Trends in Cognitive Sciences

CellPress

Opinion

The Functional Anatomy of Time: What and When in the Brain

Karl Friston[1,*] and Gyorgy Buzsáki[2,3]

This Opinion article considers the implications for functional anatomy of how we represent temporal structure in our exchanges with the world. It offers a theoretical treatment that tries to make sense of the architectural principles seen in mammalian brains. Specifically, it considers a factorisation between representations of temporal succession and representations of content or, heuristically, a segregation into when and what. This segregation may explain the central role of the hippocampus in neuronal hierarchies while providing a tentative explanation for recent observations of how ordinal sequences are encoded. The implications for neuroanatomy and physiology may have something important to say about how self-organised cell assembly sequences enable the brain to exhibit purposeful behaviour that transcends the here and now.

Trends

Recent studies of hippocampal responses suggest that they have an intrinsic dynamics that may complement (or nuance) spatiotemporal encoding, particularly the encoding of trajectories through space and time and inherent place-cell activity.

Predictive coding and the Bayesian brain now predominate as explanations for much of cognitive neuroscience and functional anatomy in the brain and have clear relevance for the encoding of trajectories through various state spaces.

Recent attempts to understand the form of ordinal or sequential processing in the brain (e.g., navigation, language) emphasise prediction and may be fundamentally informed by recent empirical findings from the study of hippocampal (and neocortical) responses.

The Principles of Functional Anatomy

There are certain architectural principles of neuroanatomy that seem amenable to explanation from a purely theoretical perspective. These range from the existence of axonal processes that form neuronal connections to macroscopic organisational principles such as functional segregation [1]. A key example is the segregation of dorsal and ventral streams into what and where streams [2]. How might these architectural features be explained from a theoretical perspective? In what follows, we appeal to active inference and the Bayesian brain hypothesis [3,4] to suggest that functional segregation emerges from statistical structure in the environment. We then consider the implications of this argument for a fundamental aspect of this structure; namely, the trajectories or ordered sequences of states that we encounter [5]. Our conclusion is that there should be a functional segregation between what and when–a conclusion that seems to explain numerous anatomical and physiological observations, particularly in the hippocampal system.

Good Enough Brains and Good Enough Models

A key theoretical development in neurobiology is the appreciation of the brain as a predictive organ generating predictions of its actions and sensations [4,6–9]. These predictions rest on an internal or **generative model** (see Glossary) of how sensory input unfolds. One can understand much of neuronal dynamics and synaptic plasticity as an optimisation of (Bayesian) **model evidence** as scored by proxies like free energy and prediction errors [9–11]. If one subscribes to this normative theory, the brain must be a good (enough) model of its environment, where recurring sequences of events are the rule. This idea dates back to notions of good regulators in self-organisation and cybernetics [12,13]. In brief, the good regulator theorem states that any system that can control its environment must be a good model of that environment. So what constitutes a good enough model?

Mathematically, a good enough model is simply a model that has sufficient evidence in light of the (sensory) data it has to explain. Evidence is the probability of sensory samples under a model of

[1]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK
[2]NYU Neuroscience Institute, School of Medicine, New York University, New York, NY 10016, USA
[3]Institute of Experimental Medicine, Hungarian Academy of Sciences, Budapest, Hungary

*Correspondence: k.friston@ucl.ac.uk (K. Friston).

 http://dx.doi.org/10.1016/j.tics.2016.05.001

Box 1. Approximate Bayesian Inference

Bayesian inference refers to optimising beliefs about a model or its hidden states (s) in the light of outcomes (o) or evidence. Formally, this can be expressed as minimising a variational free energy bound on Bayesian model evidence [91] with respect to beliefs about hidden states encoded by a probability density $Q(s)$ (with expectation $E[Q(s)] = \mathbf{s}$).

$$F(o,\mathbf{s}) = \underbrace{D[Q(s)||P(s|o)]}_{\text{relative entropy}} - \underbrace{\ln P(o)}_{\text{log evidence}} \geq \underbrace{\ln P(o)}_{\text{log evidence}}$$
$$= \underbrace{D[Q(s)||P(s)]}_{\text{complexity}} - \underbrace{E_Q[\ln P(o|s)]}_{\text{accuracy}}$$

Here, the model is specified by a joint distribution over outcomes and their causes or hidden states: $P(o, s) = P(o|s)P(s)$. The first expression for free energy shows that when free energy is minimised, the relative entropy or Kullback–Leibler (KL) divergence attains its minimum (zero) and free energy becomes the negative logarithm of model evidence. In other words, when free energy is minimised, the approximate posterior beliefs become the true posterior beliefs (i.e., the distribution of hidden states given outcomes) and free energy becomes negative log evidence.

Another way of conceptualising free energy is in terms of accuracy and **complexity**, as shown in the second equality. This equality shows that minimising free energy minimises complexity. Here, complexity is the KL divergence between posterior beliefs and **prior beliefs** (prior to any outcomes). In other words, complexity reflects the degrees of freedom–above and beyond prior beliefs–needed to provide an accurate account of observed data. It follows that when one is absolutely certain about the hidden states causing data, the complexity increases with the number of hidden states entertained by the model.

The imperative to minimise complexity is known as Occam's principle and is the basis of approximations to model evidence provided by the Akaike and Bayesian information criteria [92]. The role of complexity will become important below, when we consider models with a large number of states encoding joint distributions over two factors relative to parsimonious models (with greater model evidence) that encode just the factors or marginal densities (Box 2). In terms of the equations above, this distinction can be expressed as the mean field approximation $Q(s) = Q(s^{where})Q(s^{what})$.

how those samples were generated (Box 1). In this sense, any brain can be viewed as (self-) organising itself to maximise model evidence. Here we are implicitly appealing to the Bayesian brain hypothesis [14] while gently sidestepping big questions about its utility and falsifiability (e.g., [15,16]). In what follows, we assume that the imperative to maximise model evidence is a (possibly tautological) truism [17] and consider the implications for functional anatomy. Our focus is on the notion of a **mean field approximation** that is an integral part of **approximate Bayesian inference**.

A key conclusion–that follows from the Bayesian brain–is that the structure of a good brain will recapitulate the (statistical) structure of how sensations are caused; in short, the model resides in the structure of the brain. For example, why does the brain have extensive connections while the liver seems to operate perfectly happily without them? An obvious answer is that the brain has to model sparse dependences induced by regularities in the world. In other words, our sensory inputs are generated by a small number of underlying causes that act on each other (usually at a distance) in a lawful and structured way. This lawful structure requires a relatively sparse dependency among the causes, such as gravity causing things to fall or visual objects causing sensory impressions. In short, the probabilistic structure of our world should, in principle, provide a sufficient explanation for the structure and fabric of connections of any brain that is trying to model that world. For example, our sensations are generated in a way that conforms to logarithmic rules (e.g., Weber's law). These statistical rules may then be transcribed into the lognormal statistics of synaptic physiology (implicit in divisive normalisation [18]) or the connectome that supports this physiology [19,20]. Simply noting that causal regularities in the world are transcribed into neuronal architectures may sound self-evident. However, this conjecture does not get to the heart of principles such as functional segregation. To understand how maximising model evidence leads to functional segregation, we have to consider the constraints under which evidence is optimised. This brings us to the notion of approximate Bayesian inference (Box 1).

Good Enough Brains and Approximate Bayesian Inference

Any system or procedure that optimises (maximises) Bayesian model evidence can be regarded as implementing Bayesian inference. However, exact Bayesian inference is generally impossible

Glossary of Bayesian terms

Approximate Bayesian inference: Bayesian belief updating in which approximate posterior distributions are optimised by minimising variational free energy, ensuring that the approximate posterior convergences to the true posterior.
Bayesian belief updating: the combination of prior beliefs about the causes of an observation and the **likelihood** of that observation producing a posterior belief about its hidden causes. This updating conforms to Bayes' rule.
Bayesian model evidence: this is the probability that some observations were generated by a model. It is also known as the marginal or integrated likelihood because it does not depend upon the hidden causes.
Complexity: the difference or divergence between prior and posterior beliefs. The complexity of a model reflects the change in prior beliefs produced by Bayesian belief updating (also known as Bayesian surprise).
Expectation: the mean or average (the first-order moment of a probability distribution).
Factorisation: decomposition of a quantity into the product of factors such that multiplying the factors reproduces the original quantity.
Generative model: a probabilistic specification of the dependencies among causes and consequences; usually specified in terms of a prior belief and the likelihood of observations, given their causes.
Hidden causes or states: the unobserved (including fictive) causes of observed data. They are hidden because they are random variables that can only be inferred from observations.
Likelihood: the probability of an observation under a generative model, given its causes.
Marginal: a marginal probability distribution of a joint distribution over random variables is obtained by marginalising or averaging over all of the variables apart from the variable of interest.
Mean field approximation: approximating a joint distribution over two or more random variables with the product of their marginal distributions.
Posterior beliefs: a probability distribution over the hidden causes of

in the real world, especially when modelling data generated by hierarchically deep, dynamic, and nonlinear processes. Almost invariably, this problem is solved with approximate Bayesian inference. Approximate Bayesian inference refers to optimisation in which an approximate representation (technically, a **posterior** probability distribution or 'belief') is made as similar as possible to the exact (Bayes-optimal) belief. There are many examples of approximate Bayesian inference; for instance, Bayesian filtering (also known as predictive coding [11]), which calls on a number of approximations. These assume that probabilistic beliefs have a particular distributional form (usually a Gaussian distribution). Another important assumption–that is ubiquitous in statistical physics and data analysis–is referred to as a mean field approximation [21,22]. Combining these two approximations leads to 'variational Bayes'. The key concept here is that the brain is faced with an important choice in the way that it optimises the very structure of its generative model (i.e., its connectivity) and associated 'beliefs' or inferences (i.e., the physiology supported by its connections).

observed consequences after they are observed.
Prior belief: a probability distribution over the hidden causes of observations before they are observed.
Variational free energy: a functional of a probability distribution (and observations) that upper bounds (is always greater than) the negative log evidence for a generative model. Negative log evidence is also known as surprise, surprisal, or self-information in information theory.

Put simply, the mean field approximation approximates dependencies among multiple factors with a product of **marginal** distributions that is much easier to deal with in terms of encoding and updating. A key challenge for approximate Bayesian inference is to find the right **factorisation** or marginalisation of beliefs about the causes of sensory input. Each possible factorisation or marginal representation corresponds to a different mean field approximation and a different way of 'carving nature at its joints' [23,24]. As scientists, we use this judicious 'carving' whenever we design a factorial experiment and test for interactions. In this case, the two factors represent a parsimonious hypothesis about how our data are caused, where the interaction reflects how one factor influences the expression of the other. Can this basic tenet of good statistical modelling be applied to neurobiology?

There are two levels that immediately come to mind. The first is the perspective afforded by the Bayesian brain and, in particular, the notion of perception as hypothesis testing [7]. In this instance, efficient perceptual synthesis reduces to an efficient and good factorisation of the putative causes of sensations. In other words, the brain has to learn about statistical independencies (technically, conditional independencies) to properly approximate the probabilistic structure of the sensorium. Experience and learning play a huge role in this process [25]. In addition to individual learning, evolution may act similarly on the brain. Evolution can also be formulated as learning statistical structure in the environment and distilling that structure into the phenotype [26–28]. Formal treatments of replicator dynamics and Fisher's fundamental theorem demonstrate that these evolutionary processes are nothing more or less than **Bayesian belief updating**. Natural selection itself has been likened to Bayesian model selection, where adaptive fitness corresponds to (variational) **free energy** [9,29].

Functional Segregation and Carving Nature at its Joints

The second level at which a good (enough) factorisation might be expressed is in terms of functional anatomy and segregation. In short, evolutionary (Bayesian belief) updates have shaped the brain into an efficient (minimum free energy) mean field approximation that we know and study as functional segregation [1,30,31].

A compelling example of the implicit division of labour is the factorisation of syntax and lexicosemantic statistics of language [32]. fMRI experiments demonstrate that Brodmann areas 45 and 47 respond not only to natural sentences but also to grammatically correct sentences without semantic content or meaning. This suggests a specialised role of these brain areas in syntactical organisation of semantic information [33]. By contrast, several other neocortical areas respond selectively to meaningful sentences but not to grammatically correct sentences without semantic information [34].

Perhaps the most celebrated example of transcribing statistical independencies into neuroanatomy is the segregation of dorsal and ventral visual processing streams of the brain [2,35,36]. The argument here is straightforward: if the causes of our visual sensations are visual objects that can be in different positions, the optimal way to factorise these causes is into where an object is and what an object is. The implicit conditional independence is simply a reflection of the fact that knowing where an object is does not (generally) tell you what it is. Technically, installing this conditional independence into functional anatomy enables the brain to maximise (Bayesian) model evidence. The alternative would be to have neuronal representations of every object in every location. Clearly, this would lead to a complex generative model with redundant degrees of freedom (connections), provided our world does indeed comprise objects in various locations (Box 2).

In practice, finding the right way to carve nature into the best factors ensures that the variational free energy is a better approximation to model evidence. In this view, the functional segregation of what and where streams embodies the fact that it is more efficient to encode where an object is and what an object is than to encode every combination of what and where. The predictions of current sensations then involve multiplying the probability distribution over where an object is by the probability distribution over what an object is (we return to the importance of this multiplication or interaction below). One could take this sort of argument further, in terms of hierarchal representations and special cases of variational inference cast in terms of information theory, leading to the principle of minimum redundancy, the principle of maximum efficiency, imperatives for sparse coding, and so on [37–40].

From a neurobiological perspective, this statistical carving (factorisation) corresponds to functional segregation [1,30]. If correct, this means that conditionally independent causes of our sensations correspond to the attributes that define functional specialisation; for example, motion, colour, and form [41]. In other words, natural selection, epigenetics, and experience-dependent plasticity equip the brain with the right sort of mean field approximation to infer the factors leading to sensations. For example, knowing an object's colour does not (generically) determine its motion. One could pursue this approach down to the level of classical receptive fields [40,42] and their contextual modulation (extraclassical receptive field effects) implied by the multiplication of marginal distributions to form precise posterior (probabilistic) beliefs. Following these examples, below we consider another, potentially more fundamental carving of statistical independencies that speaks not to what and where streams but to what and when systems–a dissection that may provide organising principles for more complex brains.

What and When: Functional Segregation of the Neocortex and Hippocampus

A pervasive and simple conditional independence that we deal with at all the time in perceptual synthesis and spatial navigation is the ordinal or temporal succession of events [43]. Here, succession *per se* can be separated from the constituent events. In other words, the concepts of 'first', 'last', 'slow', and 'fast' do not specify what is happening and are not content bound. This suggests a fundamental conditional independence between the temporal structure of succession (i.e., when) and the events that succeed each other (i.e., what). In exactly the same way that the brain may factorise **hidden causes** of sensations into what and where, it may apply the same marginalisation to what and when. This distinction may be even more pervasive than what and where; it might apply at multiple levels of abstraction and the very unfolding of experience itself. The attribute of where is limited to certain causes of our sensations. By contrast, social and physiological narratives (which may not be located to a particular point in extrapersonal space) always have a sequential aspect (e.g., inference, music, mathematical reasoning, language).

To paint this picture heuristically, consider two ways of encoding sequences. First, we could have a repertoire of sequential states for every sequence encountered; in other words, a

Box 2. Mean Field Approximations in the Brain

Figure I illustrates two ways of encoding a moving object in the visual field. In both cases, the visual input corresponds to an 'H' moving downwards. The left panels show a generative model encoding a joint representation over what an object is and where it is. To generate a sequence of observations, the generative model uses state transitions, from one state to the next where each hidden state determines the observed outcome. The **B** matrices encode state transitions, while **A** encodes a probabilistic mapping from states to outcomes. In the right-hand model, there is a separate representation for each object in every position and object motion simply entails transitions from the current object in one location to another (usually the same) object in the next location.

The right panel shows the equivalent model but under a mean field approximation in which the joint distribution is approximated by the product of marginal distributions over the factors what and where. Here, motion is generated by transitions from one location to the next while the object's identity remains unchanged. Crucially, the outcome rests on a product or multiplication of the two marginal representations. This is denoted by the Kronecker tensor product $\otimes$.

So which is the better model? If observations are generated by a world in which objects are invariant, the mean field approximation provides an accurate explanation for observed outcomes with the least complexity. This is because there are fewer hidden states (or degrees of freedom) than in the joint representation. Because the same accuracy is obtained with a lower complexity, this model will have more evidence and will be selected during natural (Bayesian model) selection (Box 1). Conversely, in a system with no pressure to be efficient (i.e., no limitation in size or energy expenditure) in which the identity of a moving object can change instantaneously, the joint model can generate accurate predictions. For example, an instantaneous switch from 'H' to 'T' after the first observation cannot be modelled under the mean field approximation (indicated by the red arrow). In such an imagined world, the joint model will justify its extra complexity by providing more accurate explanations for observations. However, in a real world, it is overly complex, with a redundant or inefficient parameterisation [37].

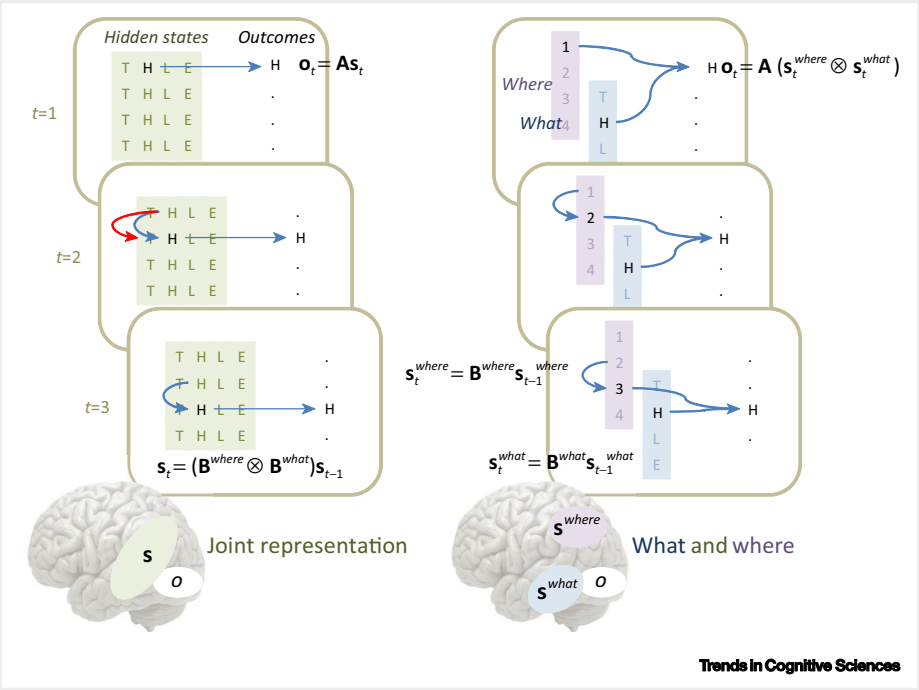

Figure I. Functional Segregation and Marginal Representations.

separate representation for every point in a sequence. This would be like having a library of sentences that we could call on to make sense of written text. The alternative to activating sequences of representations would be to have representations of sequences whose content is indexed by knowing where we are in the sequence (Box 3 and [44]). This distinction is exactly the same as the distinction between the joint and marginal representations of what and where considered above. This distinction may sound subtle; however, the marginal (mean field) approximation is substantially less complex. This is because, instead of having to represent

Box 3. What and When Architectures

The schematic (Figure I) uses the same form as Box 2. However, we have replaced where with when (and letters with words). The argument for a mean field (factorised or marginal) representation is exactly the same but in this context we are generating sequences over time at the same location. The joint representation (left panel) has an explicit representation of every possible sequence (labelled A, B, . . .). A complicated probability transition matrix then mediates jumps among hidden states to generate a sequence of outcomes.

A more parsimonious generative model–that predicts the same sequences–is shown on the right. Here, there is no explicit representation of content but simply a representation of the ordinal structure or sequence *per se* (e.g., a sentence or context). All of the heavy lifting–in terms of predicting the next outcome–is done by the connections from each representation of the sentence and their interactions with connections from representations encoding sequential transitions. As in the what and where example, the what or context factor (e.g., sentence) does not change in time. Crucially, this means that the representation of a sequence is not a sequence of representations. It is this architecture (mean field approximation) that enables sequential representations to transcend the passage of time.

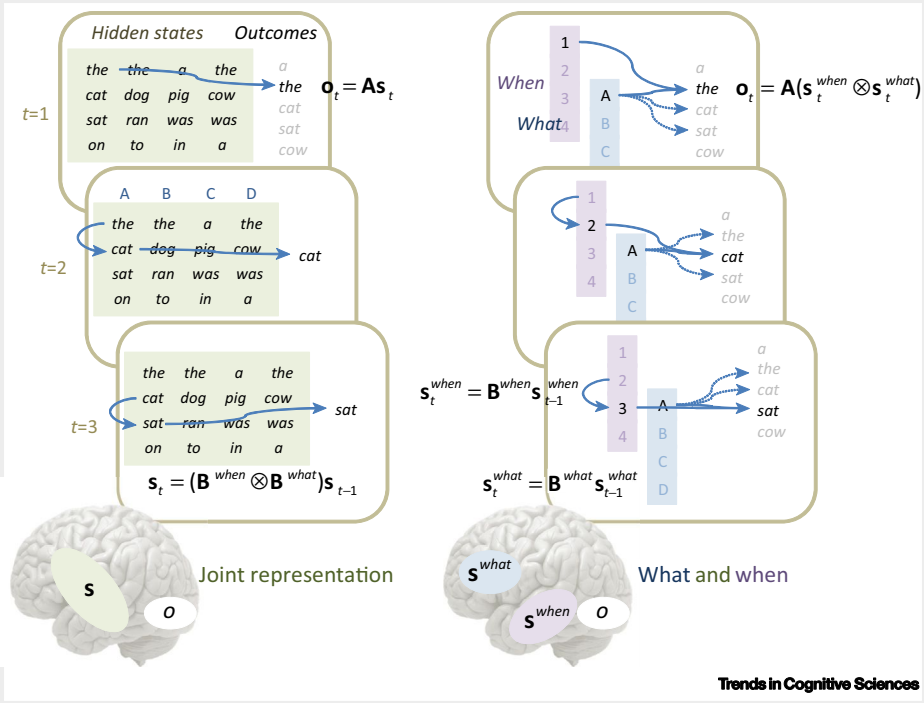

Figure I. What and When in the Brain.

hidden states or causes for every sentence (i.e., the number of words in a sentence multiplied by the number of sentences), we just have to represent sequential order and each sentence (i.e., the number of words per sentence plus the number of sentences).

If we call on a mean field approximation (functional segregation) of what and when, one would anticipate a generic architecture embodying the associated functional segregation. The natural candidate for this architecture is the distinction between brain structures of temporal succession [45,46], such as the hippocampus and cerebellum, and the content-encoding neocortex. Functional segregation also suggests that structures such as the hippocampus (when) should have the greatest divergent and convergent connectivity with representations of content (what). This may explain why the hippocampus appears to play the role of a hub [47] as opposed to modular neocortical areas. This connectivity places the hippocampus and paralimbic cortex at the centre of (centrifugal) hierarchical cortical connectivity [48,49].

Physiological Support for Model Predictions

If the hippocampus represents temporal succession or ordinal structure, one would expect to see sequential dynamics encoded by hippocampal neurons that are not bound to their content [50,51]. Self-generated sequences of neuronal firing patterns have been reported in the hippocampus [52], prefrontal cortex [53], and parietal cortex [54]. It therefore appears that the brain is (genetically) equipped with architectures that encode canonical or preconfigured sequences before those sequences are associated with (or bound to) any particular content [55]. This provides a somewhat counterintuitive prediction that one should see sequential dynamics, in systems like the hippocampus, before any particular experience [55,56]. This fits comfortably with recent observations that the neurons showing the greatest (sequential) firing rate modulations are impervious to the particular sequence of events experienced in the recent past [57]. Furthermore, experience with multiple sequences and with different content may engage the same canonical sequences, in the same way that the encoding of a spatial target in terms of its location is independent of its attributes [58].

This perspective may also explain the emergence of multiple place fields in the sequential encoding of temporal succession or order [59]. See [60,61] for related treatments of context in a Bayesian setting. In short, the picture that emerges here is of a neuronal representation of temporal succession that adumbrates any particular sequence such that content-free preexisting sequences are associated with a particular content through association with the activity of auxiliary units that show greater plasticity and context sensitivity [57].

The Statistics of Neuronal Encoding

An interesting aspect of the mean field approximation is that marginal probabilities have to be multiplied to generate joint distributions or distributions over outcomes. In statistics, a ubiquitous scheme for evaluating marginal distributions–known as belief propagation–is also called sum-product message passing. This is important because the realisation of the product of independent positive random variables is a lognormal process (this follows from the central limit theorem in the log domain). The implication for the statistics of neuronal encoding is that we might expect to see lognormal distributions of synaptic strengths, firing rates, and burst probabilities, under the assumption that spiking encodes the probability or **expectation** of occupying hidden states [19].

Predictions of the What and When Distinction: The Remembered Present

There is something quite distinct about representations with and without factorisation over time and content. An inspection of Box 3 reveals that the representations of context do not change with time. By contrast, with an exhaustive representation of both what and when, there is no temporal invariance and expectations cascade with the progress of time. Put simply, the first word in the sentence you are currently reading (i.e., 'Put') is always the same word before or after reading it. This means that sequential (when) states do all the heavy lifting in sequential processing, endowing contextual (what) representations with a form of translational invariance not in space but over time. Effectively, this converts a sequence of representations into the representation of a sequence. Heuristically, this means that the representation of a narrative, trajectory, or sequence of states is no longer tied to the present, enabling–and indeed mandating–an explicit representation of the past (i.e., memory) and future [58,62–66]. This intuition might explain why brain structures associated with memory are also implicated in planning [67,68]. This fits comfortably with recent suggestions that thinking about the past (and future) engages the same anatomical substrates and algorithms deployed for spatial navigation in the present [69–72].

In short, the factorisation into what and when necessarily entails a working memory that can accommodate postdiction and prediction. In this setting, postdiction corresponds to the

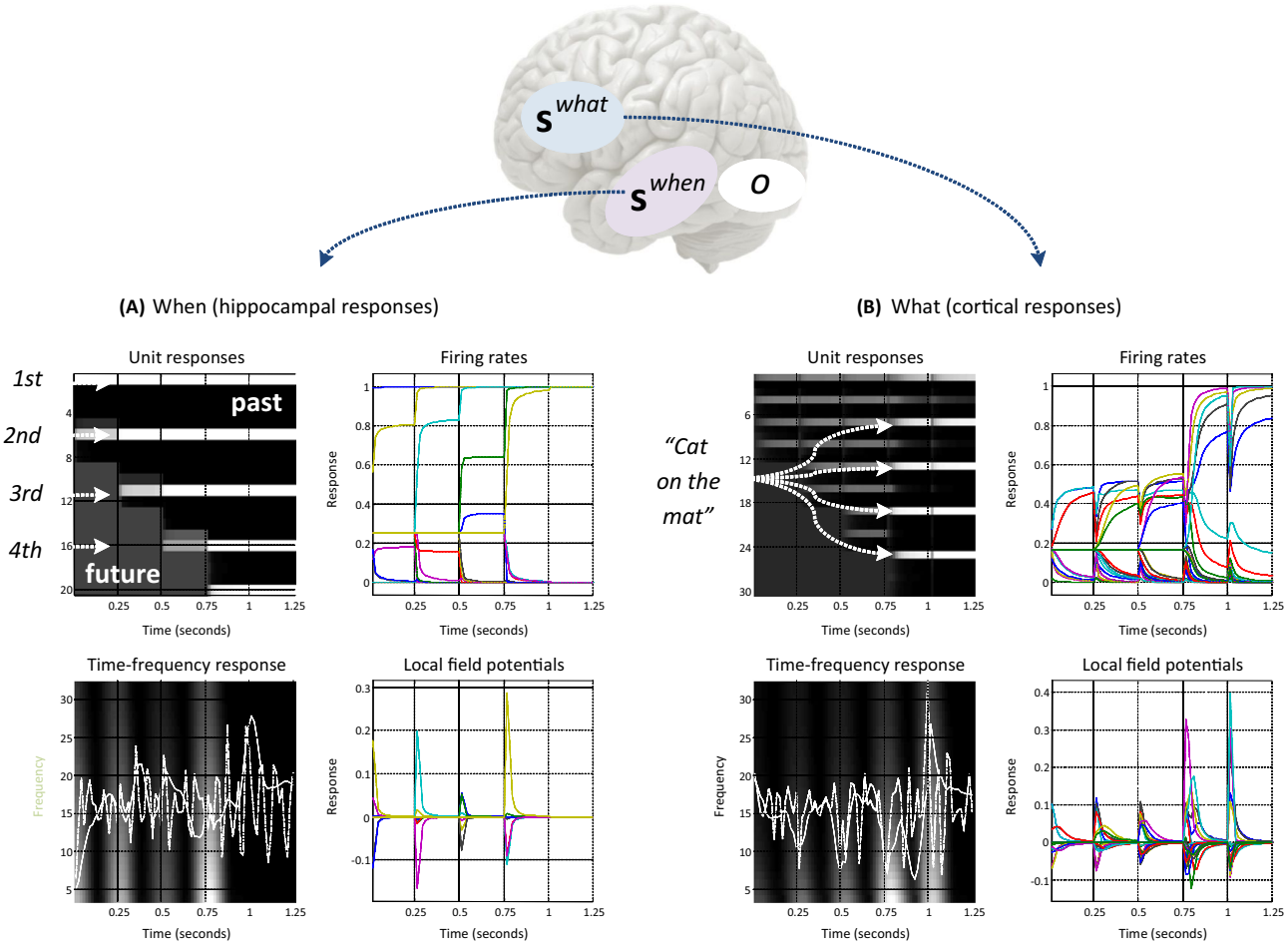

Figure 1. Simulated Electrophysiological Patterns. This figure illustrates hypothetical electrophysiological responses predicted by approximate Bayesian inference under a mean field assumption. These data report simulations of saccadic eye movements during reading, using the scheme described in [93]. In brief, we simulated saccadic eye movements sampling four successive 'words' under the hypothesis that the words were generated by one of six sentences. The generative model used to accumulate evidence assumed marginal distributions over the order of words (when) and the six alternative sentences (what). (A) Hippocampal responses. Shows responses based on a gradient descent on free energy for when expectations (with 16 iterations between each saccade). (B) Cortical Responses. Shows the equivalent results for what expectations. In this example, the first sentence was correctly inferred after the third word. The upper-left panels show the activity (firing rate) of units encoding hidden states in image (raster) format over the five epochs preceding saccades [(A) four ordinal states, (B) six sentences]. The first column reports all hidden states over all future time points at the beginning of the sequence while the rows encode their evolution as evidence is accumulated and processed. This means that the firing rates below the leading (block) diagonal effectively encode the future (planning) while those above encode beliefs about the past (i.e., memory). The upper-right panels plot the same information to illustrate evidence accumulation [89] and the resolution uncertainty about the context (i.e., sentence). The simulated the rate of change of neuronal firing is shown in the lower-right panel. The lower-left panels show average local field potentials over all spikes before (broken line) and after (unbroken line) filtering at 4 Hz superimposed on its time frequency decomposition. Please note that these simulated neuronal responses possess many features of empirical activity; for example, there is theta–gamma coupling [94–97] due to fast (gamma) activity elicited by each cue that is sampled at a slower (theta) frequency (lower-left panels). Such simulated results also show a characteristic phase precession [98,99] as predictions about the future are confirmed by sensory evidence. The above simulations can be reproduced with the DEM toolbox, available from http://www.fil.ion.ucl.ac.uk/spm.

accumulation of evidence for the current context (e.g., scene, location, sentence, story); namely, updating beliefs about the causes of previous sensory samples and, simultaneously, predictions about the future. For example, there may be sequences of neuronal populations in your brain that encode the current sentence in a way that predicts its conclusion. This representation changes much more slowly than the (e.g., predictive coding) processing of graphemes and word forms that are engaged by saccadic eye movements [73–75]. In this sense, carving the world

into ordered sequences–and the context under which those sequences unfold–provides a deep and hierarchical representation of time, as exemplified by the nested nature of the multitude of brain rhythms [76,77] (see also [78,79]). See Figure 1 for an example of simulated hippocampal responses during saccadic eye movements under this mean field assumption.

An important insight that can be drawn from what and where and what and when formulations (cf. Boxes 2 and 3) is that the coding roles of where and when become conflated in navigation (i.e., spatial sequencing). For example, the order of words during reading (but not listening) depends on where I am looking (Figure 1). This is remarkable since every principal neuron in the hippocampus can be regarded as either a 'place cell' [80] or a 'time cell' [51], as opposed to assigning time or space to distinct subsets of neurons. Ordinal sequencing (when) may include the where, with location as a special case of temporal organisation. Otherwise, one would need to postulate two separate systems or mechanisms for encoding the order of places and of time events. Whether neurons in the hippocampus and entorhinal cortex 'encode' position versus absolute time–or distance versus duration–depends largely on the testing conditions and the theoretical perspective of the observer [59,81–83]. One might anticipate that the what versus where distinction is (statistically and anatomically) conflated with the what versus when distinction, especially when dealing with trajectories in extrapersonal space (e.g., visual searches). Whether in space or time, ordinal sequences in the hippocampal system may 'index' the items (what) in the neocortex [84]. A marginal encoding of ordinal sequences (in time or space) and the semantics of ordered items (what) make the division of labour analogous to the role of a librarian (hippocampus, pointing to the items) in a library (neocortex, where semantic knowledge is stored). The organised access (in spatiotemporal trajectories) to neocortical representations (what) then becomes episodic information [85].

Concluding Remarks: Active Inference and Narratives

One could argue that we are simply putting a Bayesian gloss on hippocampal encoding of sequences. However, our proposition is simpler and subtler: the hippocampus–in contrast to other structures of succession such as the basal ganglia and cerebellum–has a privileged role; it encodes ordinal structure without reference to particular events. The content of the sequence depends on how events are 'bound' to content-free sequences through context-sensitive changes in synaptic efficacy. If this view is correct, one would expect to see intrinsic (sequential) dynamics in hippocampal activity even 'in the absence of any external memory demand or spatiotemporal boundary' [83]–a prediction that is now attracting empirical attention [50,83]. The broader empirical implication presents an avenue for falsifying the mean field hypothesis (see also [86]); namely, if a subset of hippocampal neurons encode the marginal probability of where they are in a sequence, one should be able to identify cells that are 'repurposed' for trajectories (e.g., in linear mazes) and insensitive to the particular environment or direction of travel. In other words, they should show a context invariance that speaks to conditional independence. The flipside of this thesis is that if neocortical firing patterns encode context, they should be more enduring (e.g., delay period activity in the prefrontal cortex [87], evidence accumulation in the parietal cortex [88,89]).

Several interesting predictions follow from this perspective. For example, place-cell activity is typically identified by correlating neuronal responses with the current location of an animal. However, if the hippocampus encodes both time (when) and space (what), the activity of neurons encoding the first and subsequent places visited should accumulate evidence over the duration of the sequence. This means that one should be able to find neuronal patterns whose activity is predicted not by the current location but by where the animal started–and where it is going [90].

The Bayesian brain falls short in explaining how the brain creates new knowledge and adds emotional charge to implicit 'representations'. After all, the brain does not 'represent' but creates

our world. Yet, 'good enough' brains can approximate the world in a fast and efficient way and may be a prerequisite for constructivist or enactive inference. An interesting corollary of the Bayesian perspective on mnemonic representation of sequences is that beliefs about the future are tied to beliefs about the past. If we act on these beliefs, we create a (non-Markovian) world with rich temporal structure. This follows because, in the absence of any action of the brain on the world, the succession of worldly states can be predicted completely by the laws of nature (e.g., Hamilton's principle of least action and classical mechanics). Crucially, these laws are compatible with a Markovian world in which the next state depends only on the previous state. However, if we now put history dependence into the mix the world becomes much more interesting. In short, the way we represent ordinal succession and the implicit narratives that predict and explain our senses lead inevitably to behaviour that transcends the rules of classical physics (see Outstanding Questions).

Outstanding Questions

How far can one take the mean field explanation for functional segregation in the brain? Would it help in understanding the marginalisation implicit in hierarchical cortical structures? Can it be applied at the scale of canonical microcircuitry?

Is there clear evidence for hippocampal cell activity that encodes the purely ordinal features of navigable trajectories (i.e., cells that always fire in the third location irrespective of place). Similarly, does the phase precession seen in place-cell activity represent sequential encoding or the interaction between ordinal and place representations implicit in a mean field approximation?

What is the relationship between delay period activity and (neocortical) representations of sequence content? In other words, can we understand working memory as a representation of the hidden causes that are sequentially disclosed by sensory sampling?

Could we understand evidence accumulation in terms of perceptual inference, where alternative (what) hypotheses are called on to interact with ordinal (when) representations to provide inference to the best explanation (for sensory input).

To what extent are the fundamental features of temporal succession preordained or preconfigured (genetically or epigenetically) and to what extent are they learned during neurodevelopment?

Are there natural timescales for temporal succession? For example, can the canonical bounds on working memory be related to nested rhythms in the brain? In this regard, do hippocampal–prefrontal interactions play a special role in providing neurophysiological constraints on the encoding of sequences?

If all sequences share an ordinal factor, would we expect to see neuronal structures involved in spatial navigation, language, or any activity that entails a systematic progression through ordinal sets?

Acknowledgments

K.F. is funded by the Wellcome Trust (ref/088130/Z/09/Z) and G.B. by the NIMH (MH107396, MH102840). The authors thank Sam McKenzie, Dan Levenstein, and Brendon Watson for comments on the manuscript. They also thank the Hungarian Academy of Sciences for hosting a lunch during which this Opinion article was conceived.

References

1. Zeki, S. and Shipp, S. (1988) The functional logic of cortical connections. *Nature* 335, 311–317
2. Ungerleider, L.G. and Mishkin, M. (1982) Two cortical visual systems. In *In Analysis of Visual Behavior* (Ingle, D. *et al.*, eds), pp. 549–586, MIT Press
3. Kersten, D. *et al.* (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304
4. Dayan, P. *et al.* (1995) The Helmholtz machine. *Neural Comput.* 7, 889–904
5. Hasselmo, M.E. and Stern, C.E. (2015) Current questions on space and time encoding. *Hippocampus* 25, 744–752
6. Helmholtz, H. (1866/1962) Concerning the perceptions in general. In *In Treatise on Physiological Optics, Dover*.
7. Gregory, R.L. (1980) Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197
8. Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204
9. Friston, K. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138
10. Ballard, D.H. *et al.* (1983) Parallel visual computation. *Nature* 306, 21–26
11. Rao, R.P. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87
12. Ashby, W.R. (1947) Principles of the self-organizing dynamic system. *J. Gen. Psychol.* 37, 125–128
13. Conant, R.C. and Ashby, W.R. (1970) Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* 1, 89–97
14. Knill, D.C. and Pouget, A. (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719
15. Davis, R.H. (2006) Strong inference: rationale or inspiration? *Perspect. Biol. Med.* 49, 238–250
16. Bowers, J.S. and Davis, C.J. (2012) Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414
17. Friston, K. (2013) Life as we know it. *J. R. Soc. Interface* 10, 20130475
18. Reynolds, J.H. and Heeger, D.J. (2009) The normalization model of attention. *Neuron* 61, 168–185
19. Buzsáki, G. and Mizuseki, K. (2014) The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* 15, 264–278
20. Markov, N. *et al.* (2013) Cortical high-density counterstream architectures. *Science* 342, 1238406
21. Jaakkola, T. and Jordan, M. (1998) Improving the mean field approximation via the use of mixture distributions. In *Learning in Graphical Models* (Jordan, M., ed.), pp. 163–173, Springer
22. Buice, M.A. and Cowan, J.D. (2009) Statistical mechanics of the neocortex. *Prog. Biophys. Mol. Biol.* 99, 53–86
23. Couchman, J.J. *et al.* (2010) Carving nature at its joints using a knife called concepts. *Behav. Brain Sci.* 33, 207–208
24. Gershman, S.J. and Niv, Y. (2010) Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 251–256
25. Buzsáki, G. (1998) Memory consolidation during sleep: a neurophysiological perspective. *J. Sleep Res.* 7 (Suppl. 1), 17–23
26. Paulin, M.G. (2005) Evolution of the cerebellum as a neuronal machine for Bayesian state estimation. *J. Neural Eng.* 2, S219–S234
27. Fernando, C. *et al.* (2012) Selectionist and evolutionary approaches to brain function: a critical appraisal. *Front. Comput. Neurosci.* 6, 24
28. Harper, M. (2011) Escort evolutionary game theory. *Physica D* 240, 1411–1415
29. Sella, G. and Hirsh, A.E. (2005) The application of statistical physics to evolutionary biology. *Proc. Natl Acad. Sci. U.S.A.* 102, 9541–9546
30. Tononi, G. *et al.* (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl Acad. Sci. U.S.A.* 91, 5033–5037
31. Park, H.J. and Friston, K. (2013) Structural and functional brain networks: from connections to cognition. *Science* 342, 1238411
32. Lees, R.B. (1957) Review of Chomsky, 1957. *Language* 33, 375–408
33. Tyler, L.K. *et al.* (2010) Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. *Cereb. Cortex* 20, 352–364
34. Pallier, C. *et al.* (2011) Cortical representation of the constituent structure of sentences. *Proc. Natl Acad. Sci. U.S.A.* 108, 2522–2527
35. Ungerleider, L.G. and Haxby, J.V. (1994) 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165
36. Goodale, M.A. *et al.* (2004) Two distinct modes of control for object-directed action. *Prog. Brain Res.* 144, 131–144
37. Barlow, H. (1961) Possible principles underlying the transformations of sensory messages. In *In Sensory Communication* (Rosenblith, W., ed.), pp. 217–234, MIT Press

38. Linsker, R. (1990) Perceptual neural organization: some approaches based on network models and information theory. *Annu. Rev. Neurosci.* 13, 257–281
39. Optican, L. and Richmond, B.J. (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior cortex. II Information theoretic analysis. *J. Neurophysiol.* 57, 132–146
40. Olshausen, B.A. and Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609
41. Zeki, S. (2005) The Ferrier Lecture 1995. Behind the seen: the functional specialization of the brain in space and time. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1145–1183
42. Angelucci, A. and Bressloff, P.C. (2006) Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.* 154, 93–120
43. Zucker, H.R. and Ranganath, C. (2015) Navigating the human hippocampus without a GPS. *Hippocampus* 25, 697–703
44. Dehaene, S. *et al.* (2015) The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88, 2–19
45. Verschure, P.F.M.J. and Edelman, G.M. (1992) The remembered present: a biological theory of consciousness. *Am. J. Psychol.* 105, 477
46. Buzsáki, G. (2010) Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68, 362–385
47. Wittner, L. *et al.* (2007) Three-dimensional reconstruction of the axon arbor of a CA3 pyramidal cell recorded and filled *in vivo*. *Brain Struct. Funct.* 212, 75–83
48. Felleman, D. and Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47
49. Mesulam, M.M. (1998) From sensation to cognition. *Brain* 121, 1013–1052
50. Itskov, V. *et al.* (2011) Cell assembly sequences arising from spike threshold adaptation keep track of time in the hippocampus. *J. Neurosci.* 31, 2828–2834
51. Eichenbaum, H. (2014) Time cells in the hippocampus: a new dimension for mapping memories. *Nat. Rev. Neurosci.* 15, 732–744
52. Pastalkova, E. *et al.* (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321, 1322–1327
53. Fujisawa, S. *et al.* (2008) Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–833
54. Harvey, C.D. *et al.* (2012) Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* 484, 62–68
55. Mizuseki, K. and Buzsáki, G. (2013) Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell Rep.* 4, 1010–1021
56. Dragoi, G. and Tonegawa, S. (2014) Selection of preconfigured cell assemblies for representation of novel spatial experiences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20120522
57. Grosmark, A.D. and Buzsáki, G. (2016) Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science* 351, 1440–1443
58. Manns, J.R. *et al.* (2007) Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56, 530–540
59. Buzsáki, G. (2013) Cognitive neuroscience: time, space and memory. *Nature* 497, 568–569
60. Fuhs, M.C. and Touretzky, D.S. (2007) Context learning in the rodent hippocampus. *Neural Comput.* 19, 3173–3215
61. Gershman, S.J. *et al.* (2010) Context, learning, and extinction. *Psychol. Rev.* 117, 197–209
62. Eichenbaum, H. and Fortin, N.J. (2009) The neurobiology of memory based predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1183–1191
63. Schacter, D.L. *et al.* (2008) Episodic simulation of future events–concepts, data, and applications. In *In Year in Cognitive Neuroscience 2008* (Kingstone, A. and Miller, M.B., eds), pp. 39–60, Wiley-Blackwell
64. Scoville, W.B. and Milner, B. (1957) Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21
65. Squire, L.R. (1992) Memory and the hippocampus–a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* 99, 195–231
66. Dragoi, G. and Buzsáki, G. (2006) Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron* 50, 145–157
67. Buckner, R.L. (2010) The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol.* 61, 27–48
68. Epstein, R. *et al.* (1999) The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23, 115–125
69. Buzsáki, G. and Moser, E.I. (2013) Memory, navigation and theta rhythm in the hippocampal–entorhinal system. *Nat. Neurosci.* 16, 130–138
70. Hassabis, D. and Maguire, E.A. (2007) Deconstructing episodic memory with construction. *Trends Cogn. Sci.* 11, 299–306
71. Zeidman, P. *et al.* (2015) Investigating the functions of subregions within anterior hippocampus. *Cortex* 73, 240–256
72. Izquierdo, I. and Medina, J.H. (1997) Memory formation: the sequence of biochemical events in the hippocampus and its connection to activity in other brain structures. *Neurobiol. Learn. Mem.* 68, 285–316
73. Rayner, K. (2009) Eye movements in reading: models and data. *J. Eye Mov. Res.* 2, 1–10
74. Friston, K. *et al.* (2012) Perceptions as hypotheses: saccades as experiments. *Front. Psychol.* 3, 151
75. Pierrotdeseilligny, C. *et al.* (1995) Cortical control of saccades. *Ann. Neurol.* 37, 557–567
76. Buzsáki, G. *et al.* (2013) Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764
77. Buzsáki, G. and Draguhn, A. (2004) Neuronal oscillations in cortical networks. *Science* 304, 1926–1929
78. George, D. and Hawkins, J. (2009) Towards a mathematical theory of cortical micro-circuits. *PLoS Comput. Biol.* 5, e1000532
79. Kiebel, S.J. *et al.* (2008) A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4, e1000209
80. O'Keefe, J. and Dostrovsky, J. (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175
81. Schiller, D. *et al.* (2015) Memory and space: towards an understanding of the cognitive map. *J. Neurosci.* 35, 13904–13911
82. Kraus, B.J. *et al.* (2015) During running in place, grid cells integrate elapsed time and distance run. *Neuron* 88, 578–589
83. Villette, V. *et al.* (2015) Internally recurring hippocampal sequences as a population template of spatiotemporal information. *Neuron* 88, 357–366
84. Teyler, T.J. and DiScenna, P. (1986) The hippocampal memory indexing theory. *Behav. Neurosci.* 100, 147–154
85. Tulving, E. (1987) Multiple memory systems and consciousness. *Hum. Neurobiol.* 6, 67–80
86. Hasselmo, M.E. (2015) If I had a million neurons: potential tests of cortico–hippocampal theories. *Prog. Brain Res.* 219, 1–19
87. Kojima, S. and Goldman-Rakic, P.S. (1982) Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. *Brain Res.* 248, 43–49
88. Huk, A.C. and Shadlen, M.N. (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* 25, 10420–10436
89. Kira, S. *et al.* (2015) A neural implementation of Wald's sequential probability ratio test. *Neuron* 85, 861–873
90. Wikenheiser, A.M. and Redish, A.D. (2015) Hippocampal theta sequences reflect current goals. *Nat. Neurosci.* 18, 289–294
91. Fox, C. and Roberts, S. (2012) A tutorial on variational Bayes. *Artif. Intell. Rev.* 38, 85–95
92. Penny, W.D. (2012) Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage* 59, 319–330
93. Friston, K. *et al.* (2015) Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214

94. Canolty, R.T. *et al.* (2006) High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313, 1626–1628

95. Lisman, J. and Redish, A.D. (2009) Prediction, sequences and the hippocampus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1193–1201

96. Lisman, J. and Buzsáki, G. (2008) A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophr. Bull.* 34, 974–980

97. Bragin, A. *et al.* (1995) Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* 15, 47–60

98. Skaggs, W.E. *et al.* (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6, 149–172

99. O'Keefe, J. and Recce, M.L. (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–330