

# The secret life of predictive brains: what's spontaneous activity for?

Giovanni Pezzulo<sup>1</sup>, Marco Zorzi<sup>2,3</sup>, Maurizio Corbetta<sup>4,5</sup>

1. Institute of Cognitive Sciences and Technologies, National Research Council, Roma, Italy
2. Department of General Psychology and Padova Neuroscience Center (PNC), University of Padova, Padova, Italy
3. IRCCS San Camillo Hospital, Venice, Italy
4. Department of Neuroscience and Padova Neuroscience Center (PNC), University of Padova, Padova, Italy
5. Venetian Institute of Molecular Medicine (VIMM), Fondazione Biomedica, Padova, Italy

Correspondence: giovanni.pezzulo@istc.cnr.it (G. Pezzulo)

## Abstract

Brains at rest generate dynamical activity that is highly structured in space and time. We suggest that spontaneous activity, as in rest or dreaming, underlies top-down dynamics of generative models. During active tasks, generative models provide top-down predictive signals for perception, cognition, and action. When the brain is at rest and stimuli are weak or absent, top-down dynamics optimize the generative models for future interactions by maximizing the entropy of explanations and minimizing model complexity. Spontaneous fluctuations of correlated activity within and across brain regions may reflect transitions between "generic priors" of the generative model: low dimensional latent variables and connectivity patterns of the most common perceptual, motor, cognitive, and interoceptive states. Even at rest, brains are proactive and predictive.

**Keywords:** spontaneous activity; generative models; resting state; predictive brains

## Highlights box

- Spontaneous brain dynamics are manifestations of top-down dynamics of generative models detached from action-perception cycles
- Generative models constantly produce top-down dynamics, but we call them expectations and attention during task engagement and spontaneous activity at rest
- Spontaneous brain dynamics during resting periods optimize generative models for future interactions, by maximizing the entropy of explanations in the absence of specific data and reducing model complexity
- Low-frequency brain fluctuations during spontaneous activity reflect transitions between generic priors consisting of low-dimensional representations and connectivity patterns of the most frequent behavioral states
- High-frequency fluctuations during spontaneous activity in the hippocampus and other regions may support generative replay and model learning

## What is the brain doing when it (apparently) does nothing?

Brains are constantly active, even when they receive no sensory stimuli or are not task-engaged [1,2]. Studies of the brain at rest or in periods of relative inactivity reveal that sophisticated dynamical patterns of activity emerge spontaneously across cortical and subcortical structures [3,4] (Figure 1). The functional significance of spontaneous brain activity (also called intrinsic or endogenous) is largely unknown.

Mainstream models of cortical processing describe the brain as a sensory-motor analyzer [5]. In these models, the content of sensory representations only becomes available through feedforward processing, as photons reflected from objects hit the retina and are transduced into increasingly complex features in visual cortex (e.g., edges, geometric features, object parts, faces), as exemplified in convolutional deep neural networks [6]. Endogenous processes such as attention modulate the feedforward transduction, but do not elicit or maintain representational content in the absence of stimuli [7,8]. Within this traditional view, spontaneous activity dynamics are modeled as random noise fluctuations that modulate feedforward post-synaptic recruitment [9,10].

However, it is now well established that neural noise is correlated in space and time, and different explanations have been offered to explain spontaneous patterns of correlated activity among brain regions. One idea is that they represent emergent properties of anatomical circuitries driven by noise [11,12]. Other theories suggest that they represent homeostatic mechanisms related to the history of task activation [13–15] and possibly serve as a scaffold (or prior) for the execution of behaviorally relevant tasks [3,14–18].

Nonetheless, we experience a rich mental life event in the absence of any task or sensory stimulation. For example, when dreaming of people, we engage in sophisticated forms of cognition spanning from theory of mind to moral reasoning and language. When dreaming of actions like flying or falling, we anticipate their outcome like preparing to hit the floor and even experience feelings of vertigo. Similarly, lying awake in bed in the dark, we can recall events from the past, but also create novel events in the future by flexibly recombining bits of old memories or creating novel ones never experienced before. Hence, in the absence of any task, sensory stimulus and even consciousness, the brain can generate spontaneously complex scenes and manipulate them in cognitively sophisticated ways – prompting a reconsideration of the richness of the brain’s “secret life”.

## Four indications of the functional importance of spontaneous activity

At least four research lines suggest the functional importance of **spontaneous brain activity**.

First, the brain consumes about 20% of the body energy, upwards of 50% in the developing brain, vis-à-vis only 2% of the body weight [1,19]. While 25-40% of this budget goes to house-keeping functions (e.g. microtubule remodeling, lipid turnover, and protein synthesis), most (60-75%) goes to electrochemical signals: resting potentials (neurons, glia), and signaling (spikes, subthreshold excitatory and inhibitory potentials) [20]. However spikes are metabolically expensive, and given a fixed metabolic budget, only about 10% of the energy goes to spikes with very few neurons (1-10%) simultaneously firing in cortex during sensory processing [21]. This explains the relatively small changes of metabolism and blood flow measured in PET and fMRI experiments during task activation [1]. Theoretical calculations and recent simultaneous neuronal recordings from thousands of neurons show that the activity of most neurons concentrates in a more metabolically convenient infra-slow frequency range ( $<0.1$  Hz) [22–24]. These observations agree with the prominent infra-slow fluctuations of activity ( $<0.1$  Hz) measured from the whole brain with fMRI blood oxygenation level dependent (BOLD) and EEG signals [25] (Figure 1A; Figure 2A). If most of the energy budget goes to resting potentials and infra-slow signaling, this activity must play a fundamental, yet unknown, role in function. In this article, we use “spontaneous activity” to refer to infra-slow activity, unless otherwise specified.

The second clue on the functional importance of spontaneous activity comes from studies of its spatial and temporal organization. Brain activity at rest (either eyes open or closed), measured through the BOLD fMRI signals, is organized in distinct spatiotemporal patterns known as **resting state networks** (RSNs). These networks are formed by groups of regions that show temporally correlated activity and co-activation during behavioral tasks [14,26]. The functional architecture of RSNs reflects anatomical connectivity, especially in sleep and anesthesia, but in the awake state the similarity between anatomical and functional connectivity patterns is low [11,12]. These findings suggest that arousal and cognition enrich a static anatomical architecture. In the temporal domain, RSNs fluctuate at slow frequencies ( $\sim 0.01$  Hz) as measured in EEG [27], MEG [28], local field potentials [29], and calcium imaging [23,24]. Also in time the electrophysiological topography of RSNs is relatively stable during task performance albeit frequency specific interactions between task-specific regions occur [30].

Third, spontaneous activity patterns covary with individual differences in cognitive functions [31,32], change with learning [13,33], and predict individual predisposition to a new behavior [34]. An enormous clinical literature has reported correlations between behavioral abnormalities in innumerable clinical conditions and alterations of spontaneous activity patterns. Particularly insightful are alterations that show both network- and behavior-specific correlations, such as a double dissociation in the motor and dorsal attention networks, respectively for motor and spatial attention deficits [35].

Fourth, spontaneous and task-evoked activity are similar at the level of single neurons within cortical and subcortical regions. The statistics of spontaneous activity in sensory cortex are well aligned to the statistics of visual or auditory scenes [36,37] and this alignment increases with development [38,39]. Task-evoked and spontaneous activity influence each other: visually evoked response reverberate in spontaneous activity [40], while spontaneous activity explains the variability of stimulus evoked responses [41]. This is not only a cortical phenomenon. Single cell recordings in rodents [42–45] and human neuroimaging studies [46,47] show that spontaneous activity in the hippocampus during sleep or wakeful rest resembles sequences of neuronal activation for places or events experienced in the awake state (“replays” [48]). Furthermore, in certain conditions, spontaneous activity before navigation of novel mazes resembles sequential patterns observed during subsequent navigation (“preplays” [49]). Spontaneous hippocampal sequences are not epiphenomenal. Their manipulation through optogenetics changes memory content and spatial decisions [50–52]. Replay sequences involve high frequency activity ( $>150$  Hz) but can be temporally coordinated with low frequency spontaneous dynamics in other brain structures, such as prefrontal cortex, visual cortex and ventral striatum, possibly supporting systems-level cognitive functions [53–56].

In summary, spontaneous activity (both in individual regions and whole brain networks) play a fundamental, yet poorly understood, role in brain function. Here we propose that spontaneous brain activity patterns, such as those observed during resting state and hippocampal replays, may be the signature of a class of computational models called **generative models** [57] and the top-down and recurrent dynamics that they produce.

### Top-down brain processes and spontaneous activity: a generative modeling perspective

Top-down and recurrent brain dynamics have been mostly studied during task performance, where they may support prediction, attention, and the contextualization of stimulus processing [17,58–61]. Top-down processes are triggered by internal (e.g. the memory we are out of fruit at home) or external (e.g. the sign of a grocery store) signals and can be maintained for long periods of time to guide visual exploration and perception; in part, by modulating ongoing rhythms [62,63]. At the moment of search, they modulate sensory representations toward relevant/attended stimuli (e.g. apples) and away from irrelevant ones (e.g. breads, meats) [7,8].

According to an influential perspective, top-down dynamics may reflect the brain’s generative model of the environment and its own body [64,65]. From a generative modeling perspective, perception, action selection, and learning are problems of statistical inference. For example, perception corresponds to the unconscious inference [66] of the latent (i.e., unobserved) causes of our observations; e.g., which object (e.g., apple or frog) caused our retinal activity patterns (e.g., red rounded or green elongated). This problem is solved using a generative model that comprises two elements: a prior probability of the objects in the visual scene, which specifies how probable are apples or frogs in the context; and a (likelihood) mapping between objects and stimuli, which specifies which stimuli are more likely given the presence of a specific object (e.g., that seeing red is more likely under the apple than the frog hypothesis). **Bayes’ rule** prescribes how to use these two elements plus sensory data (e.g., red stimuli) to infer the posterior probability of apple and frog hypotheses – hence adjudicating between them.

In the most widely used biological implementation of the above inferential scheme – **predictive coding** – top-down activity updates perceptual representations by integrating top-down predictions and bottom-up stimuli (Box 1). Similarly, in the hierarchical recurrent neural network known as **Deep Boltzmann Machine** (Figure 3), top-down signals generate fictive data that guide learning and later convey information that help resolve uncertainty at lower levels during inference [67]. The above theories focus on top-down processes during online tasks, not spontaneous activity at rest. However, this dichotomy may be only apparent. If brain networks implement generative models, they will constantly perform inferential and generative processes, not only for cognition and sensory-motor interactions, but also during rest [17,44,68].

Our proposal highlights however a fundamental difference between generative processes during task performance and rest. During task performance, top-down processes prepare the brain to process specific stimuli and upcoming actions. For example, top-down signals in visual cortex encode the location, feature, objects, and decisions for expected stimuli

at the expense of unexpected stimuli [8,69,70]. Hence, preparatory signals during active behavior encode context-specific task information. In contrast, spontaneous activity dynamics at rest are not confined to recapitulating the history of task activations, but prepare the brain to process a wider range of stimuli and responses for future behavior, largely in a context-independent manner.

Formally, the disengagement from current action-perception cycles allows the optimization of generative models for future use. Statistical inference balances constantly two imperatives: maximizing the accuracy of explanations (to fit data) and minimizing model complexity (to avoid overfitting). However, the balance of accuracy and complexity may change between task and resting states. During resting states, when stimuli are weak or absent, accuracy loses importance, favoring the reduction of model complexity. Furthermore, the statistical framework mandates that in the absence of data we should adopt maximally uncertain beliefs about the state of the world, that is, maximize the entropy of explanations and “keep options open” (i.e., maximum entropy principle [71]).

In the next two sections, we unpack this hypothesis. We argue that the imperative to maximize entropy of explanations leads to the formation of generic priors, which become apparent as transitions between low-dimensional brain states. Furthermore, the imperative to minimize model complexity may govern model reduction and synaptic pruning.

### **Inferring generic and low-dimensional spatiotemporal priors for future interactions**

A key function of spontaneous activity during offline periods may be the inference of generic and low-dimensional spatiotemporal priors for future interactions.

In predictive coding and similar schemes, priors encoded at high hierarchical levels are propagated top-down to predict incoming stimuli at lower levels; and are corrected if prediction errors are generated. During offline periods, bottom-up stimuli are weak or absent; they cannot elicit prediction errors to correct priors, which are therefore continuously reiterated. During these periods, the brain's generative model may recirculate (or resample) the model's priors, or the spatiotemporal patterns acquired during sensorimotor behavior and the exposition to external stimuli [13,17,72–74]. This hypothesis explains the resemblance (at the level of average statistics) between brain activations during spontaneous and evoked cortical activity [31,36–39,41,75–77] but does not fully specify the content of the priors. We argue that offline activity reflects two kinds of priors: representations and connectivity patterns. For simplicity, we discuss them separately, even though they are closely integrated in the brain [78].

#### ***Representations as generic priors***

During offline periods, there are fewer stimuli to “explain away” and this may favor the formation of generic priors, i.e., prior representations not tied to the explanation of any specific data point, but to the data distribution and their natural statistics, such as the distribution of natural image features, objects, behavioral and interoceptive patterns previously experienced [68,79]. Within generative neural networks, generic priors may correspond to information-compressed, low dimensional states that summarize a large amount of information, abstracting away from specific stimuli [57]. Because hierarchical networks operate successive information compressions, they encode simpler to more complex (e.g., time- and position-invariant) summary representations at increasingly higher hierarchical levels – and can recycle them across cognitive tasks [80]. Generic priors therefore embed low dimensional representations of perceptual, cognitive, or motor patterns that summarize the statistics of various natural tasks and have general use, for example, prototypes of object categories or features rather than individual instances, or action pattern categories and synergies rather than individual movements. While the ontology of this low dimensional representation is unknown, anatomical considerations suggest that the information compression may be operated along a hierarchy that has primary sensory and motor regions (highly granular cortex) at its bottom, and less granular heteromodal regions (including limbic cortices) at its apex. The fact that limbic cortices lie at the apex of the architectural gradient suggests that the brain's generative model may organize information along behaviorally relevant and interoceptive dimensions (Box 2) and model “the body in the world” [18] rather than just reflect exteroceptive statistics, as often assumed [17,38]. In keeping, the notion of generic prior used here is broad and encompasses environmental, behavioral, interoceptive and cognitive patterns (see Box 2).

At the neuronal level, growing evidence indicates that stimuli or movements in cortex are not coded individually (e.g. a neuron coding for a specific face [81]), but through low dimensional hidden variables that summarize the variability of individual stimuli or actions. For instance, activity in monkey inferotemporal cortex can be described as the weights of principal components that represent the variability of hundreds of face or object stimuli [82]. The same occurs in monkey motor cortex for complex movement [83]. Similarly, in humans the fMRI response of association visual cortex to many different stimuli can be summarized by relatively few dimensions (e.g. animacy vs. non-animacy, [84]). In motor cortex,

the response to principal components of complex hand movements summarize the fMRI response to individual movements [85]. In association cortex, a similar low dimensionality has been shown for semantic stimuli [86,87].

This compression strategy is efficient because the apparent heterogeneity of behavioral patterns in our everyday tasks may hide a much simpler lower dimensionality of input and output statistics (Box 2). Critically, this low dimensionality of neural (and behavioral) patterns shall be also apparent at rest. Accordingly, spontaneous activity in mouse visual cortex replays principal components of exploratory face movements [23] (Figure 2A). In human association visual cortex spontaneous multivoxel fMRI activity represents more likely domain-specific prototype-like stimuli rather than individual exemplars [77] (Figure 2B). In human motor cortex common hand movements are more likely to be represented in multi-voxel resting activity than novel uncommon hand movements [88] and learned motor sequences are replayed in multi-unit firing activity at rest [89]. Similar observations have been made in cat [36] and monkey visual cortex [76]. Critically, these spontaneous patterns are not necessarily sensory-specific in sensory cortex or motor-specific in motor cortex, but may reflect more general patterns that are statistically associated during natural behavior and linked by structural-functional connections. For instance, regions of visual cortex that respond to a specific visual category (bodies) also respond during reaching movements [90].

In summary, we posit that the low dimensionality of spontaneous neural patterns may reflect slow transitions between generic priors tuned for low dimensional, general-purpose, behavioral patterns [91,92] (Figure 2D). Itinerant dynamics across sequences of brain states may reflect the maximization of the entropy of explanations [93] (or in dynamical systems parlance, the exploration of the brain's dynamic repertoire [94]) that generative models privilege in the absence of data to explain – to remain flexible.

### *Connectivity patterns as priors*

The notion of generic priors includes not only representations but also connectivity patterns. Preconfigured brain networks apparent within spontaneous activity at rest may function as a spatiotemporal “scaffolding” or low dimensional basis set of elements that can be selected or combined to form task-specific generative architectures.

There is much evidence for this idea [14,26]. First, the organization of the resting brain can be readily summarized along few dimensions. Early accounts emphasized a distinction between so-called task-positive and task-negative networks, respectively driven by sensory-motor (external) vs. self-referential, emotional, and memory retrieval (internal) tasks [31,95,96]. More recently, a gradient of cortical organization going from sensory-motor networks, with a faster temporal scale of integration, to polymodal tertiary association networks, with a slowest temporal scale, has been proposed [97–99]. This spatiotemporal organization is exactly the one mandated by hierarchical generative models. Networks that occupy higher levels may continuously generate predictions to suppress prediction errors of lower brain networks, such as primary sensory and motor regions – which may be engaged when prediction errors cannot be readily cancelled out and disengaged otherwise [97,100,101]. Furthermore, task-specific generative architectures can be formed by selectively engaging and disengaging network elements.

Second, there is direct experimental evidence that the connectivity patterns of high-level association networks at rest are set up in space and time to be ready to process behaviorally relevant and meaningful natural stimuli. For example, during a visuospatial attention task prefrontal and posterior parietal regions of the dorsal attention network interact more strongly, as compared to rest, with visual occipital regions. This top-down modulation is necessary for the selection of relevant visual stimuli. However, while the functional architecture of occipital visual regions is reorganized when transitioning from rest to task, dorsal attention regions’ topography and directional interactions do not change as if they were already at rest pre-set for attention [102]. Other studies have shown that the most central regions of the brain (hubs) are commonly recruited across many different tasks [103], and that many different task networks reflect the combination of low dimensional rest connectivity patterns [16]. These are examples of “spatial connectivity priors”.

A similar principle may occur in the temporal domain. At rest, hub regions in the dorsal attention, default, and motor network dynamically interact with peripheral sensory and motor regions through slow fluctuations of the alpha and beta band limited power (BLP) measured with MEG [104,105]. During the observation of movies, alpha and beta BLP connectivity in the visual system decreases, whereas the beta-band connectivity and dynamics of hub regions do not change. This stability may reflect a “temporal connectivity priors” [91]. Interestingly, the topography and dynamics of hub regions is modified as compared to rest if the visual input is not ecological as when the movie is temporally scrambled. Priors in spontaneous activity are therefore tuned to both the spatial and temporal statistics of the environment. Similar observations have been made in visual cortex measuring the similarity of spontaneous neuronal activity to natural vs. synthetic stimuli [38,39].

We shall emphasize that the connectivity architecture at rest is not the same as that observed during tasks, and that behavior requires the re-organization of resting networks. Spontaneous connectivity patterns at rest are thus “scaffolds” that change during tasks. The rules governing these modifications are incompletely understood. The distinction above between more stable central hub regions and more pliable peripheral sensory regions is one of them. Moreover, it is often assumed that resting networks correspond to specific behaviors. This idea comes from the similarity of task and rest networks, and the functional nicknames resting brain networks were labeled with early on (dorsal attention, visual, motor). However, that is not the case. For instance, a “reading” network composed of visual, auditory, language, and premotor regions has been well characterized in many task experiments [106], but the “visual word form area” in occipital cortex specialized in reading is functionally connected at rest with parietal and prefrontal regions of the dorsal attention network [14]. Similarly, brain regions that fluctuate at rest in two large clusters (task-positive vs. -negative) split in multiple clusters during vision of movies [107,108]. Hence spontaneous activity patterns may not be preconfigured for specific tasks but rather in a way that afford rapid transitions to many possible task-specific configurations [16,103,109] (Figure 2C), which is another manifestation of the maximum entropy principle.

### **Learning simpler and more accurate generative models without novel data**

Another key function of spontaneous activity during offline periods may be the optimization of generative models for future interactions, by removing their redundancies and aligning them with the spatiotemporal structure of the animal's ecological niche [2,110].

Generative models can be optimized even without novel data, in at least two ways. The first kind of optimization consists in reducing model complexity by pruning unnecessary parameters, i.e., those that can be removed without losing accuracy [111,112]. This process is analogous to post-hoc (Bayesian) model selection used in data analysis, which permits evaluating the evidence for a full model against reduced versions of the model – and select the most parsimonious. Spontaneous fluctuations may explore the space of possible (reduced) models, thus producing a form of mind wandering [113]. Importantly, since model parameters are encoded in neural tissue (e.g., synapses), reducing model complexity entails metabolic parsimony. This biological consideration aligns our proposal with theories of synaptic pruning and homeostasis during sleep [114].

The second kind of optimization consists in generating (sampling) fictive data from the model probability distribution, and then using these fictive data, as they were real data, to optimize the same or other models [73,115]. This method was used in an early algorithm to train unsupervised generative models: wake-sleep. While during the “wake” phase the algorithm used real data for training, during the “sleep” phase it used fictive data generated by the model itself [116]. Recent advancements in “generative replay” show that sampling fictive data from a generative model and then using them to train the same or other models is an effective strategy to learn multiple tasks, while avoiding catastrophic forgetting (i.e., the tendency of neural networks to forget old tasks when they learn novel ones) [73].

Neurophysiological support from this idea comes from studies of “replay” from memory. Replays are observed in several cortical and subcortical areas, such as prefrontal cortex [117] and visual cortex [118], but are studied more extensively in the hippocampus, where they correspond to high-frequency reactivations of previously experienced patterns, e.g., sequences of place cells along spatial trajectories [48]. To explain these findings, an influential framework posits that the hippocampus is a memory system that stores experiences rapidly and then replays them during off-line periods, to train a separate (cortical) memory system that learns more slowly but generalizes better [119,120]. This theory inspired the widespread use of “experience replays” from a memory buffer to improve learning in deep networks [121]. The notion of “generative replay” suggests instead that hippocampal replays are generated by resampling from a generative model, not rehearsed from verbatim memories [73]. This perspective explains more naturally the fact that hippocampal replays can recombine experiences and reorganize them to follow novel learned rules [47,122] – possibly, by cross-talking with prefrontal cortex and other areas [44,54]. The possibility to recombine experiences may be crucial for prospective functions commonly associated to the hippocampus, such as planning and imagination [48,123].

Hippocampal replays occur at very high frequencies [42–45], which seems at odds with our emphasis on low frequency dynamics in the previous section. However, the learning mechanisms discussed in this section require specific episodic sequences (e.g., place cell sequences corresponding to specific trajectories, not average trajectories). Encoding such episodic details may require high-frequency activations observed in hippocampal replays, in contrast to the low frequencies encoding generic priors. Furthermore, both high and low frequencies are elicited during coordinated cortical-subcortical reactivations that occur both during sleep and the awake state, potentially favoring the bidirectional integration of more generic (cortical) and specific (hippocampal) representations [54,55,124,125]. An fMRI study suggests that cortical-hippocampal dialogue could follow a “sender-receiver” architecture, with high frequency (delta

band) activity conveying information from sender to receiver; and low frequency (infra-slow) activity propagating from receiver to sender, to coordinate the timing of information transfer [126]. Interestingly, the study shows that during wakefulness, the cortex may act as sender and the hippocampus as receiver; but during slow-wave sleep, propagation directions reverse – suggesting that the cortex and the hippocampus may exchange (sender-receiver) roles when the brain’s generative model is task-engaged or optimized, respectively.

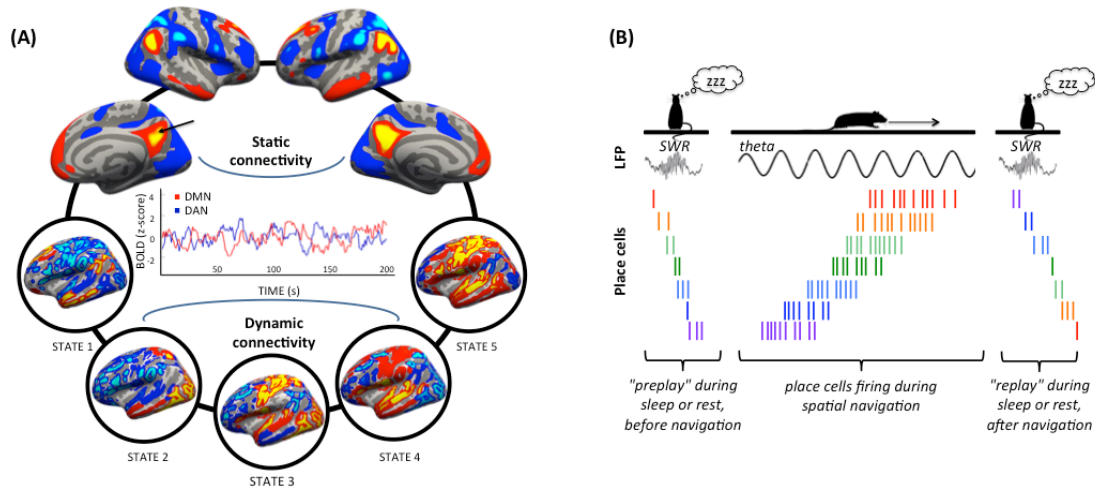
## **Concluding remarks**

As our vivid dreams exemplify, spontaneous activity in the resting brain can be cognitively sophisticated. We proposed that spontaneous activity in cortical and subcortical brain networks could be understood in terms of top-down computational processes that generative models use to learn and infer states of the world. During offline periods, these processes are most useful to refine the model, as opposed to explain data accurately. Spontaneous activity is therefore characterized by low-frequency transitions that prepare generic priors, explore the brain’s dynamic repertoire, and reduce the complexity of brain generative models; and are sometimes punctuated by high-frequency transitions that selectively update the model with fictive data. This perspective reconciles two varieties of spontaneous activity (resting state activity and replays) that are often studied separately, suggesting that they may operate in a coordinated manner to optimize brain generative models.

This view contributes to an emerging framework that puts spontaneous activity at the center stage of brain processing [2,39,127,128]. It posits that model learning starts from preconfigured circuit dynamics that produce spontaneous activity. This activity is initially meaningless and acquires meaning through the constant alignment to statistics and dynamics of the external world, realized by acting and perceiving action consequences. During task-evoked activity, we observe the modulation of spontaneous activity by external inputs. Rather, during spontaneous activity at rest, we see manifestations of the internal model that – our analysis suggests – prepares us for future interactions (see the Outstanding Questions).

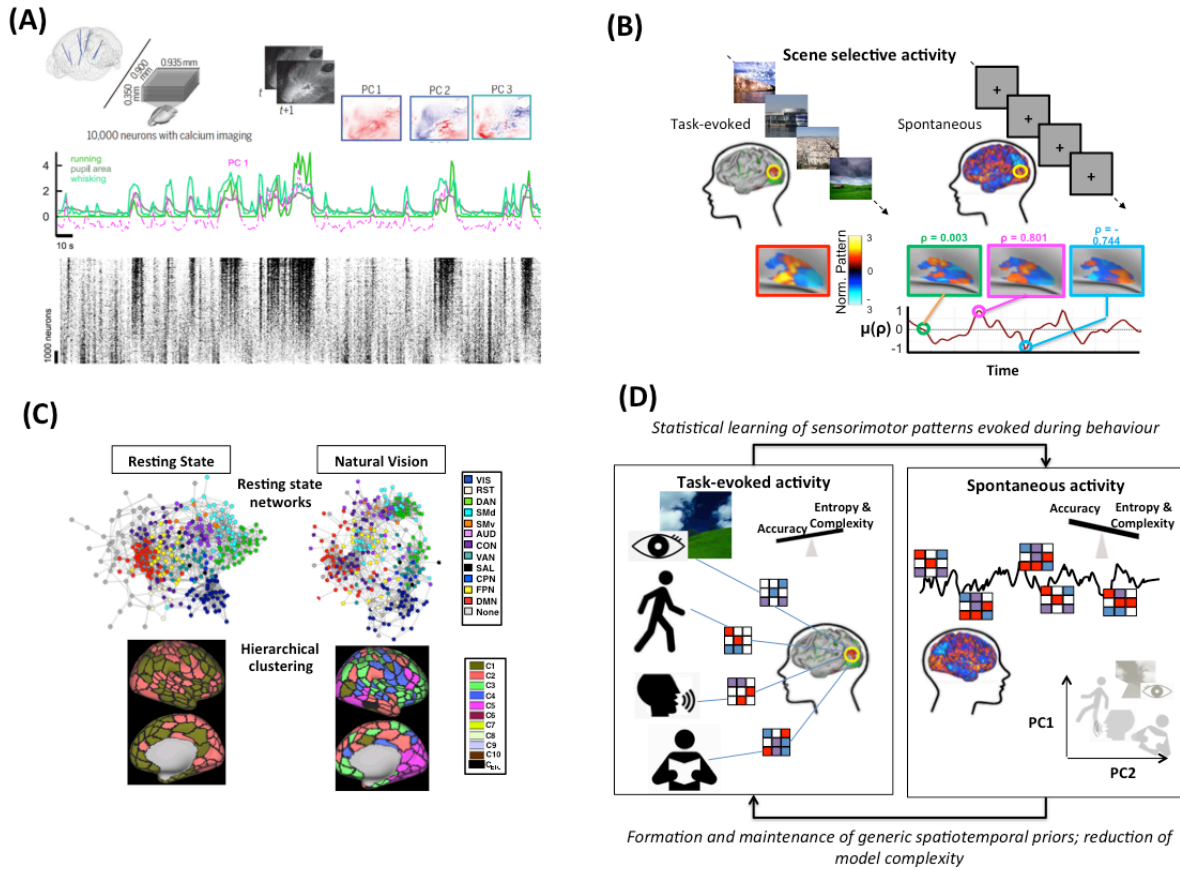
In sum, this perspective implies that the main function of spontaneous brain activity is the maintenance and optimization of the brain’s generative models for future interactions. While not all our dreams become reality, we always dream for the future.

## Figures

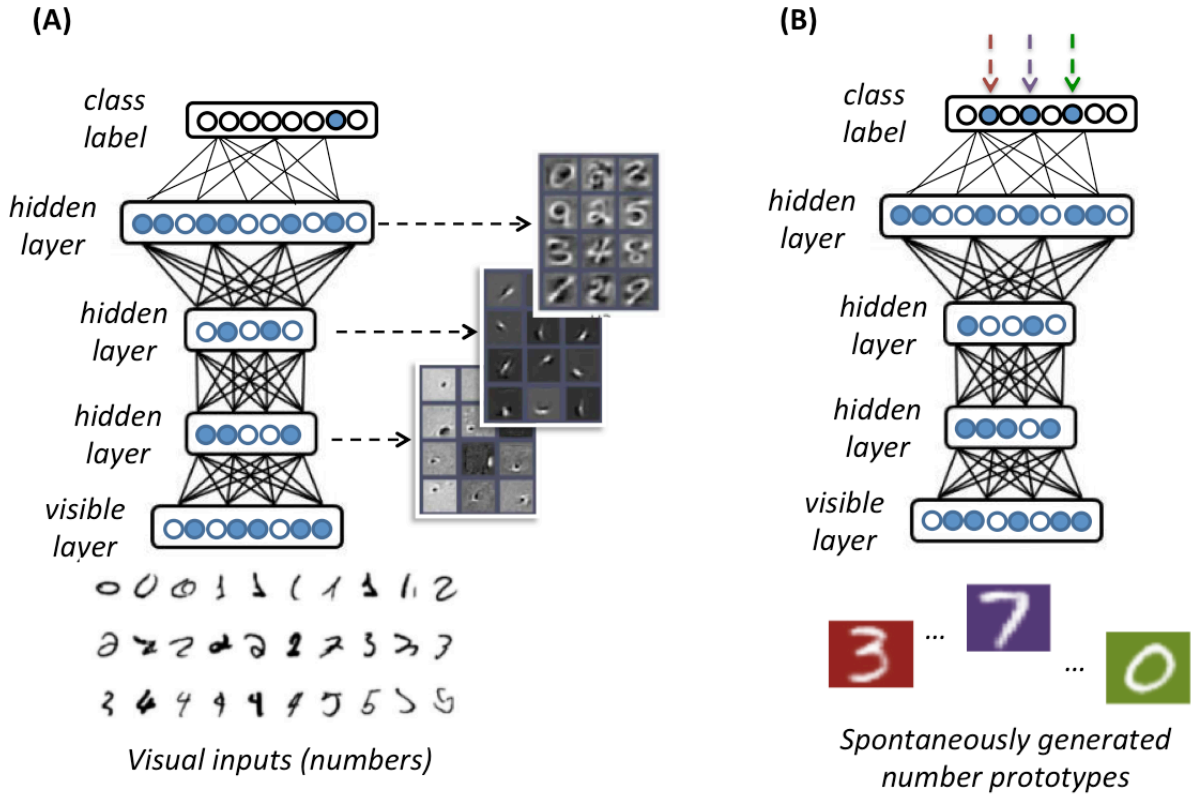


**Figure 1. Two examples of spontaneous brain dynamics. (A)** An example of resting state spontaneous activity measured with fMRI. Static connectivity: a map of the ‘static’ (time-averaged) temporal correlation of the blood oxygenation level dependent (BOLD) signal between a region in medial parietal cortex (black arrowhead), and the rest of the cortex over many minutes. Inset: fluctuations of the BOLD signal in two cortical networks: DMN-default mode network, and DAN-dorsal attention network. Dynamic connectivity: the maps represent the main patterns of correlated activity across the brain as identified through a sliding window analysis, a winner-take-all classification, and the projection on the cortical surface of the first eigenvector. The fluctuations (time-course) have a frequency of about 1/10 seconds (0.1 Hz). Low frequency activity is also evident in single cell recordings from rat neocortex, where they appear to encode behaviorally relevant information (**Figure 2A**). **(B)** A schematic illustration of internally generated hippocampal sequences. The middle part of the figure shows a (fictive) spatiotemporal sequence of spikes from seven hippocampal place cells (represented by different colors), whose place fields are located in different portions of the corridor. These sequences are visible within the hippocampal theta rhythm, while the animal navigates through the corridor. However, sequential activity from the same ensemble – sometimes in the same (or reverse) order as during navigation – can be decoded during animal sleep or awake rest before (“preplays”) and after (“replays”) navigation, respectively. These internally generated hippocampal sequences are often embedded in network events called sharp-wave ripple (SWR) complexes [42–45].





**Figure 2. Supporting evidence and schematic of the novel proposal.** (A) Similarity of spontaneous activity in visual cortex and low dimensional behavioral components of mouse movements. Calcium imaging recordings of 10,000 neurons from mouse visual cortex during natural exploration (from courtesy of Matteo Carandini). Facial movements are summarized by a few patterns of correlated movements of whiskers, facial expressions, eye movements, and pupil area. Bottom shows raster plots of activity in visual cortex without visual stimulation (spontaneous). Note slow correlated fluctuations ( $\sim 1/10$  seconds). The intermediate panel shows how the first principal component of activity (violet timeseries) correlates with behavioral components (green timeseries). (B) Spontaneously emerging patterns in human visual cortex and their functional connectivity are linked to the patterns evoked by visual stimuli. Visual objects (faces, scenes, bodies, words) produce multivariate patterns of activity recorded with fMRI in visual association cortex. The left inset shows the multivariate pattern in a scene-processing region of human cortex. At rest, scene specific patterns, i.e. yielding strong spatial correlation with the task pattern—see time-course of correlation values—occur more frequently than patterns representing other objects or null patterns. The same occur in other regions of visual cortex, e.g. more face patterns at rest in face specific regions and so on. Critically, object specific patterns emerge at rest in a synchronized manner across multiple regions [77]. (C) Resting networks as connectional priors of task networks, from [107]. Reorganization of cortical regions and networks when observers go from visual fixation (rest) to movie watching. Top: Spring embedded representation of the temporal correlation strength of the BOLD fMRI signal between individual regions of RSNs according to the parcellation of [129]. Note that networks that are segregated at rest combine during tasks. Bottom: data driven hierarchical clustering. Two large resting communities (task-positive, task-negative) split in multiple communities during movie watching. (D) Schematic illustration of the novel hypothesis. Brain generative models continuously incorporate the statistical history of brain co-activation patterns experienced during behavioral experience. The endogenous regeneration of brain co-activation patterns at rest supports the offline optimization (e.g., compression) of the brain’s generative model, and the preparation of “generic spatiotemporal priors” for future tasks. These comprise both representations and connectivity patterns; and become apparent as transitions between low-dimensional states of the network. These are shown in the right part of the figure a series of brain co-activation patterns (which possibly average across multiple episodes) nested within low-frequency fluctuations of the resting brain. The inset suggests that a large variety of task representations can be summarized in low dimensional states.



**Figure 3. Hierarchical generative neural networks. (A)** Deep Boltzmann Machines (DBMs) are deep neural networks of symmetrically coupled stochastic neurons organized in a hierarchy of multiple layers. The DBM learns a hierarchical generative model of the input (e.g., sensory) data presented on the “visible layer”. “Hidden layers” contain neurons that encode latent causes of the data: when trained on images (here, handwritten digits) they become tuned to visual features that are increasingly more complex in deeper layers (examples of receptive fields of individual neurons are shown in the right panels [130]). Learning is unsupervised (i.e., it does not require external teaching or reward signal): its objective is to learn a probability distribution that approximates the true probability distribution of the training data. Recurrent connections convey the information sampled from the upper layers downstream to generate data on the visible layer, in a top-down fashion. The divergence between real input and its top-down reconstruction drives the change of connection weights during learning, using Hebbian rules. After learning, recurrent interactions support stochastic inference that leads to denoising, completion or “filling in” of ambiguous (or missing) inputs, in the same sensory modality or in different modalities if the architecture is multimodal (e.g., learns visual and linguistic inputs). Discriminative tasks (here, digit classification) can be learned by adding a layer of neurons representing the class labels. During learning, the model acquires “generic priors”: here, prototypical digit shapes that abstract away from many input details [57,130]. **(B)** After learning, sampling can be conditioned by the class labels to generate prototypical digit shapes “spontaneously”, i.e., in the absence of sensor inputs.

## Boxes

### Box 1. Examples of biological implementations of generative modeling

A popular biological implementation of perceptual inference is predictive coding [60,131]. Predictive coding uses a hierarchical architecture, where different perceptual hypotheses at higher levels (e.g., about which objects exist in a visual scene) compete to explain bottom-up stimuli. The competition requires reciprocal interactions between top-down and bottom-up streams. Top-down predictions from hierarchically higher levels (i.e., what object one expects to see) are continuously compared with incoming stimuli. Prediction errors emerging from this comparison are propagated bottom-up, to revise perceptual hypotheses (e.g., change mind about what object one is seeing); and the process continues until prediction error is minimal – and the best interpretation of the perceptual scene emerges.

In predictive coding, top-down dynamics are key to maintain the representation of a task, predict sensory inputs at lower hierarchical levels and replace them if missing [132]. Furthermore, they play a modulatory role, by conveying precision signals through neuromodulation and NMDAR-dependent plasticity. Precision signals are second-order predictions about the precision (inverse variance) of prediction errors. They render the inference flexible by modulating the gain of prediction error units in proportion to their reliability, permitting (for example) to ignore noisy stimuli [133,134].

Active Inference extends the above scheme to the inference of the sequence of actions or policy expected to minimize the discrepancy between current and desired (goal) states [111]. It requires endowing the generative model with a (transition) mapping between actions and their expected outcomes; and a prior preference over outcomes, which encodes the agent's goals (e.g., homeostatic goals, such as having optimal glucose levels in the blood). This prior preference is analogous to a set-point in cybernetics: it acts as a constant source of (high-precision) top-down predictions, which the agent fulfills by acting (e.g., eating a cake) [61,135].

Top-down dynamics also support unsupervised learning, and the update of the parameters of generative models. One example is the Restricted Boltzmann Machine (RBM): a stochastic recurrent neural network that learns to reconstruct its sensory input by gradually adjusting the connection weights, so that the top-down signals better match the inputs. Learning consists in maximizing the likelihood of the observed data under the generative model, which is equivalent to discovering efficient ways of coding the sensory data. The RBM is also the building block of Deep Belief Networks (DBNs) and Deep Boltzmann Machines (DBMs) [57,67], a class of deep networks that learn a hierarchical generative model with latent features that are increasingly more complex and abstract as a function of network depth (Figure 3). DBNs/DBMs are appealing because they entail a more psychologically plausible learning regimen [64,130] compared to classic deep feed-forward networks that rely on discriminative learning and therefore require an external teaching signal (i.e., labeled data) at each learning event.

### Box 2. Brain generative models beyond the natural statistics of external stimuli: behaviorally relevant and interoceptive states

Mounting evidence suggests that the patterns of activity of the brain are guided by control demands rather than just the imperative to represent the external world, as in classical interpretations of predictive coding [136–139]. Similarly, spontaneous activity can represent the ensemble of brain states driven by behavior [23,75] rather than just the natural statistics of external stimuli [37–39,41,75]. Behaviorally-relevant brain patterns may result from the progressive internalization of action-perception loops, which can be later spontaneously reactivated to support future-oriented cognition [127].

If behaviorally relevant states are internalized in the brain's generative model, they should undergo compression. This would explain why many task patterns can be summarized with low dimensional vectors of neural activities [16] and low dimensional dynamics for the control of movements [140]. Such compression may be effective because also behavior, despite its apparent heterogeneity and individual variability, is low dimensional. Theoretically, thousands of linear combinations of degrees of freedom of joints and muscles during arm and hand movements can be summarized by a dozen or so of correlated patterns of joint movements and muscle activation (motor synergies) [141]. These synergies are represented in motor cortex [142] and at rest [88]. The pattern of eye movement exploration during laboratory and ecological exploration tasks can be described with a small number of correlated features that explain most of the inter-subject variability [143]. Notably, components identified during visual exploration also describe spontaneous eye movements to a blank screen suggesting the role of intrinsic dynamics in shaping active behavior [144]. Two styles across hundreds of participants (returners, explorers) describe human navigation in natural surroundings [145]. The ability to

inhibit immediate reward is also described by three archetypes across more than a thousand subjects [146]. Finally, behavioral deficits after focal lesions are correlated across patients identifying two axes—action/attention vs. language/memory—[147] that nicely match key dimensions of inter-hemispheric activation [148]. In summary, low dimensional neural patterns discussed in the main article nicely match low dimensional behavioral organization.

Furthermore, the brain's generative model may support interoceptive inference, based on interoceptive signals from the internal of the body [149,150]. Interoceptive inference may be guided by strong, primordial priors (e.g., about viable ranges for body temperature and sugar level in the blood) that ensure allostasis. Interoceptive prediction errors (signaling that e.g., the body is too hot) may steer a cascade of error-correction processes, ranging from autonomic reflexes such as ventilation to goal-directed plans such as buying a cold drink. These activities may manifest themselves also during spontaneous activity (also given that the brain remains relatively more connected to interoceptive than exteroceptive signals) and possibly contribute to consolidate interoceptive models and priors.

Finally, an intriguing possibility is that spontaneous activity connectivity patterns may reflect not only a generative neural model, but also a metabolic prior to anticipate future activity and energetic demands in the same circuitry. In keeping, a recent study reveals a deep connection between neural activity and metabolism, by showing that neural activity drives metabolic signals like intracellular pyruvate and ATP; and that transient sensory stimulation causes an allocation of energy to recently activated connections [24].

### Outstanding Questions Box

- Do low- and high-frequency spontaneous activity serve separable functional roles? Do they reflect a difference between coarse-grained, prototype-like patterns that average across many stimuli, responses and events vs. detailed, episodic-like patterns encoding single events or sequences? Do they reflect key differences in spontaneous activity between cortical and subcortical systems?
- What is the relation between neural priors and metabolism? A recent study shows that metabolism prospectively increases in recently active circuits, raising the possibility that the brain encodes “metabolic priors” that help allocating energy in anticipation of future needs. Moreover, theoretical calculations indicate that in the cerebellum, most energy is spent in maintaining potential sensory-motor associations on-line rather than controlling current movements – does the same happen in cortex?
- Do development and learning shape not only novel functional connections, but also multivariate prior representations? What is the ontogeny of spontaneous activity patterns? Brain generative models may reflect more than the natural statistics of exteroceptive signals; for example, low dimensional memory and interoceptive patterns. Does spontaneous activity reflect a weighted contribution of these different signals?
- What are the implications of our hypothesis for education, learning theory and addiction? In education and drug rehabilitation the notion of task control is often emphasized, e.g., the idea to use ‘slow’ thinking over ‘fast’ reflexive behavior. However, if it is true that spontaneous activity patterns entrained by the frequency of a certain behavior shape future behavior, then the challenge is not to train control, but to train good adaptive habits.
- Is it possible that some disorders such as depression, obsessive-compulsive disorder or schizophrenia reflect excessively strong priors that render patients insensitive to low-level stimuli or abnormally sensitive to prediction error signals? This idea might explain why brain networks are dysfunctional across a variety of disorders, even in the presence of relatively local damage. A paucity or simplification of priors may also be an explanation for loss of memory or semantic information in neurodegeneration, trauma, and stroke.

## Glossary

- **Bayes' rule.** Bayes' rule provides us with a way to update a prior belief (the probability of an event before new data is collected) into a posterior belief, after collecting new, relevant data.
- **Boltzmann machine.** A popular biological implementation of unsupervised learning of generative models. It is a stochastic neural network of symmetrically connected neurons, whose dynamics is governed by an energy function. One layer or (visible) neurons encodes the input data pattern whereas another layer of (hidden) neurons encodes the latent causes of the data, see Figure 3.
- **Resting state networks.** Networks of brain regions that activate or deactivate together during spontaneous brain activity at rest and are jointly active during behavioral tasks. However the equivalence rest-task is not complete as most tasks involve the recombination of resting state networks.
- **Predictive coding.** A popular biological implementation of perceptual inference using a hierarchical generative model. It explains how perceptual inference can be cast as the minimization of prediction errors, which stem from the continuous comparison of top-down sensory predictions and stimuli; see Box 1.
- **Generative vs. discriminative models:** A generative model is a probabilistic model that describes how a set of data is generated. By sampling from this model, we can generate new data instances. In contrast, discriminative models discriminate between different kinds of data instances (e.g., in a classification task).
- **Task-evoked versus spontaneous brain activity:** This dichotomy usually refers to neuronal activations recorded when a creature is engaged in some task, in the presence of external stimuli (task-evoked) versus at rest, when external stimuli are weak or absent (spontaneous activity). However, this dichotomy may be slightly misleading, if both spontaneous and task-evoked activities are manifestations of generative models. Spontaneous activity at rest may reflect the activity of generative models (and especially their top-down dynamics) in the absence of external stimuli to "explain away". Generative models can also produce spontaneous dynamics immediately before and during task-engagement: these create the context for stimulus processing. Accordingly, the evidence we review shows that stimuli only explain a portion of task-evoked brain activity, whereas a larger portion can be explained by spontaneous activity immediately before stimuli onset.

## Acknowledgements

GP was funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement Nos. 785907 and 945539 (Human Brain Project SGA2 and SGA3) and the European Research Council under the Grant Agreement No. 820213 (ThinkAhead). MZ was funded by the Italian Ministry of Health (RF-2013-02359306 and Ricerca Corrente IRCCS Ospedale San Camillo), CARIPARO Foundation Excellence Grant, and MIUR (Dipartimenti di Eccellenza DM 11/05/2017 n. 262 to the Department of General Psychology). MC was funded by FC-Neuro Progetto Strategico University of Padova, NIH NS095741, FLAG-ERA JTC 2017-2020, BIAL Foundation grant (No. 361/18), Dept. Excellence Italian Ministry of Research (MIUR), CARIPARO foundation Padova grant 2019, and Celeghin Foundation grant 2019.

## References

- 1 Raichle, M.E. and Mintun, M.A. (2006) Brain Work and Brain Imaging. *Annual Review of Neuroscience* 29, 449–476
- 2 Buzsaki, G. (2019) *The brain from inside out*, Oxford University Press, USA.
- 3 Raichle, M.E. (2011) The Restless Brain. *Brain Connect* 1, 3–12
- 4 Buckner, R.L. *et al.* (2013) Opportunities and limitations of intrinsic functional connectivity MRI. *Nat Neurosci* 16, 832–837
- 5 Barlow, H.B. (1961) Possible principles underlying the transformation of sensory messages. *Sensory communication*
- 6 Yamins, D.L. and DiCarlo, J.J. (2016) Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience* 19, 356
- 7 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18, 193–222
- 8 Corbetta, M. and Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 3, 201–215
- 9 Tollhurst, D.J. *et al.* (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research* 23, 775–785
- 10 Shadlen, M.N. and Newsome, W.T. (1998) The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. *J. Neurosci.* 18, 3870–3896
- 11 Deco, G. *et al.* (2013) Resting-State Functional Connectivity Emerges from Structurally and Dynamically Shaped Slow Linear Fluctuations. *J. Neurosci.* 33, 11239–11252
- 12 Barttfeld, P. *et al.* (2015) Signature of consciousness in the dynamics of resting-state brain activity. *PNAS* 112, 887–892
- 13 Lewis, C.M. *et al.* (2009) Learning sculpts the spontaneous activity of the resting human brain. *PNAS* 106, 17558–17563
- 14 Petersen, S.E. and Sporns, O. (2015) Brain Networks and Cognitive Architectures. *Neuron* 88, 207–219
- 15 Harmelech, T. and Malach, R. (2013) Neurocognitive biases and the patterns of spontaneous correlations in the human cortex. *Trends in Cognitive Sciences* 17, 606–615
- 16 Shine, J.M. *et al.* (2019) Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nature neuroscience* 22, 289–296
- 17 Fiser, J. *et al.* (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn Sci* 14, 119–130
- 18 Chanes, L. and Barrett, L.F. (2016) Redefining the Role of Limbic Areas in Cortical Processing. *Trends Cogn. Sci. (Regul. Ed.)* 20, 96–106
- 19 Howarth, C. *et al.* (2012) Updated Energy Budgets for Neural Computation in the Neocortex and Cerebellum. *J Cereb Blood Flow Metab* 32, 1222–1232
- 20 Engl, E. and Attwell, D. (2015) Non-signalling energy use in the brain. *The Journal of Physiology* 593, 3417–3429
- 21 Harris, J.J. *et al.* (2012) Synaptic energy use and supply. *Neuron* 75, 762–777
- 22 Mitra, A. *et al.* (2018) Spontaneous Infra-slow Brain Activity Has Unique Spatiotemporal Dynamics and Laminar Structure. *Neuron* 98, 297–305.e6
- 23 Stringer, C. *et al.* (2019) Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364,
- 24 Mann, K. *et al.* (2021) Coupling of activity, metabolism and behaviour across the Drosophila brain. *Nature* DOI: 10.1038/s41586-021-03497-0
- 25 Palva, J.M. and Palva, S. (2012) Infra-slow fluctuations in electrophysiological recordings, blood-oxygenation-level-dependent signals, and psychophysical time series. *NeuroImage* 62, 2201–2211
- 26 Smith, S.M. *et al.* (2009) Correspondence of the brain’s functional architecture during activation and rest. *PNAS* 106, 13040–13045
- 27 Mantini, D. *et al.* (2007) Electrophysiological signatures of resting state networks in the human brain. *PNAS* 104, 13170–13175
- 28 Pasquale, F. de *et al.* (2010) Temporal dynamics of spontaneous MEG activity in brain networks. *PNAS* 107, 6040–6045
- 29 Leopold, D.A. *et al.* (2003) Very Slow Activity Fluctuations in Monkey Visual Cortex: Implications for Functional Brain Imaging. *Cereb Cortex* 13, 422–433
- 30 Betti, V. *et al.* (2018) Topology of Functional Connectivity and Hub Dynamics in the Beta Band As Temporal Prior for Natural Vision in the Human Brain. *J Neurosci* 38, 3858–3871

- 31 Smith, S.M. *et al.* (2015) A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience* 18, 1565–1567
- 32 Finn, E.S. *et al.* (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience* 18, 1664–1671
- 33 Albert, N.B. *et al.* (2009) The resting human brain and motor learning. *Curr. Biol.* 19, 1023–1027
- 34 Baldassarre, A. *et al.* (2012) Individual variability in functional connectivity predicts performance of a perceptual task. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3516–3521
- 35 Baldassarre, A. *et al.* (2016) Dissociated functional connectivity profiles for motor and attention deficits in acute right-hemisphere stroke. *Brain* 139, 2024–2038
- 36 Kenet, T. *et al.* (2003) Spontaneously emerging cortical representations of visual attributes. *Nature* 425, 954–956
- 37 Luczak, A. *et al.* (2009) Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. *Neuron* 62, 413–425
- 38 Berkes, P. *et al.* (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87
- 39 Fiser, J. *et al.* (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* 431, 573–578
- 40 Han, F. *et al.* (2008) Reverberation of recent visual experience in spontaneous cortical waves. *Neuron* 60, 321–327
- 41 Arieli, A. *et al.* (1996) Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871
- 42 Buzsáki, G. (2015) Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* 25, 1073–1188
- 43 Diba, K. and Buzsáki, G. (2007) Forward and reverse hippocampal place-cell sequences during ripples. *Nat Neurosci* 10, 1241–1242
- 44 Pezzulo, G. *et al.* (2017) Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Ann. N.Y. Acad. Sci.* 1396, 144–165
- 45 Pfeiffer, B.E. and Foster, D.J. (2013) Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* 497, 74–79
- 46 Kurth-Nelson, Z. *et al.* (2016) Fast Sequences of Non-spatial State Representations in Humans. *Neuron* 91, 194–204
- 47 Liu, Y. *et al.* (2019) Human Replay Spontaneously Reorganizes Experience. *Cell* 178, 640–652.e14
- 48 Foster, D.J. (2017) Replay Comes of Age. *Annual Review of Neuroscience* 40, 581–602
- 49 Dragoi, G. and Tonegawa, S. (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469, 397–401
- 50 Jadhav, S.P. *et al.* (2012) Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. *Science* 336, 1454–1458
- 51 Liu, X. *et al.* (2012) Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* 484, 381–385
- 52 Ramirez, S. *et al.* (2013) Creating a False Memory in the Hippocampus. *Science* 341, 387–391
- 53 Ji, D. and Wilson, M.A. (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat Neurosci* 10, 100–107
- 54 Penagos, H. *et al.* (2017) Oscillations, neural computations and learning during wake and sleep. *Curr. Opin. Neurobiol.* 44, 193–201
- 55 Shin, J.D. and Jadhav, S.P. (2016) Multiple modes of hippocampal-prefrontal interactions in memory-guided behavior. *Curr. Opin. Neurobiol.* 40, 161–169
- 56 Sirota, A. *et al.* (2003) Communication between neocortex and hippocampus during sleep in rodents. *PNAS* 100, 2065–2069
- 57 Hinton, G.E. (2007) Learning multiple layers of representation. *Trends Cogn Sci* 11, 428–434
- 58 Gilbert, C.D. and Sigman, M. (2007) Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696
- 59 Ringach, D.L. (2009) Spontaneous and driven cortical activity: implications for computation. *Curr Opin Neurobiol* 19, 439–444
- 60 Friston, K. (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360, 815–836
- 61 Pezzulo, G. *et al.* (2018) Hierarchical active inference: a theory of motivated control. *Trends in cognitive sciences* 22, 294–306

- 62 Fries, P. (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences* 9, 474–480
- 63 Jensen, O. *et al.* (2015) Oscillatory mechanisms of feedforward and feedback visual processing. *Trends Neurosci* 38, 192–194
- 64 Testolin, A. and Zorzi, M. (2016) Probabilistic Models and Generative Neural Networks: Towards an Unified Framework for Modeling Normal and Impaired Neurocognitive Functions. *Front Comput Neurosci* 10,
- 65 Friston, K. *et al.* (2015) Active inference and epistemic value. *Cogn Neurosci* 6, 187–214
- 66 Helmholtz, H. von (1866) Concerning the perceptions in general. In *Treatise on physiological optics* 3 (Southall, J. P. C., ed), Dover
- 67 Salakhutdinov, R. (2015) Learning Deep Generative Models. *Annual Review of Statistics and Its Application* 2, 361–385
- 68 Barrett, L.F. (2017) *How emotions are made: The secret life of the brain*, Houghton Mifflin Harcourt.
- 69 Ress, D. *et al.* (2000) Activity in primary visual cortex predicts performance in a visual detection task. *Nature Neuroscience* 3, 940–945
- 70 Carrasco, M. (2011) Visual attention: The past 25 years. *Vision Research* 51, 1484–1525
- 71 Jaynes, E.T. (1957) Information theory and statistical mechanics. II. *Physical review* 108, 171
- 72 Buesing, L. *et al.* (2011) Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons. *PLoS Computational Biology* 7, e1002211
- 73 Stoianov, I. *et al.* (2020) The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning. *bioRxiv* DOI: 10.1101/2020.01.16.908889
- 74 Stoianov, I. and Zorzi, M. (2012) Emergence of a “visual number sense” in hierarchical generative models. *Nat Neurosci* 15, 194–196
- 75 Romano, S.A. *et al.* (2015) Spontaneous Neuronal Network Dynamics Reveal Circuit’s Functional Adaptations for Behavior. *Neuron* 85, 1070–1085
- 76 Omer, D.B. *et al.* (2019) Dynamic Patterns of Spontaneous Ongoing Activity in the Visual Cortex of Anesthetized and Awake Monkeys are Different. *Cereb Cortex* 29, 1291–1304
- 77 Kim, D. *et al.* (2020) Spontaneously emerging patterns in human visual cortex and their functional connectivity are linked to the patterns evoked by visual stimuli. *Journal of Neurophysiology* 124, 1343–1363
- 78 Kamps, F.S. *et al.* (2020) Connectivity at the origins of domain specificity in the cortical face and place networks. *Proc Natl Acad Sci U S A* 117, 6163–6169
- 79 Kaiser, D. *et al.* (2019) Object Vision in a Structured World. *Trends in Cognitive Sciences* 23, 672–685
- 80 Testolin, A. *et al.* (2017) Letter perception emerges from unsupervised deep learning and recycling of natural image features. *Nature Human Behaviour* 1, 657
- 81 Quiroga, R.Q. *et al.* (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107
- 82 Chang, L. and Tsao, D.Y. (2017) The Code for Facial Identity in the Primate Brain. *Cell* 169, 1013–1028.e14
- 83 Churchland, M.M. *et al.* (2012) Neural population dynamics during reaching. *Nature* advance online publication,
- 84 Kriegeskorte, N. *et al.* (2008) Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron* 60, 1126–1141
- 85 Leo, A. *et al.* (2016) A synergy-based hand control is encoded in human motor cortical areas. *eLife* 5, e13420
- 86 Huth, A.G. *et al.* (2012) A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron* 76, 1210–1224
- 87 Lescroart, M.D. and Gallant, J.L. (2019) Human Scene-Selective Areas Represent 3D Configurations of Surfaces. *Neuron* 101, 178–192.e7
- 88 Livne, T. *et al.* (2020) Spontaneous emergence of behaviorally relevant motifs in human motor cortex. *bioRxiv* DOI: 10.1101/2020.10.25.353326
- 89 Eichenlaub, J.-B. *et al.* (2020) Replay of Learned Neural Firing Sequences during Rest in Human Motor Cortex. *Cell Reports* 31, 107581
- 90 Astafiev, S.V. *et al.* (2004) Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nat Neurosci* 7, 542–548
- 91 Betti, V. *et al.* (2020) Spontaneous Beta Band Rhythms in the Predictive Coding of Natural Stimuli. *Neuroscientist* DOI: 10.1177/1073858420928988
- 92 Corbetta, M. *et al.* (2018) On the low dimensionality of behavioral deficits and alterations of brain network connectivity after focal injury. *Cortex* 107, 229–237
- 93 Friston, K.J. *et al.* (2014) On nodes and modes in resting state fMRI. *Neuroimage* 99, 533–547
- 94 Ghosh, A. *et al.* (2008) Noise during Rest Enables the Exploration of the Brain’s Dynamic Repertoire. *PLoS*



- 95 Buckner, R.L. *et al.* (2008) The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38
- 96 Fox, M.D. *et al.* (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *PNAS* 102, 9673–9678
- 97 Margulies, D.S. *et al.* (2016) Situating the default-mode network along a principal gradient of macroscale cortical organization. *PNAS* 113, 12574–12579
- 98 Raut, R.V. *et al.* (2020) Hierarchical dynamics as a macroscopic organizing principle of the human brain. *PNAS* 117, 20890–20897
- 99 Mitra, A. *et al.* (2016) Human cortical–hippocampal dialogue in wake and slow-wave sleep. *PNAS* 113, E6868–E6876
- 100 Carhart-Harris, R.L. and Friston, K.J. (2010) The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain* 133, 1265–1283
- 101 Zhang, J. *et al.* (2019) Intrinsic Functional Connectivity is Organized as Three Interdependent Gradients. *Sci Rep* 9, 15976
- 102 Spadone, S. *et al.* (2015) Dynamic reorganization of human resting-state networks during visuospatial attention. *PNAS* 112, 8112–8117
- 103 Cole, M.W. *et al.* (2014) Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83, 238–251
- 104 Betti, V. *et al.* (2013) Natural Scenes Viewing Alters the Dynamics of Functional Connectivity in the Human Brain. *Neuron* 79, 782–797
- 105 de Pasquale, F. *et al.* (2012) A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron* 74, 753–764
- 106 Price, C.J. (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847
- 107 Kim, D. *et al.* (2018) A New Modular Brain Organization of the BOLD Signal during Natural Vision. *Cereb Cortex* 28, 3065–3081
- 108 Meer, J.N. van der *et al.* (2020) Movie viewing elicits rich and reliable brain state dynamics. *Nature Communications* 11, 5004
- 109 Krienen, F.M. *et al.* (2014) Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philos Trans R Soc Lond B Biol Sci* 369,
- 110 Northoff, G. (2018) *The Spontaneous Brain: From the Mind–Body to the World–Brain Problem*, MIT Press.
- 111 Friston, K. *et al.* (2017) Active Inference, Curiosity and Insight. *Neural Comput* DOI: 10.1162/neco\_a\_00999
- 112 Hobson, J.A. and Friston, K.J. (2012) Waking and dreaming consciousness: neurobiological and functional considerations. *Progress in neurobiology* 98, 82–98
- 113 Sadaghiani, S. *et al.* (2010) The relation of ongoing brain activity, evoked neural responses, and cognition. *Front. Syst. Neurosci.* 4,
- 114 Cirelli, C. and Tononi, G. (2017) The Sleeping Brain. *Cerebrum* 2017,
- 115 Barron, H.C. *et al.* (2020) Prediction and memory: a predictive coding account. *Progress in Neurobiology* DOI: 10.1016/j.pneurobio.2020.101821
- 116 Hinton, G.E. *et al.* (1995) The “wake-sleep” algorithm for unsupervised neural networks. *Science* 268, 1158–1161
- 117 Peyrache, A. *et al.* (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature neuroscience* 12, 919–926
- 118 Wittkuhn, L. and Schuck, N.W. (2021) Dynamics of fMRI patterns reflect sub-second activation sequences and reveal replay in human visual cortex. *Nature Communications* 12, 1795
- 119 McClelland, J.L. *et al.* (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* 102, 419–457
- 120 Kumaran, D. *et al.* (2016) What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences* 20, 512–534
- 121 Mnih, V. *et al.* (2015) Human-level control through deep reinforcement learning. *Nature* 518, 529–533
- 122 Gupta, A.S. *et al.* (2010) Hippocampal replay is not a simple function of experience. *Neuron* 65, 695–705
- 123 Buckner, R.L. (2010) The role of the hippocampus in prediction and imagination. *Annu Rev Psychol* 61, 27–48, C1-8
- 124 Tang, W. *et al.* (2020) Multiple time-scales of decision making in the hippocampus and prefrontal cortex.

- 125 Yu, J.Y. *et al.* (2018) Specific hippocampal representations are linked to generalized cortical representations in memory. *Nature Communications* 9, 2209
- 126 Mitra, A. *et al.* (2016) Human cortical-hippocampal dialogue in wake and slow-wave sleep. *Proc Natl Acad Sci U S A* 113, E6868–E6876
- 127 Buzsáki, G. *et al.* (2014) Emergence of Cognition from Action. *Cold Spring Harb. Symp. Quant. Biol.* 79, 41–50
- 128 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
- 129 Gordon, E.M. *et al.* (2016) Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb Cortex* 26, 288–303
- 130 Zorzi, M. *et al.* (2013) Modeling language and cognition with deep unsupervised learning: a tutorial overview. *Front. Psychol.* 4,
- 131 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
- 132 Muckli, L. *et al.* (2015) Contextual Feedback to Superficial Layers of V1. *Current Biology* 25, 2690–2695
- 133 Feldman, H. and Friston, K.J. (2010) Attention, uncertainty, and free-energy. *Front. Hum. Neurosci* 4, 215
- 134 Bastos, A.M. *et al.* (2012) Canonical microcircuits for predictive coding. *Neuron* 76, 695–711
- 135 Pezzulo, G. *et al.* (2015) Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology* 136, 17–35
- 136 Schneider, D.M. (2020) Reflections of action in sensory cortex. *Current Opinion in Neurobiology* 64, 53–59
- 137 Egger, S.W. *et al.* (2019) Internal models of sensorimotor integration regulate cortical dynamics. *Nature Neuroscience* 22, 1871–1882
- 138 Engel, A.K. *et al.* (2013) Where’s the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences* 17, 202–209
- 139 Musall, S. *et al.* (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience* 22, 1677–1686
- 140 Shenoy, K.V. *et al.* (2013) Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* 36, 337–359
- 141 d’Avella, A. *et al.* (2003) Combinations of muscle synergies in the construction of a natural motor behavior. *Nature Neuroscience* 6, 300–308
- 142 Leo, A. *et al.* (2016) A synergy-based hand control is encoded in human motor cortical areas. *Elife* 5,
- 143 Poynter, W. *et al.* (2013) Individuals exhibit idiosyncratic eye-movement behavior profiles across tasks. *Vision Research* 89, 32–38
- 144 Zangrossi, A. *et al.* (2020) , Eye Movements Abnormalities as Early Biomarker of Alzheimer’s Disease: An Ecological Approach. , presented at the Behavioral Neurology
- 145 Pappalardo, L. *et al.* (2015) Returners and explorers dichotomy in human mobility. *Nature Communications* 6, 8166
- 146 Cona, G. *et al.* (2019) Archetypes of human cognition defined by time preference for reward and their brain correlates: An evolutionary trade-off approach. *Neuroimage* 185, 322–334
- 147 Corbetta, M. *et al.* (2015) Common behavioral clusters and subcortical anatomy in stroke. *Neuron* 85, 927–941
- 148 Karolis, V.R. *et al.* (2019) The architecture of functional lateralisation and its relationship to callosal connectivity in the human brain. *Nature Communications* 10, 1417
- 149 Seth, A.K. and Friston, K.J. (2016) Active interoceptive inference and the emotional brain. *Phil. Trans. R. Soc. B* 371, 20160007
- 150 Barrett, L.F. and Simmons, W.K. (2015) Interoceptive predictions in the brain. *Nat Rev Neurosci* 16, 419–429