mind that these forward and reverse replays are not mutually exclusive [56]. Indeed, Wu & Foster [57] found that forward, reverse and mixed replays frequently stitched together to form spatially consistent trajectories, and, further, the ‘shortcut’ sequences reported by Gupta *et al.* [19] may result from the juxtaposition of spatially contiguous forward and reverse events. Ultimately, both types of sequences likely aid the animal in evaluating the reward structure of an environment to benefit future spatial decisions.

The main question for these analyses involved comparing whether a temporal event was better matched to one template (e.g. run from left-to-right) or to a second template (run from right-to-left) in either forward or reverse. Though there are several different ways to perform such a comparison [15,56,57], evaluation in these studies has generally been straightforward because (i) the templates compared (forward and reverse) are balanced—that is, subjects run through each template trajectory an equal number of times—and (ii) they are applied to the same SWR data. Yet, even this relatively simple case is not immune to possible biases due to different numbers or different properties of cells in each of the templates (discussed in the next section).

### (b) Sequences at time-1 versus time-2

A different second-order question involves a comparison of sequences that occur at different time points. A salient example of this type of question is preplay versus replay, which requires comparing sequences before experience with sequences following experience. Such a comparison between two time points circumvents potential confounds from shuffling methods alone, although the quantification will still depend on the choice of shuffle. As a result, virtually all investigators agree that replay in post-task sleep is stronger than preplay in sleep before the task, although by exactly how much is still under investigation and debate [13,72].

Further examples of this kind of question focus on the conditions under which replay at one time is enhanced relative to

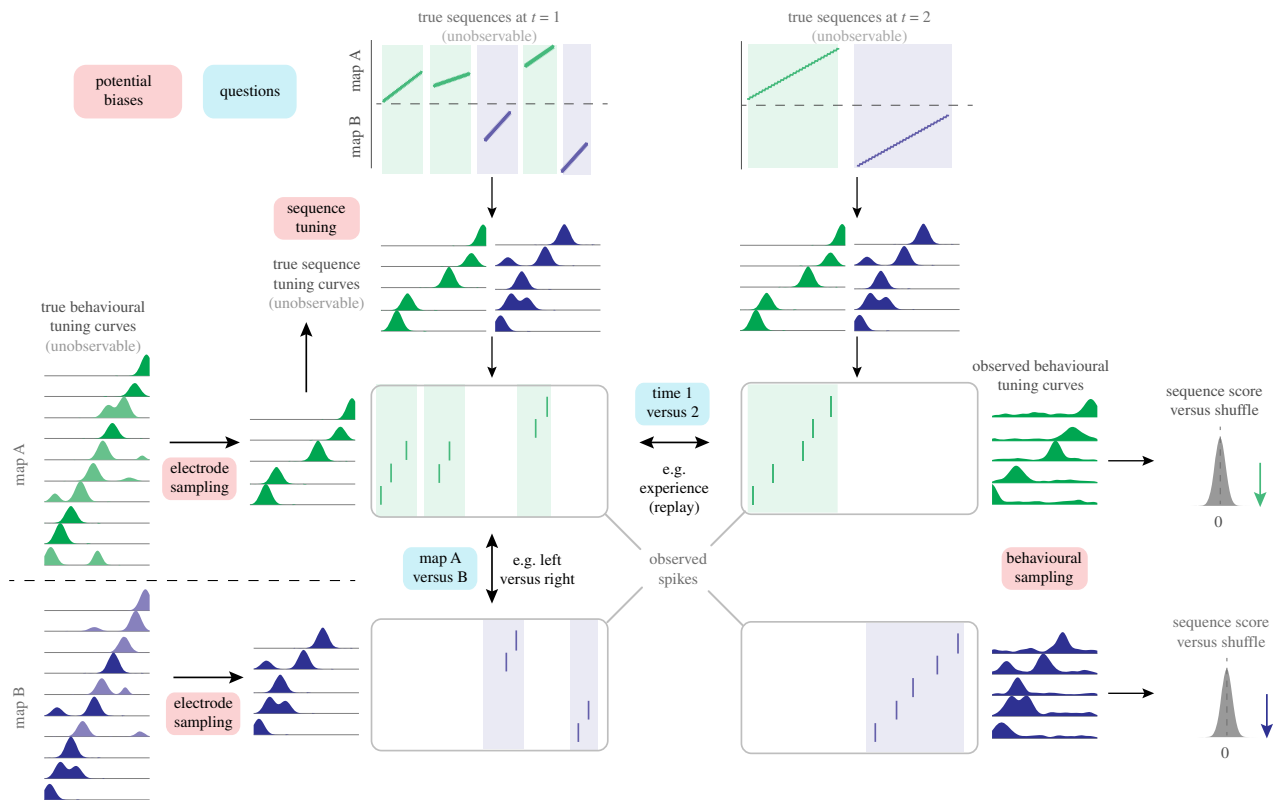
replay at other times. For example, Singer *et al.* [22] observed increased SWR reactivation of place-cells following rewarded trials, compared to unrewarded trials, indicating an important role for reward processing during these events. Note, however, that because reverse and forward sequences were not detected or separated in this study, it is unknown if this role was attributable exclusively to reverse replay [14,23]. In an alternation task on a W-shaped maze, Singer *et al.* [44] reported increased SWR reactivation of place-cells prior to correct trials versus incorrect trials. This provided support for the role of replay in planning and effective execution of upcoming routes, though this study similarly did not examine whether forward sequences alone could account for the planning component [15]. In a recent study, Shin *et al.* [29] found that reverse replay decreases in frequency with time on the maze, whereas forward replay increases in frequency during this same period, indicating that SWR content can indeed change dynamically with experience.

Time-1 versus time-2 comparisons are similar to forward versus reverse questions because they use the *same templates* for analysis, but are distinct because the templates are applied to the analysis of *different data* (e.g. pre- versus post-task). While this approach has clear benefits, it does also introduce some possible issues, which we discuss in the section below.

### (c) Replay of track-1 versus track-2

The more complicated version of second-order sequence analysis involves comparing replay of different tracks or mazes, on which the animals potentially have different amounts of experience, different behaviour and different numbers of place-fields. For example, Olafsdottir *et al.* [73] observed greater sequence preplay of an arm of the T-maze that the animal was cued to enter, compared to the uncued arm, indicating anticipation of the future path during SWRs. Wu *et al.* [74] examined different regions of a linear track after delivering electric shocks in one of the segments. They found more replay of the shock zone that the animal actively avoided, in comparison to a control region at the opposite end of the track that the animal also avoided, but that represented less danger. Xu *et al.* [69] compared replays of different arms of a multi-arm maze. Remarkably, the arm that was replayed varied according to the cognitive demands of the task. In reference memory at the choice point, where the animal needed to remember the rewarded arms, forward replays tended to predict the upcoming path. On the other hand, during working memory on the same maze, when the animal needed to remember where it came from, the previously visited arm was replayed in reverse. The observations in these studies provide strong support for the flexibility of hippocampal sequences in supporting cognition and decision making (see also [8]).

However, other second-order examinations present findings that are more challenging to reconcile with this simple picture. In the first study of its kind to question the relationship between experience and replay, Gupta *et al.* [19] found that when rats were rewarded on only the left arm of a continuous T-maze, they replayed the opposite (right) arm more often than if they alternated between left and right. This finding is in conflict with an experience-driven account of replay, which would predict that the more frequently visited arm should be replayed more often. In a free-choice variant of this task, where the animal could choose whether to run for a water reward if



**Figure 2.** Framework for identifying issues in second-order sequence analysis. Underlying ‘true’ sequences (top, shown here as linear changes in position over time, though nonlinear changes can be similarly represented) occur on either of two maps (A, green shading; B, blue shading). These are unobservable ‘ground truth’ sequences that the data analysis seeks to recover. True sequences are postulated to generate observable spiking activity (shown in the rasterplots) through *sequence tuning curves*, which relate an internal position signal to firing rate, and are also not directly observable. True sequence content may differ between time points ( $t = 1$  shown in the left column versus  $t = 2$  shown on the right) and between maps (A versus B). Such analyses are inherently restricted because they sample neural activity from a finite number of cells (*electrode sampling*, left), and sample *behavioural tuning curves* (firing rate as a function of position, used to determine the sequence template) based on finite behavioural data (*behavioural sampling*, illustrated by the noisy tuning curves on the right). Sequence tuning curves, electrode sampling and behavioural sampling may all differ between time points and/or maps, potentially resulting in biases in resulting sequence scores. (Online version in colour.)

thirsty, or a food reward if hungry, Carey *et al.* [21] remarkably found that replay was biased toward the arm less visited, even in rest before the actual task. In their study, Xu *et al.* [69] saw that when the animal was at the reward site and simply had to return back to the choice point, reverse replays tended to start remotely from the choice point and propagate over to the animal’s current location. These types of trajectory sequences do not fit readily within the planning/reward framework and present a challenge to simple models of the cognitive or computational benefits of sequence replay.

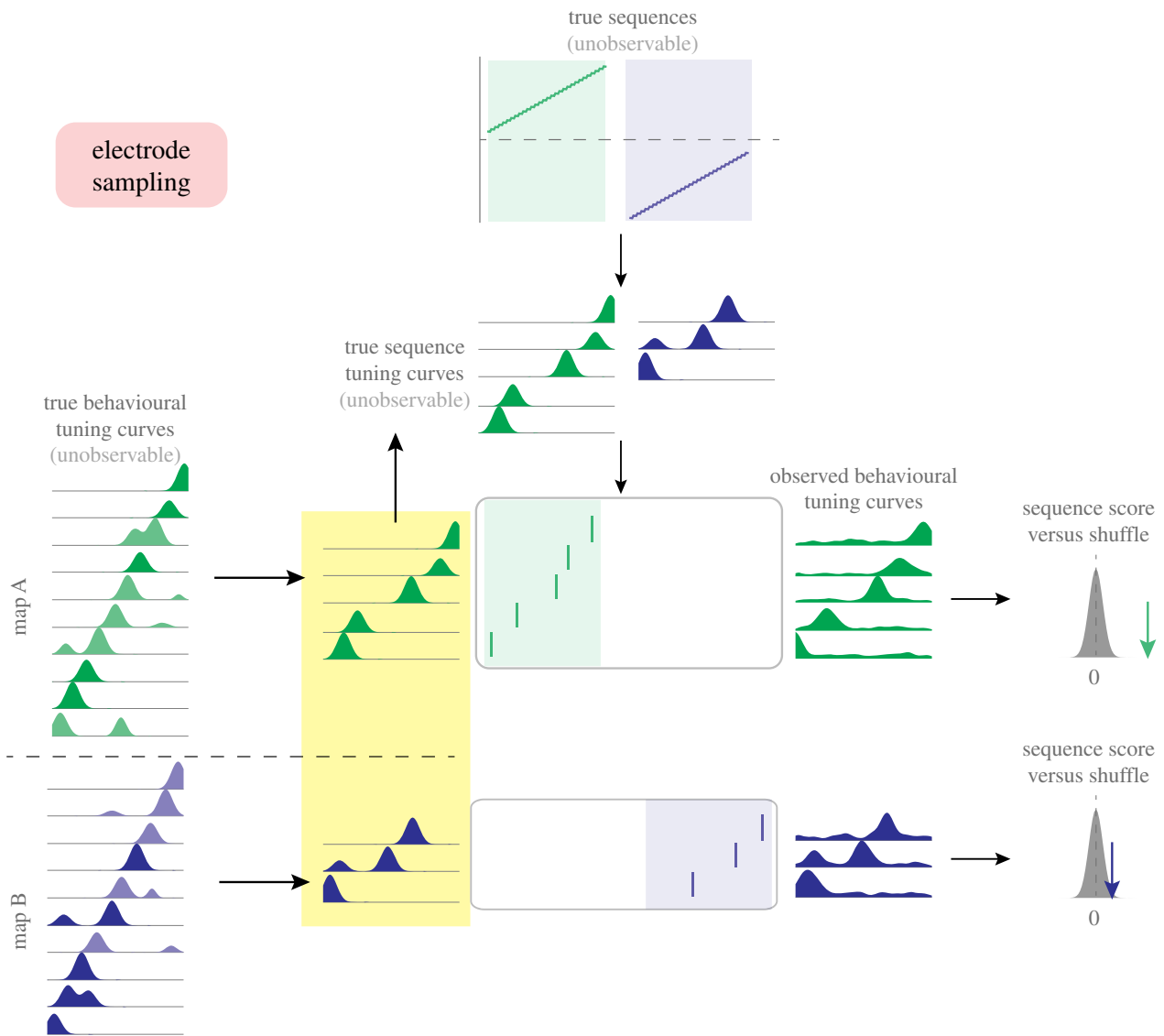
Overall, the above types of second-order comparisons are more complex than the previous forward/reverse comparisons because they require different templates for each of the arms. This introduces possible biases due to differential electrode sampling (different numbers of cells recorded, cells with different firing rates, etc.) and behavioural sampling of the environment (when the animal’s behaviour involves an experiential bias towards segments of the environment, e.g. by spending more time, running through them faster, etc.). We discuss such biases in the next section.

## 5. Framework for issues in second-order sequence analysis

A number of potential analysis and interpretation issues arise when addressing seemingly straightforward A versus B or

time-1 versus time-2 questions comparing replay under different conditions. In general, under these conditions, the detection and scoring of replay is susceptible to biases that are unrelated to underlying sequential content, but may nevertheless result in differences between observed replays of A versus B. These false positives can lead to incorrect interpretations of results. Fortunately, such biases can be prevented or minimized by experimental design and careful analysis. In addition, some diagnostics are available to determine if these biases are occurring so that the interpretation can be modified accordingly. We conceptualize these biases within an overall conceptual framework for replay generation (figure 2).

In this framework, the hippocampus maps any given environment with a set of *true behavioural tuning curves*, which describe the relationship between locations in that environment and the firing rates of neurons. These are not directly observable for two reasons. First, any given experiment records from only a limited number of neurons (*cell sampling*). Second, tuning curve estimates are inherently based on finite and variable *behavioural sampling* of locations in the environment, and are therefore susceptible to various covariates and confounds. For example, animals may spend more time at some locations than others. Their running speeds and acceleration can also vary, not only according to the spatial configuration in the maze, but also because of behavioural factors such as familiarity, anxiety or expected reward. Furthermore, there are nonstationarities in the true tuning curves (e.g. emerging, drifting and expanding



**Figure 3.** Schematic illustration of bias in second-order sequence analysis due to unequal electrode sampling. Following the general flow introduced in figure 2, unobservable underlying sequences (top) generate sequences of spiking activity (rasterplots) across two different hippocampal maps (equivalently, different arms of a maze). Differences in the number of recorded neurons between the two maps (here: five cells for map A, green; three cells for map B, blue) will result in differences in the sequence scores for the two maps, even though there is no underlying difference in replay content. (Online version in colour.)

place-fields), yet the experimenter must average over multiple trials to estimate tuning curves. Thus the observed tuning curves are inherently imperfect approximations of the true behavioural tuning curves.

Likewise, in our framework observed sequences of spikes (rasterplots in figure 2) during candidate replays arise from *true sequences* generated by the brain from a set of *true sequence tuning curves* that are also not directly observable. The true sequence tuning curves are presumably translated by the brain from the true behavioural tuning curves and are therefore related but not necessarily identical to them; for instance, instantaneous firing rates during SWRs can differ from those during behaviour [75]. Moreover, sequence tuning curves presumably represent ‘mental’ time travel through space, as opposed to physical travel during the experience. As such, they may substantially differ from those obtained from behavioural sampling. The spikes during SWRs are nevertheless analysed by comparison with the *observed behavioural tuning curves*, resulting in sequence scores. We note that because of uncertainty regarding the transformation between behavioural and sequential tuning curves, initial studies often used

nonlinear metrics for quantifying replay (e.g. Spearman rank-order correlation coefficients), whereas studies using Bayesian decoding methods assume identical sequence and behavioural tuning curves with interchangeable firing rates ([56] and others). In our framework, we remain agnostic concerning the nature of this transform.

This conceptual framework allows us to highlight a number of possible biases that are particularly relevant for analyses involving second-order questions about sequences:

- (1) Electrode sampling from limited numbers of place cells.
- (2) Imperfect estimates of behavioural tuning curves, which can arise from (a) non-uniform behavioural sampling of the environment, and (b) drift in place cell tuning over time.
- (3) Changes in the sequence tuning curves.

### (a) Biases in cell sampling

In the scenario illustrated in figure 3, limited sampling of neurons results in an unequal distribution of place cells recorded for two maps (A and B; these could be different arms of a maze or



two different environments entirely; 5 cells for map A, 3 cells for map B in this example). It may be that the different cell counts arise from random chance (shot noise) or that units representing one map are more noisy and/or less isolated. As the number of recorded cells increases, the magnitude of this bias will decrease correspondingly. However, the non-linear nature of Bayesian decoding methods, where posterior probabilities from spiking neurons are multiplied together, means that a small difference in the number of recorded cells can lead to large differences in decoded output, even when there is no underlying difference in replay content.

Alternatively, it may be that different cell counts result not from finite sampling but from inherent differences in the number of cells representing maps A and B. Such differences could be important, i.e. larger rewards associated with maze A, resulting in increased density of place-fields (e.g. [67]), or they may be trivial (e.g. maze B may be smaller in size). Increased replay as a consequence of such underlying differences in place field density may or may not be functionally important, but is conceptually distinct from increases in replay that result from other factors not associated with place field density, such as memory prioritization and recall bias.

Regardless of how a difference in the number of recorded place-fields comes about, in the null case when the true replay distribution has equal sequence frequency and strength for both A and B, the observed spike trains will appear unequal: the 5 ordered cells for A result in higher replay scores than the 3 ordered cells for B (right panel). Taken in isolation, this difference in replay scores may be (incorrectly) interpreted as replay content favouring A over B.

### (b) Biases in behavioural sampling

In addition to unequal sampling of neurons, differences in the animal's behaviour under two conditions introduces concerns that require careful consideration. Even if the true neuronal representations are similar in two mazes, if the animal spends more time in A, the experimenter's reconstructed template for B will be more noisy simply because less data was available, which in turn can lead to corresponding bias in replay analysis. Other potential sources of unequal behaviour between different conditions include differences in running speed or acceleration through maze locations and the degree of path stereotypy.

As an example of these possibilities, under the null scenario depicted in figure 4, A and B maps are represented by an equal number of cells, but the *quality* of their observed tuning curves is different: messy, multi-peaked fields for B, clean, unimodal fields for A. The true behavioural tuning curves are not systematically different for A and B (left panel); rather, the behavioural tuning curves for B can deviate from the true tuning curves due to differences in behavioural sampling (e.g. 50 trials for A, 5 trials for B), differences in the animal's behaviour (e.g. more consistent runs in A, but stoppages or changing running speeds in B, etc.) and/or differences in nonspatial cues influencing place cell activity between the arms. In a similar way to figure 3 above, this difference in the observed tuning curves between the A and B maps will lead to a difference in replay scores for A and B because the noisier 'B' template is now out of order (note red dots). This difference may be incorrectly interpreted as a difference in replay content. Furthermore, for questions that require trial-unique analysis of replay content (e.g. in a reward devaluation or replanning scenarios where the key manipulation can

only be done once) behavioural sampling bias remains a particular concern.

### (c) Non-stationary tuning curves

A different source of bias in replay content arises from imperfect behavioural tuning curve estimates related to temporal drift, i.e. the known tendency for at least some cells to change their place tuning over time [76–78], particularly in new environments [79]. This problem is further compounded by instabilities in unit recordings, during which electrode drift or cell attrition can affect the isolation distance and cluster quality of a unit over time. Imagine that behavioural tuning curves are estimated from an initial task epoch and first applied to sequences that occur shortly after (figure 5, left; 'time 1'). In this case, the behavioural tuning curves are a good match for the true tuning curves. However, if the same behavioural tuning curves are used for analysing sequences that occur sometime later (time 2) the true tuning curves may have shifted (figure 5, right) causing a difference in observed replay score even in the absence of a difference in true underlying sequences.

If this scenario seems far-fetched, consider that van der Meer *et al.* [80] showed clear differences in decoding accuracy when decoding trials that were 1 versus 10 trials apart. Additionally, observed tuning curves are known to change over time as a result of several factors, such as experience-dependent place field expansion [65], look-ahead at decision points [81], rapid switching between multiple maps [82–84] and the presence of different gamma rhythms [85].

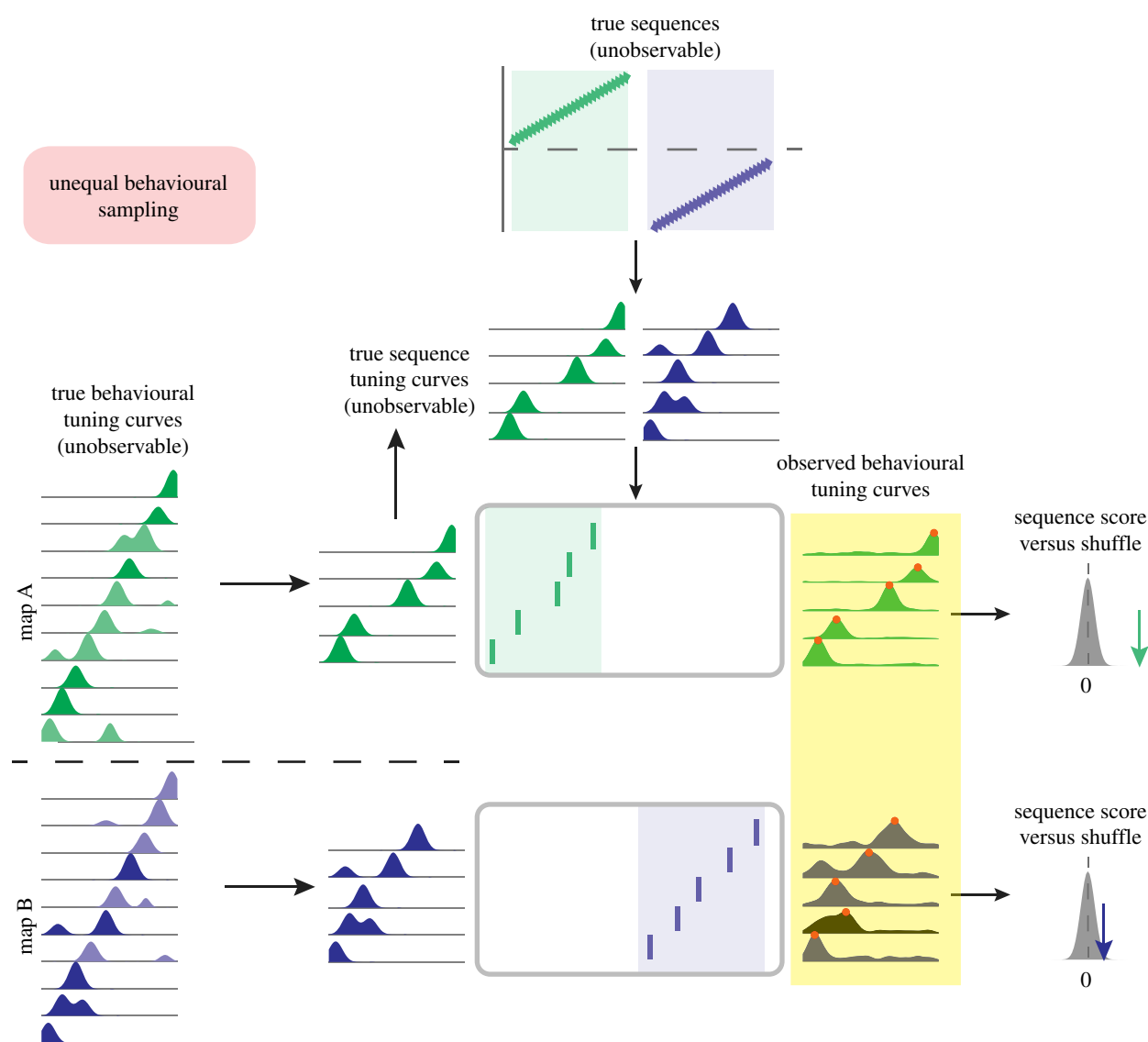
On the other hand, for instances in which 'time 1' and 'time 2' are closer together (e.g. rewarded versus unrewarded trials, or correct versus error trials), examining the reactivation or co-activation probabilities of place cells or ensembles (i.e. 'what cells') can potentially provide the desired information without the need to evaluate their temporal sequences *per se* (e.g. [20]).

### (d) Changes in sequence generation tuning curves

A final source of bias can arise from differences or changes over time in the true sequence tuning curves. In our generative framework, sequence tuning curves underlie the instantaneous firing rates that occur during SWRs. Figure 6 illustrates the possibility that for map A (green), the sequence generation tuning curves are more precise than for map B (blue), resulting in higher sequence scores for map A. This could occur, for instance, because of differences related to deep versus superficial areas of the CA1 pyramidal cell layer [86]. However, it presents an even more significant concern when comparing replays across time points that might be accompanied by changes in the excitability of cells. For example, neuromodulatory changes between awake and sleep states, or across the sleep/circadian cycle, can potentially affect the likelihood of cells to fire during SWRs, and their true sequence tuning curves will vary accordingly across time points. Furthermore, as noted above, the decoded posterior probabilities during SWRs are sensitive to the number of co-active cells. Thus changes in SWR excitability alone can impact replay scores.

### (e) Relevance of these biases for ensemble reactivation measures

In some instances, investigators may simply be interested in second-order questions involving cells or cell ensembles, rather than temporal sequences. While we have highlighted



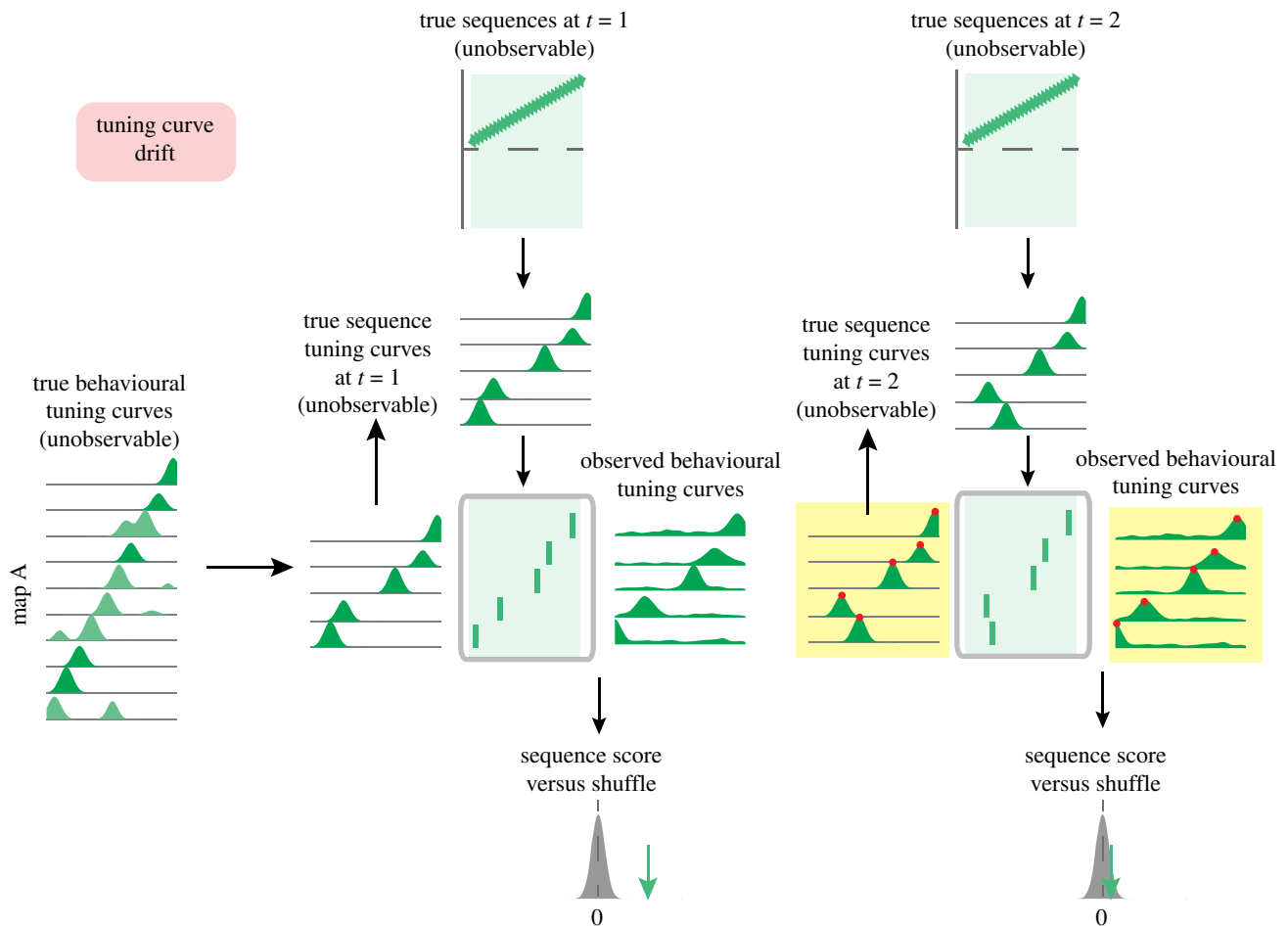
**Figure 4.** Schematic illustration of bias in second-order sequence analysis due to unequal behavioural sampling. Following the general flow introduced in figure 2, unobservable underlying sequences (top) generate sequences of spiking activity (rasterplots) across two different hippocampal maps (equivalently, different arms of a maze). Differences in the data used to estimate tuning curves for the two maps, such as the number of trials and/or the amount of trial-to-trial variability in behaviour, can result in ‘clean’ tuning curves for one map (green tuning curves) and ‘noisy’ tuning curves for another map (blue tuning curves). Note that for the noisy blue map tuning curves the sequence of observed behavioural tuning curve peaks is now in a different order (red dots), resulting in differences in the sequence scores for the two maps, even though there is no underlying difference in replay content. (Online version in colour.)

these issues and potential biases with regards to sequence analysis, it is important to recognize that they also factor into second-order reactivation analyses using pairwise or PCA/ICA ensemble methods. For example, less sampled regions of a maze or fewer sampled neurons also inherently result in poorer estimates for the co-activation patterns of neurons. For ensemble analyses in particular, because these methods typically involve a predetermined time bin, temporal compression between behaviour and replay presents an additional significant issue. Since many ensembles are most activated during SWRs, the common approach is to actually use SWR activity during immobility or sleep, rather than activity during behaviour alone, to define the ensembles, and then to examine their instatement *post hoc* during the behaviour and/or sleep periods. That neurons contribute both positive and negative weights, and that reactivation strength can also take positive and negative values introduce other points worthy of careful consideration. Finally, it should be kept in mind that correlation methods measure coordinated changes in firing rates and are, therefore, affected not only by global

activity but also by global silences, such as during DOWN states [87], LOW states [88] or infra-slow oscillations [89]. If these silent or low-activity periods vary across the periods under comparison, they will inevitably produce further confounds.

## 6. Strategies for diagnosis and prevention of second-order sequence analysis issues

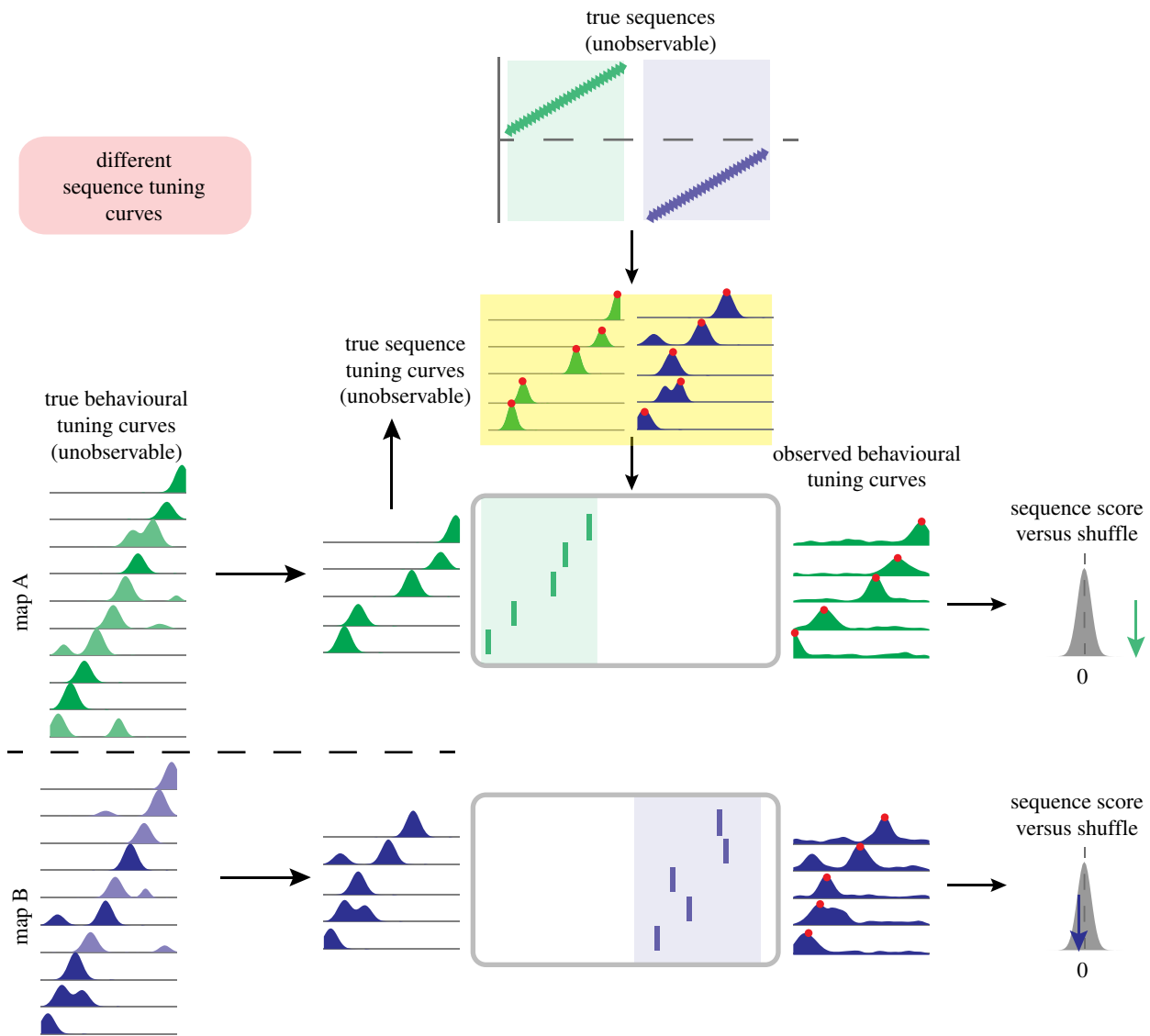
We believe that the above biases resulting from unequal sampling of place cells and unequal estimates of tuning curves are pervasive. It is rare that animal behaviour is exactly equal across an environment. So, what is a replay researcher to do? We suggest two main categories: *bias minimization*, which can occur by experimental design and analysis, and *bias diagnostics*, to determine what biases exist so that the interpretations can be appropriately qualified. Strategies for bias minimization include:



**Figure 5.** Schematic illustration of bias in second-order sequence analysis due to non-stationary tuning curves. Following the general schematic introduced in figure 2, unobservable underlying sequences (top) generate sequences of spiking activity (rasterplots), illustrated here for a single map. Behavioural tuning curves are estimated from an initial task epoch and first applied to sequences that occur shortly after (time 1). In this case, the behavioural tuning curves are a good match for the true tuning curves. However, if the same behavioural tuning curves are used for analysing sequences that occur sometime later (time 2) the true tuning curves may have shifted (note the difference between true and behaviourally observed tuning curves, highlighted in yellow). Thus, outdated tuning curves are inadvertently used for replay scoring, leading to a reduction in observed replay score even in the absence of a difference in true underlying sequences. (Online version in colour.)

- In experimental design, take steps to promote equal behavioural sampling of the environment wherever possible. This can include not only ensuring that the animal samples different locations as equally as possible, but also standardizing when and how (under different behavioural conditions) the animal samples these locations.
- In analysis, subsample numbers of trials to equalize for the different locations of interest. For instance, if there are 50 A trials and 10 B trials, subsample the A trials to make the A–B comparison more equitable. Below a certain number, nevertheless, decoding accuracy will likely drop; for instance, van der Meer *et al.* [80] found that decoding accuracy dropped substantially when fewer than 5 trials were used to construct tuning curves for decoding.
- If uneven numbers of cells are recorded in the conditions to be considered, similar-sized sections of the environment (e.g. lengths of track) should be compared against each other. This approach will therefore avoid inequalities arising from cell sampling while preserving more important inequalities that may be due to differences in the true behavioural tuning curves, such as a higher density of place-fields in one condition versus the other.
- Future replay scoring methods should hopefully be able to take into account not only the mean firing rate in turning

- curves, but also the uncertainty in that measure. This would help mitigate biases due to unequal sampling that underlie mean firing rate estimations for different locations. The commonly used Bayesian decoding framework for replay analysis [90,91] uses a Poisson process limited to a mean firing rate ( $\lambda$ ) for each location, discarding information about the amount of data this mean estimate was derived from (more uncertain for low sampling, less uncertain for high sampling). Ghanbari *et al.* [92] show that incorporating such uncertainty improves decoding accuracy of reaching movements from motor cortex data; a similar approach will likely be fruitful for hippocampal data under limited sampling conditions. Unit isolation and stability on an event-by-event basis could likewise be incorporated into a more comprehensive probabilistic framework.
- Future replay scoring methods should also aim to avoid the conflation between how strong a replay is (which should be independent of the number of cells) and how much evidence/uncertainty there is about that estimate (which does depend on the number of cells). Current methods such as z-scoring relative to a shuffled distribution mix these two together.
- Alternate methods based on a probabilistic framework such as Hidden Markov Models [59,93] can be valuable



**Figure 6.** Schematic illustration of bias in second-order sequence analysis due to non-stationary sequence tuning curves. Unobservable underlying sequences (top) generate sequences of spiking activity (rasterplots) across two different hippocampal maps (equivalently, different arms of a maze). The tuning curves for the generation of map A sequences (green), may be more precise than those for the generation of map B sequences (blue; note wider tuning curves whose peaks are now out of order, red dots). As a consequence, the generated sequences for map B are now out of order compared to the behaviourally observed tuning curves, resulting in lower sequence scores for map B. (Online version in colour.)

for incorporating uncertainties regarding the factors mentioned above, since in these methods observation states and transition matrices are constructed from distributions observed in the data in a template-free manner. These methods can also be helpful when constructing a template from place fields is not desirable, for example, during one-trial learning or non-spatial behaviours.

Diagnostics for detection of replay content bias include:

- Compute a cross-validated decoding error on behavioural data where a true correct answer, such as the subject's position in space, is available. Importantly, the decoding error on data not included in the training set (i.e. the data from which the decoder is obtained) is sensitive to all biases discussed above [80]. Thus, this error can be used to obtain a null hypothesis for replay content given that the true underlying replay distribution is uniform, i.e. it can reveal the replay content differences that would be expected from experimental biases unrelated to true replay content.

- Generative models of replay can generate spike sequences from a specific model of ground truth. Such synthetic data can then be used to quantify the expected bias due to the various factors discussed above (e.g. limited behavioural sampling of true underlying tuning curves; different numbers and/or firing rates of place cells). Of course, the relevance of the results from such a model is proportional to how accurately the synthetic data captures the properties of real data; ongoing development of such generative models of replay is likely to be a fruitful area of research for multiple reasons.

## 7. Conclusion

The temporally compressed sequential firing patterns of hippocampal neurons during SWRs have rightfully sparked widespread fascination and attention from researchers with a broad range of scientific interests. This is due to the potential role of hippocampal sequences not only in learning and memory, but also in complex decision making, planning and



imagination. Furthermore, recent years have seen (i) the emergence of a productive interaction with reinforcement learning and artificial intelligence systems that incorporate replay [8,94], (ii) the development of behavioural paradigms and fMRI/MEG analysis methods to study not only reinstatement, but also sequential replay in humans [95,96], and (iii) the development of real-time causal interventions targeting SWRs and SWR content [2–4,97,98].

These different lines of work drive current questions about SWR content in increasingly complex experimental designs, such as those involving more richly structured spatial environments with multiple possible trajectories, representation of not only the self but other agents and multi-faceted non-spatial behaviour. The continuing evolution of wireless and recording technologies further enable the exploration of larger environments, and longer continuous recordings. As these developments push the boundaries of replay questions, potential issues, such as unequal behavioural sampling and temporal drift, become increasing concerns. These are likely to be

combined with closed-loop approaches that will modify SWRs and their underlying sequence tuning curves. Yet, as we have argued here, these concerns are already prevalent today, even in seemingly innocuous settings such as comparing two arms of a maze or two different time points. There is ample potential for biases in replay analysis that can lead to erroneous inferences. While the potential for confounds and biases are pronounced, by careful experimental design and consideration of the assumptions underlying null hypotheses, these issues can be understood and reasonably managed.

**Data accessibility.** This article has no additional data.

**Authors' contributions.** All authors contributed equally.

**Competing interest.** We declare we have no competing interests.

**Funding.** This work was funded by HFSP grant no. RGY0088/2014 (M.A.A.v.d.M. and C.K.), NSF grant no. CAREER IOS-1844935 (M.A.A.v.d.M.), NINDS grant no. 1R01NS115233 (C.K. and K.D.) and NIMH grant no. 1R01MH109170 (K.D.).

## References

- Buzsaki G. 2015 Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073–1188. (doi:10.1002/hipo.22488)
- Girardeau G, Benchenane K, Wiener SI, Buzsaki G, Zugaro MB. 2009 Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* **12**, 1222–1223. (doi:10.1038/nn.2384)
- Ego-Stengel V, Wilson MA. 2010 Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* **20**, 1–10. (doi:10.1002/hipo.20707)
- Jadhav SP, Kemere C, German PW, Frank LM. 2012 Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454–1458. (doi:10.1126/science.1217230)
- Michon F, Sun J.-J., Kim CY, Ciliberti D, Kloosterman F. 2019 Post-learning hippocampal replay selectively reinforces spatial memory for highly rewarded locations. *Curr. Biol.* **29**, 1436–1444. (doi:10.1016/j.cub.2019.03.048)
- Sutton RS. 1990 First results with Dyna, an integrated architecture for learning, planning and reacting. In *Neural networks for control* (eds WT Miller, RS Sutton, PJ Werbos), pp. 179–189. Cambridge, MA: MIT Press.
- Johnson A, Redish AD. 2005 Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Netw.* **18**, 1163–1171. (doi:10.1016/j.neunet.2005.08.009)
- Mattar MG, Daw ND. 2018 Prioritized memory access explains planning and hippocampal replay. *Nat. Neurosci.* **21**, 1609–1617. (doi:10.1038/s41593-018-0232-z)
- Tingley D, Peyrache A. 2020 On the methods for reactivation and replay analysis. *Phil. Trans. R. Soc. B* **375**, 20190231. (doi:10.1098/rstb.2019.0231)
- Foster DJ. 2017 Replay comes of age. *Annu. Rev. Neurosci.* **40**, 581–602. (doi:10.1146/annurev-neuro-072116-031538)
- Dragoi G, Tonegawa S. 2010 Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* **469**, 397–401. (doi:10.1038/nature09633)
- Dragoi G, Tonegawa S. 2013 Distinct preplay of multiple novel spatial experiences in the rat. *Proc. Natl Acad. Sci. USA* **110**, 9100–9105. (doi:10.1073/pnas.1306031110)
- Farooq U, Sibille J, Liu K, Dragoi G. 2019 Strengthened temporal coordination within pre-existing sequential cell assemblies supports trajectory replay. *Neuron* **103**, 719–733. (doi:10.1016/j.neuron.2019.05.040)
- Foster D, Wilson M. 2006 Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683. (doi:10.1038/nature04587)
- Diba K, Buzsaki G. 2007 Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* **10**, 1241–1242. (doi:10.1038/nn1961)
- Karlsson MP, Frank LM. 2009 Awake replay of remote experiences in the hippocampus. *Nat. Neurosci.* **12**, 913–918. (doi:10.1038/nn.2344)
- Silva D, Feng T, Foster DJ. 2015 Trajectory events across hippocampal place cells require previous experience. *Nat. Neurosci.* **18**, 1772–1779. (doi:10.1038/nn.4151)
- Wilson MA, McNaughton BL. 1994 Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679. (doi:10.1126/science.8036517)
- Gupta AS, van der Meer MAA, Touretzky DS, Redish AD. 2010 Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705. (doi:10.1016/j.neuron.2010.01.034)
- Singer AC, Carr MF, Karlsson MP, Frank LM. 2013 Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. *Neuron* **77**, 1163–1173. (doi:10.1016/j.neuron.2013.01.027)
- Carey AA, Tanaka Y, van der Meer MAA. 2019 Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nat. Neurosci.* **22**, 1450–1459. (doi:10.1038/s41593-019-0464-6)
- Singer AC, Frank LM. 2009 Rewarded outcomes enhance reactivation of experience in the hippocampus. *Neuron* **64**, 910–921. (doi:10.1016/j.neuron.2009.11.016)
- Ambrose RE, Pfeiffer BE, Foster DJ. 2016 Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron* **91**, 1124–1136. (doi:10.1016/j.neuron.2016.07.047)
- Suh J, Foster DJ, Davoudi H, Wilson MA, Tonegawa S. 2013 Impaired hippocampal ripple-associated replay in a mouse model of schizophrenia. *Neuron* **80**, 484–493. (doi:10.1016/j.neuron.2013.09.014)
- Middleton SJ, Kneller EM, Chen S, Ogiwara I, Montal M, Yamakawa K, McHugh TJ. 2018 Altered hippocampal replay is associated with memory impairment in mice heterozygous for the Scn2a gene. *Nat. Neurosci.* **21**, 996–1003. (doi:10.1038/s41593-018-0163-8)
- Lansink CS, Goltstein PM, Lankelma JV, McNaughton BL, Pennartz CMA. 2009 Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biol.* **7**, e1000173. (doi:10.1371/journal.pbio.1000173)
- Sosa M, Joo HR, Frank LM. In press. Dorsal and ventral hippocampal sharp-wave ripples activate distinct nucleus accumbens networks. *Neuron*.
- Jadhav SP, Rothschild G, Roumis DK, Frank LM. 2016 Coordinated excitation and inhibition of prefrontal ensembles during awake hippocampal sharp-wave

- ripple events. *Neuron* **90**, 113–127. (doi:10.1016/j.neuron.2016.02.010)
29. Shin JD, Tang W, Jadhav SP. 2019 Dynamics of awake hippocampal-prefrontal replay for spatial learning and memory-guided decision making. *Neuron* **104**, 1110–1125. (doi:10.1016/j.neuron.2019.09.012)
30. Todorova R, Zugaro M. 2019 Isolated cortical computations during delta waves support memory consolidation. *Science* **366**, 377–381. (doi:10.1126/science.aay0616)
31. Olafsdottir HF, Carpenter F, Barry C. 2016 Coordinated grid and place cell replay during rest. *Nat. Neurosci.* **19**, 792–794. (doi:10.1038/nn.4291)
32. Trimmer JB, Trettel SG, Hwaun E, Colgin LL. 2017 Methodological caveats in the detection of coordinated replay between place cells and grid cells. *Front. Syst. Neurosci.* **11**, 57. (doi:10.3389/fnsys.2017.00057)
33. Tingley D, Buzsaki G. 2019 Routing of hippocampal ripples to subcortical structures via the lateral septum. *Neuron* **105**, 138–149. (doi:10.1016/j.neuron.2019.10.012)
34. Rothschild G, Eban E, Frank LM. 2017 A cortical–hippocampal–cortical loop of information processing during memory consolidation. *Nat. Neurosci.* **20**, 251–259. (doi:10.1038/nn.4457)
35. Gomperts SN, Kloosterman F, Wilson MA. 2015 VTA neurons coordinate with the hippocampal reactivation of spatial experience. *Elife* **4**, 321–352. (doi:10.7554/eLife.05360)
36. Gutig R, Sompolinsky H. 2006 The tempotron: a neuron that learns spike timing-based decisions. *Nat. Neurosci.* **9**, 420–428. (doi:10.1038/nn1643)
37. Lansink CS, Goltstein PM, Lankelma JV, Joosten RNJMA, McNaughton BL, Pennartz CMA. 2008 Preferential reactivation of motivationally relevant information in the ventral striatum. *J. Neurosci.* **28**, 6372–6382. (doi:10.1523/JNEUROSCI.1054-08.2008)
38. Nadel L. 2008 The hippocampus and context revisited. In *Hippocampal place fields* (ed. SJY Mizumori), pp. 3–15. (doi:10.1093/acprof:oso/9780195323245.003.0002)
39. Colgin LL, Moser EI, Moser M-B. 2008 Understanding memory through hippocampal remapping. *Trends Neurosci.* **31**, 469–477. (doi:10.1016/j.tins.2008.06.008)
40. Kubie JL, Levy ERJ, Fenton AA. In press. Is hippocampal remapping the physiological basis for context? *Hippocampus*. (doi:10.1002/hipo.23160)
41. Pavlides C, Winson J. 1989 Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *J. Neurosci.* **9**, 2907–2918. (doi:10.1523/JNEUROSCI.09-08-02907.1989)
42. Kudrimoti HS, Barnes CA, McNaughton BL. 1999 Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *J. Neurosci.* **19**, 4090–4101.
43. Cheng S, Frank LM. 2008 New experiences enhance coordinated neural activity in the hippocampus. *Neuron* **57**, 303–313. (doi:10.1016/j.neuron.2007.11.035)
44. Singer AC, Carr MF, Karlsson MP, Frank LM. 2013 Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. *Neuron* **77**, 1163–1173. (doi:10.1016/j.neuron.2013.01.027)
45. McNamara CG, Tejero-Cantero A, Trouche S, Campo-Urriza N, Dupret D. 2014 Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nat. Neurosci.* **17**, 1658–1660. (doi:10.1038/nn.3843)
46. Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP. 2009 Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* **12**, 919–926. (doi:10.1038/nn.2337)
47. Benchenane K, Peyrache A, Khamassi M, Tierney PL, Gioanni Y, Battaglia FP, Wiener SI. 2010 Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron* **66**, 921–936. (doi:10.1016/j.neuron.2010.05.013)
48. van de Ven GM, Trouche S, McNamara CG, Allen K, Dupret D. 2016 Hippocampal offline reactivation consolidates recently formed cell assembly patterns during sharp wave-ripples. *Neuron* **92**, 968–974. (doi:10.1016/j.neuron.2016.10.020)
49. Skaggs WE, McNaughton BL. 1996 Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271**, 1870–1873.
50. Euston DR, Tatsuno M, McNaughton BL. 2007 Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science* **318**, 1147–1150. (doi:10.1126/science.1148979)
51. Giri B, Miyawaki H, Mizuseki K, Cheng S, Diba K. 2019 Hippocampal reactivation extends for several hours following novel experience. *J. Neurosci.* **39**, 866–875. (doi:10.1523/JNEUROSCI.1950-18.2018)
52. Nadasdy Z, Hirase H, Czurko A, Csicsvari J, Buzsaki G. 1999 Replay and time compression of recurring spike sequences in the hippocampus. *J. Neurosci.* **19**, 9497–9507. (doi:10.1523/JNEUROSCI.19-21-09497.1999)
53. Lee AK, Wilson MA. 2002 Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194. (doi:10.1016/S0896-6273(02)01096-6)
54. Louie K, Wilson MA. 2001 Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* **29**, 145–156. (doi:10.1016/S0896-6273(01)00186-6)
55. Wang Y, Romani S, Lustig B, Leonardo A, Pastalkova E. 2014 Theta sequences are essential for internally generated hippocampal firing fields. *Nat. Neurosci.* **18**, 282–288. (doi:10.1038/nn.3904)
56. Davidson TJ, Kloosterman F, Wilson MA. 2009 Hippocampal replay of extended experience. *Neuron* **63**, 497–507. (doi:10.1016/j.neuron.2009.07.027)
57. Wu X, Foster DJ. 2014 Hippocampal replay captures the unique topological structure of a novel environment. *J. Neurosci.* **34**, 6459–6469. (doi:10.1523/JNEUROSCI.3414-13.2014)
58. Grosmark AD, Buzsaki G. 2016 Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science* **351**, 1440–1443. (doi:10.1126/science.aad1935).
59. Maboudi K, Ackermann E, de Jong LW, Pfeiffer BE, Foster DJ, Diba K, Kemere C. 2018 Uncovering temporal structure in hippocampal output patterns. *Elife* **7**, e34467. (doi:10.7554/eLife.34467)
60. Mackevicius EL, Bahle AH, Williams AH, Gu S, Denisenko NI, Goldman MS, Fee MS. 2019 Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *Elife* **8**, e38471. (doi:10.7554/eLife.38471)
61. Rubin A, Sheintuch L, Brande-Eilat N, Pincasof O, Rechavi Y, Geva N, Ziv Y. 2019 Revealing neural correlates of behavior without behavioral measurements. *Nat. Commun.* **10**, 1–14. (doi:10.1038/s41467-018-07882-8)
62. Wang Y, Romani S, Lustig B, Leonardo A, Pastalkova E. 2014 Theta sequences are essential for internally generated hippocampal firing fields. *Nat. Neurosci.* **18**, 282–288. (doi:10.1038/nn.3904)
63. Ranck JB. 1973 Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. Behavioral correlates and firing repertoires. *Exp. Neurol.* **41**, 461–531.
64. Mizuseki K, Buzsaki G. 2013 Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell Rep.* **4**, 1010–1021. (doi:10.1016/j.celrep.2013.07.039)
65. Mehta MR, Barnes CA, McNaughton BL. 1997 Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921. (doi:10.1073/pnas.94.16.8918)
66. Hollup SA, Molden S, Donnett JG, Moser MB, Moser EI. 2001 Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *J. Neurosci.* **21**, 1635–1644. (doi:10.1523/JNEUROSCI.21-05-01635.2001)
67. Dupret D, O'Neill J, Pleydell-Bouverie B, Csicsvari J. 2010 The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat. Neurosci.* **13**, 995–1002. (doi:10.1038/nn.2599)
68. Ólafsdóttir HF, Carpenter F, Barry C. 2017 Task demands predict a dynamic switch in the content of awake hippocampal replay. *Neuron* **96**, 925–935. (doi:10.1016/j.neuron.2017.09.035)
69. Xu H, Baracskey P, O'Neill J, Csicsvari J. 2019 Assembly responses of hippocampal CA1 place cells predict learned behavior in goal-directed spatial tasks on the radial eight-arm maze. *Neuron* **101**, 119–132. (doi:10.1016/j.neuron.2018.11.015)
70. Pfeiffer BE, Foster DJ. 2013 Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79. (doi:10.1038/nature12112)
71. Stella F, Baracskey P, O'Neill J, Csicsvari J. 2019 Hippocampal reactivation of random trajectories resembling Brownian diffusion. *Neuron* **102**, 450–461. (doi:10.1016/j.neuron.2019.01.052)
72. Silva D, Feng T, Foster DJ. 2015 Trajectory events across hippocampal place cells require previous

- experience. *Nat. Neurosci.* **18**, 1772–1779. (doi:10.1038/nn.4151)
73. Olafsdottir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ. 2015 Hippocampal place cells construct reward related sequences through unexplored space. *Elife* **4**, e06063. (doi:10.7554/eLife.06063)
74. Wu C-T, Haggerty D, Kemere C, Ji D. 2017 Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* **20**, 571. (doi:10.1038/nn.4507)
75. Csicsvari J, Hirase H, Czurko A, Mamiya A, Buzsaki G. 1999 Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J. Neurosci.* **19**, 274–287.
76. Mankin EA, Sparks FT, Slayyeh B, Sutherland RJ, Leutgeb S, Leutgeb JK. 2012 Neuronal code for extended time in the hippocampus. *Proc. Natl Acad. Sci. USA* **109**, 19 462–19 467. (doi:10.1073/pnas.1214107109)
77. Ziv Y, Burns LD, Cocker ED, Hamel EO, Ghosh KK, Kitch LJ, El Gamal A, Schnitzer MJ. 2013 Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266. (doi:10.1038/nn.3329)
78. Rubin A, Geva N, Sheintuch L, Ziv Y. 2015 Hippocampal ensemble dynamics timestamp events in long-term memory. *Elife* **4**, 723–727. (doi:10.7554/eLife.12247)
79. Frank LM, Stanley GB, Brown EN. 2004 Hippocampal plasticity across multiple days of exposure to novel environments. *J. Neurosci.* **24**, 7681–7689. (doi:10.1523/JNEUROSCI.1958-04.2004)
80. van der Meer MAA, Carey AA, Tanaka Y. 2017 Optimizing for generalization in the decoding of internally generated activity in the hippocampus. *Hippocampus* **27**, 580–595. (doi:10.1101/066670)
81. Johnson A, Redish AD. 2007 Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12 176–12 189. (doi:10.1523/JNEUROSCI.3761-07.2007)
82. Fenton AA, Muller RU. 1998 Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc. Natl Acad. Sci. USA* **95**, 3182–3187. (doi:10.1073/pnas.95.6.3182)
83. Jackson J, Redish AD. 2007 Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. *Hippocampus* **17**, 1209–1229. (doi:10.1002/hipo.2035)
84. Dvorak D, Radwan B, Sparks FT, Talbot ZN, Fenton AA. 2018 Control of recollection by slow gamma dominating mid-frequency gamma in hippocampus CA1. *PLoS Biol.* **16**, e2003354. (doi:10.1371/journal.pbio.2003354)
85. Bieri KW, Bobbitt KN, Colgin LL. 2014 Slow and fast gamma rhythms coordinate different spatial coding modes in hippocampal place cells. *Neuron* **82**, 670–681. (doi:10.1016/j.neuron.2014.03.013)
86. Mizuseki K, Diba K, Pastalkova E, Buzsaki G. 2011 Hippocampal CA1 pyramidal cells form functionally distinct sublayers. *Nat. Neurosci.* **14**, 1174–1181. (doi:10.1038/nn.2894)
87. Mochol G, Hermoso-Mendizabal A, Sakata S, Harris KD, la Rocha J. 2015 Stochastic transitions into silence cause noise correlations in cortical circuits. *Proc. Natl Acad. Sci. USA* **112**, 3529–3534. (doi:10.1073/pnas.1410509112)
88. Miyawaki H, Billeh YN, Diba K. 2017 Low activity microstates during sleep. *J. Sleep Sleep Disord. Res.* **40**, zsx066. (doi:10.1093/sleep/zsx066)
89. Watson BO. 2018 Cognitive and physiologic impacts of the infraslow oscillation. *Front. Syst. Neurosci.* **12**, 44. (doi:10.3389/fnsys.2018.00044)
90. Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ. 1998 Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044. (doi:10.1152/jn.1998.79.2.1017)
91. Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. 1998 A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* **18**, 7411–7425.
92. Ghanbari A, Lee CM, Read HL, Stevenson IH. 2019 Modeling stimulus-dependent variability improves decoding of population neural responses. *J. Neural Eng.* **16**, 066018. (doi:10.1088/1741-2552/ab3a68)
93. Chen Z, Grosmark AD, Penagos H, Wilson MA. 2016 Uncovering representations of sleep-associated hippocampal ensemble spike activity. *Sci. Rep.* **6**, 32193. (doi:10.1038/srep32193)
94. Mnih V *et al.* 2015 Human-level control through deep reinforcement learning. *Nature* **518**, 529.
95. Kurth-Nelson Z, Economides M, Dolan RJ, Dayan P. 2016 Fast sequences of non-spatial state representations in humans. *Neuron* **91**, 194–204. (doi:10.1016/j.neuron.2016.05.028)
96. Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TE. 2019 Human replay spontaneously reorganizes experience. *Cell* **178**, 640–652.e14. (doi:10.1016/j.CELL.2019.06.012)
97. Ciliberti D, Kloosterman F. 2017 Falcon: a highly flexible open-source software for closed-loop neuroscience. *J. Neural Eng.* **14**, 045004. (doi:10.1088/1741-2552/aa7526)
98. Fernandez-Ruiz A, Oliva A, de Oliveira EF, Rocha-Almeida F, Tingley D, Buzsaki G. 2019 Long-duration hippocampal sharp wave ripples improve memory. *Science* **364**, 1082–1086. (doi:10.1126/SCIENCE.AAX0758)