Title

The ascending arousal system shapes low-dimensional brain dynamics to mediate awareness of changes in intrinsic cognitive states

One Sentence Summary

Fluctuations in brainstem arousal structures coincide with whole-brain topological network fluctuations and low-dimensional state space dynamics that facilitate alterations in attentional awareness.

Authors

Brandon Munn[1,2,†], Eli J. Müller[1,2,†], Gabriel Wainstein[1,2], and James M. Shine[1,2,*]

Affiliations

1 Complex Systems Research Group, The University of Sydney, Sydney, NSW, Australia

2 Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia

* Corresponding author: mac.shine@sydney.edu.au

† Denotes equal contribution

Abstract

Models of cognitive function typically focus on the cerebral cortex, ignoring functional links to subcortical structures. This view neglects the highly-conserved ascending arousal system's role and the computational capacities it provides the brain. In this study, we test the hypothesis that the ascending arousal system modulates cortical neural gain to alter brain dynamics' low-dimensional attractor landscape. Our analyses of spontaneous functional magnetic resonance imaging data and phasic bursts in both locus coeruleus and basal forebrain demonstrate precise time-locked relationships between brainstem activity, low-dimensional energy landscapes, network topology, and spatiotemporal travelling waves. We extend our analysis to a cohort of experienced meditators and demonstrate locus coeruleus-mediated network dynamics were associated with internal shifts in conscious awareness. Together, these results present a novel view of brain organization that highlights the ascending arousal system's role in shaping both the dynamics of the cerebral cortex and conscious awareness.

Main Text

It is often difficult to see the forest for the trees, but to fully understand a concept typically involves an accurate depiction of both. That is, we need to comprehend not only the detailed workings of a specific system, but also how that system functions within a broader context of interacting parts. Modern theories of whole-brain function exemplify this challenge. For instance, activity in the brain has been shown to incorporate signatures of both local computational specificity (e.g., specialized regions within the cerebral cortex) as well as system-wide integration (e.g., the interactions between the cortex and the rest of the brain)[1,2]. Anatomical evidence suggests that the balance between integration and segregation is mediated in part by the relatively fixed white matter connections between cerebral cortical regions[1] – local connectivity motifs support segregated activity, whereas the axonal, re-entrant connections between regions act to integrate the distributed signals via a highly interconnected structural backbone[3]. However, how the human brain is also capable of remarkable contextual flexibility given this relatively fixed connectivity remains poorly understood.

During cognitive tasks, neural activity rapidly reconfigures the functional large-scale network architecture of the brain in order to facilitate coordination between otherwise segregated cortical regions. Precisely how this flexibility is implemented in the brain without altering structural connectivity remains an open question in systems neuroscience. Although it is often overlooked in theories of whole brain function, the neuromodulatory ascending arousal system is well-placed to mediate this role[4]. The arousal system is comprised of a range of nuclei spread across the brainstem and forebrain that send wide-reaching axons to the rest of the central nervous system[5]. At their target sites, arousal neurons release neuromodulatory neurotransmitters that shape and constrain a region's processing mode – altering their excitability and responsivity without necessarily causing them to fire an action potential[4,6]. As a result, subtle changes in the concentration of neuromodulatory chemicals can cause massive alterations in the dynamics of the target regions, leading to nonlinear effects on the coordinated patterns of activity that emerge from 'simple' neuronal circuits[4].

The ascending arousal system also contains substantial heterogeneity – unique cell populations project in diverse ways to the cerebral cortex, and also release distinct neurotransmitters. One key dichotomy is the distinction between adrenergic neuromodulation (predominantly via the

locus coeruleus [LC]), which promote arousal and exploratory behaviour[7], and cholinergic neuromodulation (such as via the basal nucleus of Meynert [BNM]), which are associated with attentional focus and vigilance[8]. These highly interconnected[9] structures both promote wakefulness and arousal[10,11], albeit via distinct topological projections to the cerebral cortex: the LC projects in a diffuse manner that crosses typical specialist boundaries, whereas the BNM projects in a more targeted, region-specific manner[12] (Fig. 1A). The two systems have also been linked with distinct and complimentary computational principles: the noradrenergic LC is presumed to modulate interactions between neurons (response gain)[13], whereas the cholinergic BNM is presumed to facilitate divisive normalization (multiplicative gain)[14]. Based on these anatomical and computational features, we have hypothesized that the interaction between these two neuromodulatory systems is crucial for mediating the dynamic, flexible balance between integration and segregation in the brain[15].

Another crucial feature of the ascending arousal system is that the number of neurons that project to the cerebral cortex is several orders of magnitude smaller than those that project back to the brainstem and forebrain[16–18]. Based on this feature, we further hypothesize that shifts in arousal are realized through a low-dimensional modulation of the ongoing neural activity ('brain state')[17]. Conceptually, low-dimensional brain state dynamics can be depicted as evolving on a brain state energy landscape[29], where the energy of a given state corresponds to the occurrence probability, e.g. high energy brain states have a low occurrence probability (and *v. v.*). That is brain states evolve along the energy landscape topography, much like a ball rolls under the influence of gravity down a valley and requires energy to traverse up a hill, this corresponds to an evolution towards an attractive or repulsive brain state, respectively. This technique can resolve what might otherwise be obscured states of attraction (and repulsion) in a multi-stable system and has been successfully applied to the dynamics of spiking neurons[19,20], blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI)[21,22], and magnetoencephalography (MEG)[23]. The approach offers a number of conceptual advances, but perhaps most importantly, it renders the otherwise daunting task of systems-level interpretation relatively intuitive. Importantly, this framework is not a mere analogy[30], as the topography of the attractor landscape shares a 1-to-1 correspondence with the generative equations required to synthesize realistic neural timeseries data[24]. In this

manuscript, we test these ideas by combining high-resolution resting state fMRI data with analytic techniques from the study of complex systems.

Results

To begin with, we extracted time series data from major subcortical hubs within the noradrenergic LC[9] (Fig. 1A, red) and cholinergic BNM[25] (Fig. 1A, green) systems from 59 healthy participants who had undergone high-resolution, 7T resting-state functional magnetic resonance imaging (fMRI; 2 mm^3 voxels; TR = 586 ms repetition time). Given the known spatiotemporal interactions between the ascending arousal system and fluctuations in cerebrospinal fluid, we first controlled for activity fluctuations in the nearby fourth ventricle, which contains no neural structures, but nonetheless can cause alterations in the BOLD signal over time. Using the residuals from this regression, we calculated the difference between the signals from the LC and BMN ($\tau_{LC\text{-}BNM}$; concatenated across subjects) and then identified time points associated with phasic bursts of LC activity that led to sustained adrenergic (versus cholinergic) influence over evolving brain state dynamics (see Methods). Importantly, the phasic mode of firing within the noradrenergic arousal system has been specifically linked to systemic influences that occur on time-scales relevant to cognitive function[8,26]. Tracking the mean cortical BOLD response around these peaks identified a spatiotemporal travelling wave (Fig. 1B; velocity = $0.13ms^{-1}$) that propagated from frontal to sensory cortices and tracked closely with the known path of the dorsal noradrenergic bundle[9], albeit with a preserved 'island' within the parietal operculum (Fig. 1B). These results confirm that coordinated macroscale activity patterns align to fluctuations in activity within the ascending arousal system of the brainstem[27].

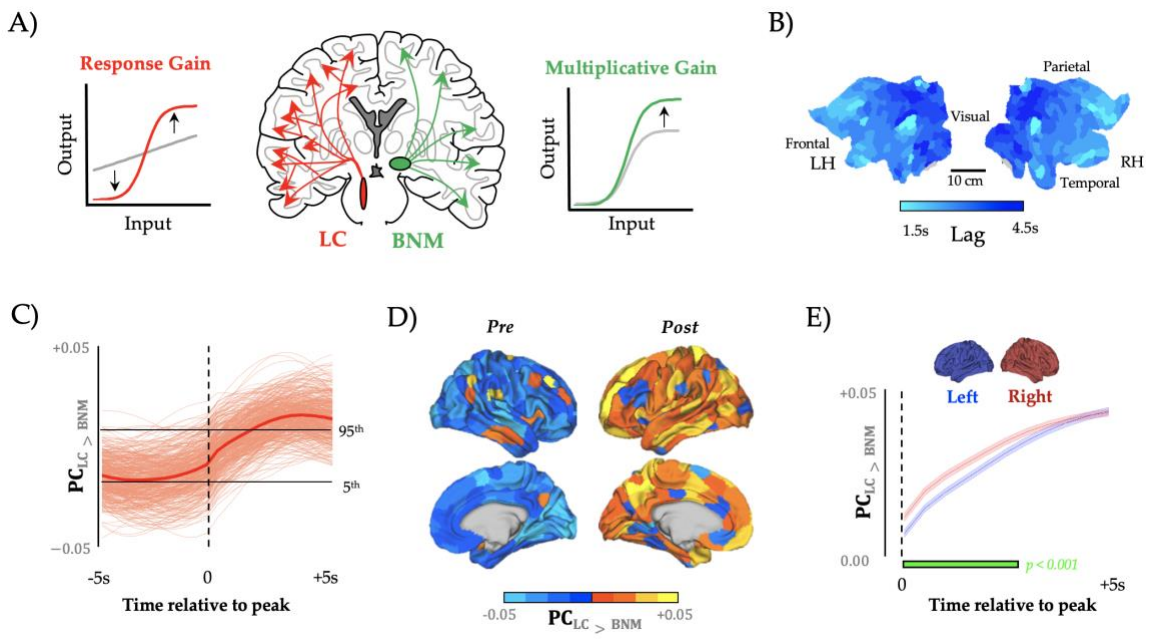

Figure 1. Sympathetic activity precedes network-level integration. A) regional time series were extracted from the locus coeruleus (red), which is thought to alter response gain, and the basal nucleus of Meynert (green), which is thought to alter multiplicative gain, and compared to BOLD signal and topological signatures during the resting state; B) we observed a anterior-to-posterior traveling wave (velocity ~ 0.13ms^{-1}) following peaks in $\tau_{LC\text{-}BNM}$, which are shown on both the left (LH) and right (RH) hemispheres of a cortical flat map; C) the lagged cross-correlation between τ_{LC-BNM} and PC – dotted line depicts the zero-lag correlation, and the black lines depict the upper (lower) bounds of a block-resampled null model; D) mean cortical PC preceding (left) and following (right) the zero-lagged τ_{LC-BNM} value; E) the participation coefficient following peak τ_{LC-BNM} was higher in the right- (red) vs. the left- (blue) hemisphere ($p < 0.001$; green bar); F) τ_{LC-BNM} was preceded by activity in the posterior > anterior hypothalamus.

Time-varying network topology

Based on previous empirical[28], modelling[29] and theoretical[15] work, we predicted that phasic bursts in $\tau_{LC\text{-}BNM}$ would facilitate network-level integration by modulating increased neural gain among regions distributed across the cerebral cortex. As predicted, we observed a strong positive relationship between $\tau_{LC\text{-}BNM}$ and network-level integration ($p < 0.05$; Fig. 1C) across the brain (Fig. 1D). An increase in phasic activity within the LC (relative to the BNM) consistently preceded an increase in the level of integration within the cerebral cortex that was maximal in frontoparietal cortices (and *v.v.*; Fig. S1). Regional integration occurred earlier in the right-hemisphere (Fig. 1E), which is consistent with the known anatomical bias of the LC system[30,31]. A block-resampling null model was applied to ensure that the results were not due

to spatial-autocorrelation ($p < 0.05$; see Methods). Together, these findings provide robust evidence for the hypothesis that the balance between ascending noradrenergic and cholinergic tone facilitates a transition towards topological integration across the distributed network of the brain[15].

Neuromodulation of the Attractor Landscape

The results of our initial analysis confirm that distributed activity in the brain is modulated by small groups of neuromodulatory cells in the brainstem and forebrain, which in turn are proposed to constrain brain dynamics onto a low-dimensional attractor landscape (Fig. 2A). The effects of noradrenaline and acetylcholine can also be easily viewed through this lens: by integrating the brain, noradrenaline should flatten the attractor landscape (Fig. 2A, red), whereas in contrast, the segregative nature of cholinergic activity should act to deepen attractor valleys (Fig. 2A, green). In previous work, we have shown a correspondence between low-dimensional brain state dynamics across multiple cognitive tasks and the heterogenous expression of metabotropic neuromodulatory receptors[17]. This implies that neuromodulators act similar to catalysts in chemical reactions, which lower (or raise) the activation energy (E_A) required to transform chemicals from one steady state (or energy well) to another (Fig. 2B). In the context of the interconnected, heterarchical networks that comprise the cerebral cortex, this would have the effect of flattening (or deepening) the attractor landscape, promoting variable (or rigid) brain states[32] (Fig. 2A).

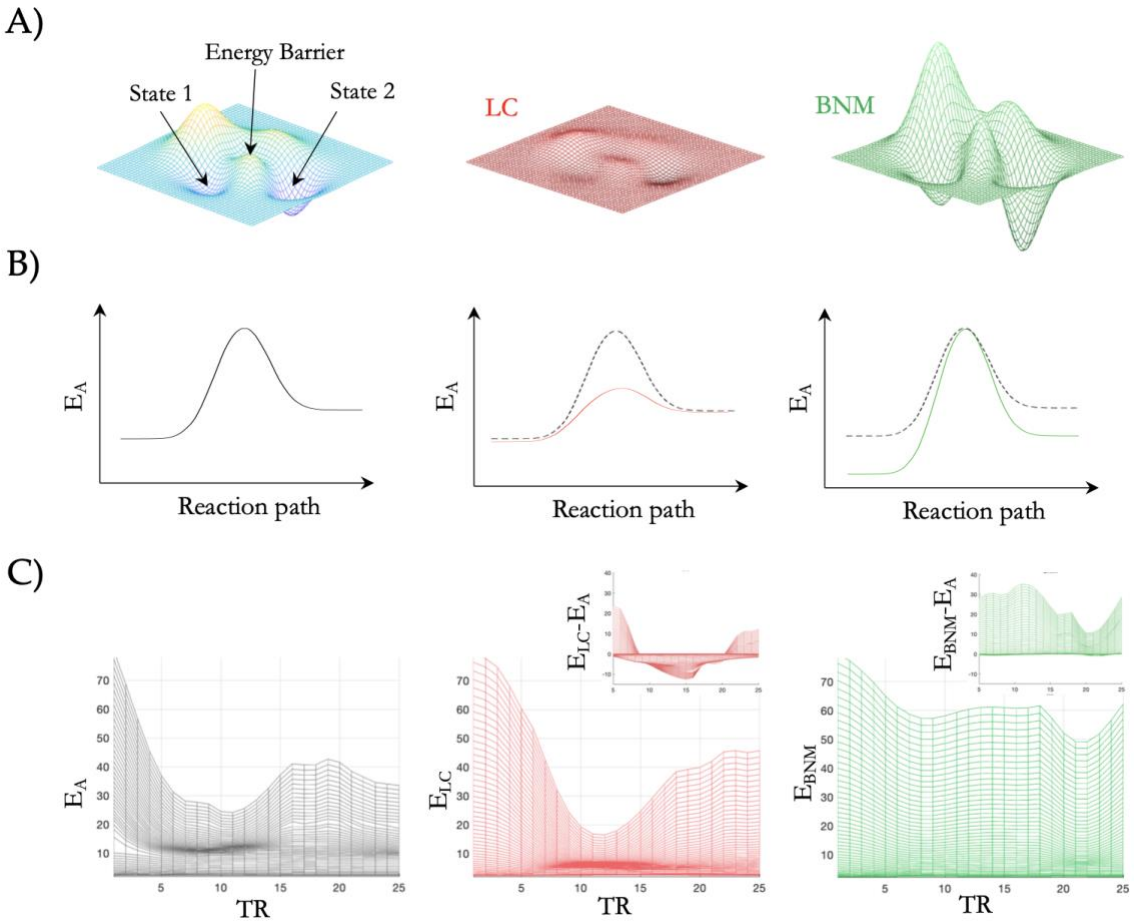

Figure 2 – LC and BNM mediated shifts in attractor landscape brain state space dynamics. A) an attractor landscape, which defines the energy required to move between different brain states: by increasing response gain, the LC should flatten the attractor landscape (red); by increasing multiplicative gain, the BNM should deepen the attractor wells (green); B) the topography of the attractor landscape can be conceptualized as similar to the activation energy (E_A) that must be overcome in order to catalyze the conversion of one chemical to another; C) Empirical brain state transition energy as a function of MSD and TR of the baseline activity (E_A, Left) and after phasic bursts in LC (red) and BNM (green) – relative to the baseline energy landscape phasic bursts in LC (red inset) lead to a flattening or reduction of the energy landscape, whereas peaks in BNM (green inset) lead to a raising of the energy landscape.

To elucidate the role of phasic activity from the neuromodulatory system in modifying the attractor landscape, we first estimated the energy landscape of BOLD signal dynamics across the cerebral cortex. Importantly, the term 'energy' here is used in reference to its definition in statistical physics and hence does not represent the biological use of the term, which instead stands for the energy used by the brain to maintain or change neural activity. Specifically, we define the energy, $E(ds, dt)$, as the inverse probability of observing a given BOLD state

change, $E(ds, dt) = 1/\mathrm{P}(ds, dt)$, where $ds = \langle |\boldsymbol{x}_{t+dt} - \boldsymbol{x}_t|^2 \rangle_r$ is the mean-squared displacement (MSD) of BOLD neural activity, $\boldsymbol{x}_t = [x_{1,t}, x_{2,t}, \dots, x_{r,t}]$across r voxels and dt is the number of timesteps of size TR from the phasic burst at time t. In this framework, a highly probable relative change in BOLD (as quantified by the MSD) corresponds to a relatively low energy transition (i.e., low E_A), whereas an infrequently visited state will require the most energy (i.e., high E_A).

By treating energy as inversely proportional to the probability of brain state occurrence, our approach resembles other studies that have been applied to spiking dynamics of neuronal populations, spiking neurons[19,20], BOLD fMRI[21,22], MEG[23], and natural scene[33]. However, these studies binarized continuous signals in order to reduce the brain state space (to 2^r states), however this approach requires the fitting of a threshold, which can be problematic in continuously recorded data. In contrast, our approach reduces the dimensionality by analysing the likelihood of a change in BOLD activity (i.e., the MSD), and thus retains the dimensionality of the underlying signal without the need for thresholding. This coordinate transformation – which switches from an absolute state space to a relative state space – is akin to switching from a 'birds-eye' or 3rd person view to one which views the attractor landscape from the perspective of the current brain state (i.e., a 'brain-centric' or 1st person view). This approach overcomes a major limitation inherent to the birds-eye view approach which requires a large sample size to sufficiently sample the brain state space.

With this in mind we turned our attention to the relationship between attractor landscape dynamics and the ascending arousal system. To test the hypothesis that the neuromodulatory system alters the topography of the attractor landscape, we calculated BOLD MSD landscape energetics following phasic bursts of both LC and BNM (see Methods) and compared these relative to sampled brain evolutions without LC and BNM arousal (baseline). The attractor landscapes for these three states are defined by the energy for a given MSD and TR. Figure. 2C demonstrates the baseline attractor landscape, which corresponds to the reaction pathway in Fig. 2B. As expected, for short TR intervals, small MSD were energetically preferable to large MSD – i.e., there were only small deviations in the BOLD signal (MSD is projected into the page see Fig. S3 for alternate view). Further analysis revealed an energetically preferable

(i.e., attractor) state at 10 TR and 30 MSD (Fig. 2B, black). Confirming our hypothesis for the role of LC and BNM (Fig. 2A), the energy for this secondary attractor is lowered (raised) after an LC (BNM) phasic burst, relative to baseline (Fig. 2B, red and green, respectively). Specifically, LC activity (relative to baseline; Fig. 2C, red inset) flattened the energy landscape, thus making the second attractor state and previously unlikely trajectories far more probable, whereas BNM activity (relative to baseline; Fig. 2C, green inset) caused the energy landscape to be elevated, thus promoting local trajectories and making large MSD trajectories energetically expensive. These patterns are analogous to modulating a physical landscape in which towns sit within valleys separated by impassable mountains – when BNM is high, the towns remain isolated, whereas when LC is high, the towns are separated by easily navigated rolling plains and novel combinations of consecutive brain states can be realised.

We next asked whether LC and BNM combined together synergistically to alter the attractor landscape. To achieve this, we isolated simultaneous phasic peaks in both LC and BNM (τ_{LC+BNM}). We found that the LC + BNM attractor landscape differed markedly from either independent LC or BNM activation, shifting the brain state into divergent regimes than could be explained by the HRF. By comparing the MSD energy topography for a given TR slice we found that the landscape switched from an anti-correlation to uncorrelated with the HRF. In other words, the cooperative behaviour between the noradrenergic and cholinergic systems allowed the brain to reach unique states that neither could facilitate individually. To examine how simultaneous LC+BNM activity altered the attractor landscape, we compared the energy relative to the two individual landscapes. As demonstrated in Fig. S4, the attractor landscape following phasic bursts of LC+BNM differed in magnitude from that expected from a linear superposition of the LC and BNM attractor landscape – i.e., LC+BNM $\neq$ (LC) + (BNM). Furthermore, to explore the dominance of either LC or BNM in this signal, we minimised the relationship LC+BNM $= \alpha$ LC $+ \beta$ BNM (conditional upon α and β being positive constants) and found that $\alpha = 0.16$ and $\beta = 0.84$ gave the best match to the LC+BNM energy landscape. That is, the BNM attractor landscape dominates the simultaneous LC+BNM attractor, which is consistent with the unidirectional synaptic projections from the LC that synapse upon the BNM on their way through to the cortex[12], and suggests that phasic LC+BNM bursts may be initiated by the LC in order to elicit a cascade of BNM activity.

Conscious awareness of shifts in attractor landscape

Interpreting the relationship between neuroimaging data and conscious awareness is notoriously challenging. For instance, it is currently not possible to directly determine the contents of self-directed thought without intervening, and thus altering, the contents of consciousness[34]. Although we can't determine the contents of consciousness directly, we can use task designs to modulate the state of consciousness. To this end, we leveraged data from a group of 14 expert meditators who were asked to meditate during an fMRI scanning session[35], and to press a button when they noticed that their focus had drifted from their breath (Fig. 3A). At this point, there is a mismatch between expectation and conscious awareness, which is an internal state that has been previously linked to the activation of the noradrenergic system, both in theoretical[36,37]and computational[38] work. Based on these studies, we predicted that the switch in internal conscious awareness would be facilitated by increases in locus coeruleus-mediated integration and subsequent reconfiguration of low-dimensional brain states. Analysing time-resolved network data with a finite impulse response model, we observed a peak in locus coeruleus activity (Fig. 3B), TR-to-TR mean squared displacement (Fig. 3C) and elevated network-level integration (Fig. 3D) surrounding the change in conscious awareness (all $p_{PERM} < 0.01$). These results confirm that the locus coeruleus mediates energy landscape reconfigurations and that these changes modulate internal states of conscious awareness.

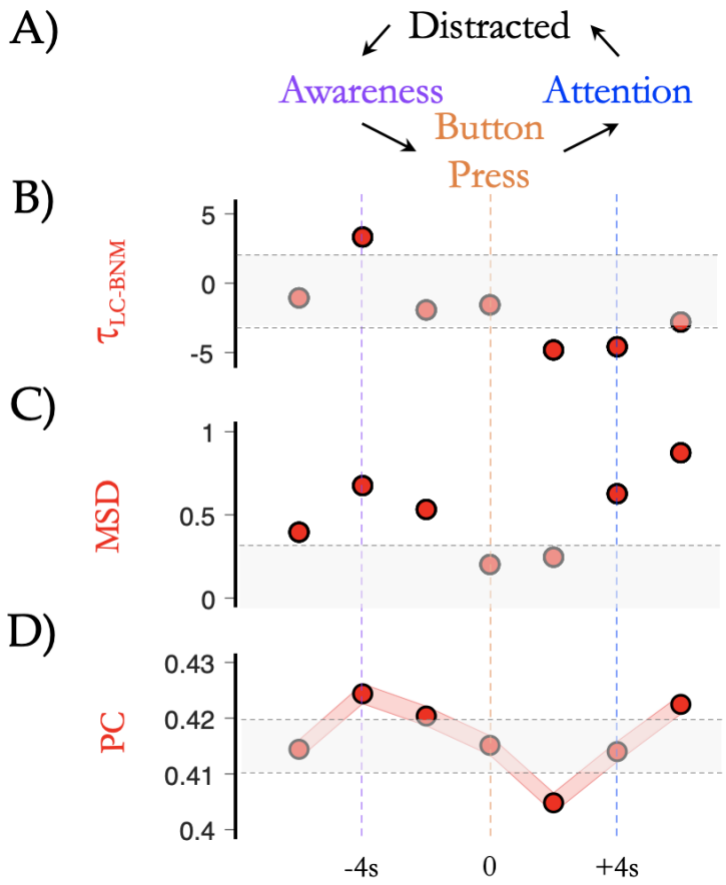

Figure 3 – Awareness of intrinsic state changes. A) participants performing breath-awareness meditation (Focus; blue) were trained to respond with a Button Press (orange) when they became Aware (purple) that they had become Distracted (i.e., their attention had wandered from their breath) and to then re-focus their Attention (blue) on their breath; B) we observed a peak in τ_{LC-BNM} (i.e., LC > BNM; red) ~4 seconds before the button press, which then returned to low levels (i.e., BNM > LC) in the 2-4 seconds following the button press; C) the Mean Squared Displacement (MSD; dark orange) of TR-to-TR BOLD signal was increased above null values around the peak in τ_{LC-BNM}, as well as following the re-establishment of attentional focus (in panels B & C, grey shading depicts 95th percentile of block-resampled null distribution); D) we observed a peak in mean participation coefficient (PC) ~4 seconds (2 TRs) prior to the Button Press during the task. All: grey shading depicts 95th percentile of null distribution i.e., outside grey shading indicates a value different than null [$p < 0.05$]; and lower: red shading represents SEM error bars).

Discussion

Our results provide evidence for an arousal-mediated macroscopic network and attractor landscape reconfiguration which tracks with moment-to-moment alterations in conscious

awareness. By tracking fluctuations in BOLD within the noradrenergic LC and the cholinergic BNM, we were able to demonstrate fundamental ways in which the low-dimensional, dynamic and topological signature of cortical dynamics was directly related to changes within the ascending arousal system. Furthermore, we demonstrated a link between these dynamic reconfigurations and alterations in conscious awareness in a cohort of experienced meditators. In this way, our results provide a novel, systems-level perspective on the distributed dynamics of the human brain.

There is growing evidence that distributed neural dynamics in the brain are well described by relatively low-dimensional models[17,18,39–41], however the biological constraints that impose these features on the brain remain poorly understood. Due to the low number of cells in the arousal system and their broad projections to the rest of the brain, we theorized that neuromodulatory regions are well-placed to shape and constrain the vast number of neurons in the cerebral cortex into low-dimensional dynamic modes. Our results confirm this prediction by showing that patterns of activity in key regions within the brainstem and forebrain relate to fundamental alterations in a dynamically evolving attractor landscape. In other words, a brain's state space is a powerful framework that extends beyond that of mere analogy, and the ascending arousal system is well-placed to mediate deformations in the attractor landscape.

Much in the same way that there are many different projections of physical maps, each with their own strengths and weaknesses (e.g., Mercator projections that preserve shape at the cost of size and distance). In this study, we utilized a novel 'brain-state centric' or '1st person' view for tracking the architecture of the attractor landscape over time. This approach conceptualized attractor landscape trajectories relative to the current state, with the energy required to navigate the current brain-state to a different spatiotemporal brain-state inversely proportional to how frequently this transition occurred. Through this coordinate transformation, the role of adrenergic and cholinergic neuromodulation in flattening and deepening the landscape was clearly visible. An alternative approach as is typically performed is to reduce the dimensionality of brain-state dynamics, coarse-grain this low dimensional embedding space, and then count the occurrence of each novel state into a 'birds-eye' or '3rd person' view. A significant downside of this technique is it requires significant sampling of the

parameter space, i.e., long recordings, or heavy coarse-graining, i.e., binarization. As our novel technique pools over states, it requires significantly less sampling of the brain-state space. Furthermore, the clear result as hypothesized suggests that the 1st person view may be the intrinsic view for the ascending arousal system. This relative 1st person view is consistent with the role of the phylogenetically ancient system, in overriding ongoing activity to avoid imminent danger from the zebrafish to the human. However, a pitfall of this 1st person view is that the flight-or-flight response can be engaged unnecessarily overriding ongoing dynamics, irrespective of their contents, e.g., during public speaking. In contrast, spatially grounded cortical areas such as the early visual system or hippocampal place cells would likely follow the 3rd person view as the relative location of neural dynamics matters. Given improvements in recording length and novel analytic techniques to probe the brains dynamical landscape, we expect that the field will ultimately discover even more optimal mappings between neurobiology and low-dimensional brain state dynamics.

The results of our state-space analysis have important implications for the biological mechanisms underlying cognition. For instance, the concept of locus coeruleus-mediated attractor landscape flattening is reminiscent of the α1 receptor-mediated notion of a 'network reset'[36]. By increasing response gain (Fig. 1A) through the modulation of second-messenger cascades[4], noradrenaline released by the LC would augment inter-regional coordination[29]. Importantly, this capacity could confer adaptive benefits across a spectrum, potentially facilitating the formation of flexible coalitions in precise cognitive contexts[42], while also forcing a broader landscape flattening (i.e., a 'reset') in the context of large, unexpected changes[26,36]. Similarly, the idea that phasic cholinergic bursts deepens attractor wells is consistent with the idea that the cholinergic system instantiates divisive normalization within the cerebral cortex[14]. Numerous cognitive neuroscience studies have shown that heightened acetylcholine levels correspond to improvements in attentional precision[8,8]. By deepening the wells of a specific attractor, acetylcholine from the BNM could in effect be acting so as to ensure that the brain remains trapped within a particular state and is hence not diluted by other (potentially distracting) brain states. Determining the specific rules that govern the links between the neuromodulation of the attractor landscape and cognitive function[43–45] is of paramount importance, particularly given the highly integrated and degenerate nature of the ascending arousal system[46].

Our results also provide a systems-level perspective on an emerging corpus of work that details the microscopic circuit level mechanisms responsible for conscious phenomena[47]. In particular, a number of recent studies have highlighted the intersection between the axonal projections of the ascending arousal system and pyramidal cell dendrites in the supragranular regions of the cerebral cortex as a key site for mediating conscious awareness. For instance, optogenetic blockage of the connections between the cell bodies and dendrites of thick-tufted layer V pyramidal cells in the sensory cortex causally modulated conscious arousal in mice[48]. Other work has shown that both the noradrenergic[49] and cholinergic[50] systems alter this mechanism, albeit in distinct ways: noradrenaline would promote burst firing due to the α2a receptor-mediated closure of *Ih* HCN leak-channels[49], whereas the cholinergic system instead prolongs the time-scale of firing via M1 cholinergic receptor activation on pyramidal cell dendrites[50]. In this way, coordinated activity in the ascending arousal system can mediate alterations in microcircuit processing that ultimately manifest as alteration in macroscopic brain network dynamics.

The vascular nature of the T2* fMRI signal is such that it is impossible to rule out the role of haemodynamics in the results we obtained in our analysis. Indeed, there is evidence that noradrenaline causes a targeted hyperaemia through the augmentation of G-protein-coupled receptors on vascular smooth muscle cells[51,52]. However, it is also clear that the haemodynamics and massed neural action in the cerebral cortex are inextricably linked[53,54]. In addition, there is evidence that stimulation of the locus coeruleus leads to the high-frequency, low-amplitude electrophysiological activity patterns characteristic of the awake state[10]. Together, these results argue that the locus coeruleus mediates a combination of haemodynamic and neural responses that facilitate integrative neural network interactions and subsequently mediate alterations in conscious awareness.

In this manuscript, we have argued that the ascending arousal system provides crucial constraints over normal brain function, however there are numerous examples wherein pathology within the ascending arousal system leads to systemic impairments in cognition. In addition to disorders of consciousness[55], dementia syndromes are also crucially related to dysfunction within the ascending arousal system. For instance, Alzheimer's disease has been

linked to tau pathology within the BNM[25], however individuals with Alzheimer's disease also often have pathological involvement of the LC as well[56]. Similarly, individuals with Parkinson's disease often have extra-dopaminergic pathology in the LC[57], as well as in the cholinergic tegmentum[58]. Given the pathological processes at play in these disorders, we expect that other neuromodulatory systems will also be impaired, and in turn effect the macroscopic dynamics of the system in ways that remain to be elucidated.

In conclusion, we leveraged a high-resolution 7T resting state fMRI dataset to test the hypothesis that activity within the ascending arousal system shapes and constrains patterns of systems-level network reconfiguration. Our results confirm specific predictions from a recent hypothetical framework[15], and further delineate the manner in which the autonomic nervous system shapes and constraints ongoing, low-dimensional brain state dynamics in the central nervous system in a manner that supports changes in conscious awareness.

Methods

7T resting state fMRI

Sixty-five healthy, right-handed adult participants (mean, 23.35 years; SD, 3.6 years; range 18–33 years; 28 females) were recruited, of whom 60 were included in the final analysis (four participants were excluded due to MR scanning issues, one participant was excluded due to an unforeseen brain structure abnormality). Participants provided informed written consent to participate in the study. The research was approved by The University of Queensland Human Research Ethics Committee. These data were originally described in Hearne et al., 2017. 1050 (~10 minutes) whole-brain 7T resting state fMRI echo planar images were acquired using a multiband sequence (acceleration factor = 5; 2 mm^3 voxels; 586 ms TR; 23 ms TE; 40^0 flip angle; 208 mm FOV; 55 slices). Structural images were also collected to assist functional data pre-processing (MP2RAGE sequence – 0.75 mm^3 voxels 4,300 ms TR; 3.44 ms TE; 256 slices).

DICOM images were first converted to NIfTI format and realigned. T1 images were reoriented, skull-stripped (FSL BET), and co-registered to the NIfTI functional images using statistical parametric mapping functions. Segmentation and the DARTEL algorithm were used to improve the estimation of non-neural signal in subject space and the spatial normalization. From each grey-matter voxel, the following signals were regressed: linear trends, signals from the six head-motion parameters (three translation, three rotation) and their temporal derivatives, white matter, and CSF (estimated from single-subject masks of white matter and CSF). The aCompCor method (Behzadi et al., 2007) was used to regress out residual signal unrelated to neural activity (i.e., five principal components derived from noise regions- of-interest in which the time series data were unlikely to be modulated by neural activity). Participants with head displacement > 3 mm in > 5% of volumes in any one scan were excluded (n = 5). A temporal band pass filter ($0.071 < f < 0.125$ Hz) was applied to the data.

Brain parcellation

Following pre-processing, the mean time series was extracted from 400 pre-defined cortical parcels using the Schaefer atlas (Schaefer et al., 2018). Probabilistic anatomical atlases were used to define the location of the noradrenergic LC[59] and the cholinergic BNM (Ch4 cell group)[60]. The mean signal intensity from each region was extracted and then used for

subsequent analyses. To ensure that the BOLD data were reflective of neuronal signals, we statistically compared LC and BNM time series with a number of potential nuisance signals from: i) the cerebrospinal fluid; ii) the cortical white matter; iii) mean framewise displacement; and iv) a $2mm^3$ sphere in the fourth ventricle (centred at MNI co-ordinates: 0 -45 -30)[61]. All signals were unrelated to LC and BNM activity ($|r| < 0.05$ in each case), however given the spatial proximity of the LC to the fourth ventricle, we opted to use a linear regression to residualize the signal from the fourth ventricle.

Phasic increases in neuromodulatory BOLD signal

To identify phasic increases in neuromodulatory BOLD signal, we calculated the second derivative (i.e., the acceleration) of the LC and BNM time series, and then identified points in time that fulfilled three criteria: 1) value greater than or equal to 2 s.d. above the mean acceleration; 2) value of the original time series, i.e., LC or BNM, was greater than or equal to 2 s.d. above the mean of the time series within the following 10 TRs (i.e., 5.8 seconds); and 3) the time point was not present within the first or last 20 TRs of an individual subjects' trial (so as to avoid potential boundary effects). Using these criteria, we identified 148 $\tau_{LC\text{-}BNM}$ time points, 130 $\tau_{BNM\text{-}LC}$ time points and 316 τ_{LC+BNM} time points across all 59 subjects. To ensure that the choice of 2 s.d. threshold was reflective of the underlying dynamics, we altered this threshold between 1-3 s.d. and found robustly similar patterns. For subsequent analyses, we identified time points in the 21 TR window surrounding these peaks, and then used these to conduct statistical comparisons of the low-dimensional, complex network signature of brain network dynamics as a function of phasic ascending arousal system activity. Each of these patterns was confirmed using a lag-based cross-correlation analysis, which demonstrated similar phenomena to those that we present in the manuscript.

To monitor the propagation of cortical signals with respect to $\tau_{LC\text{-}BNM}$, $\tau_{BNM\text{-}LC}$ and τ_{LC+BNM}, we extracted the time-to-peak of the cross-correlation between these signals and each of the 400 cortical parcels within the 10 TR (i.e., 5.8 second) windows following each identified phasic peak. These patterns were mapped onto the cortex (Fig. 1B) for visualization and clearly demonstrated anterior-to-posterior direction for the wave. We then used the volumetric MNI co-ordinates of the Schaefer parcellation scheme to calculate the average velocity of the travelling wave ($0.13m\ s^{-1}$).

In order to obtain an appropriate null model against which to compare our data, we identified 5,000 random timepoints within the concatenated dataset that did not substantially overlap with the already identified $\tau_{LC\text{-}BNM}$, $\tau_{BNM\text{-}LC}$ and τ_{LC+BNM} time series, and used these to populate a null distribution[62]. Outcome measures were deemed significant if they were more extreme than the 95th (or 5th) percentile of the null distribution. Crucially, this ensured that our data could not be explained by the characteristic spatial and temporal autocorrelation present in BOLD timeseries data.

Modularity Maximization

The Louvain modularity algorithm from the Brain Connectivity Toolbox (BCT[73]) was used on the neural network edge weights to estimate community structure. The Louvain algorithm iteratively maximizes the modularity statistic, Q, for different community assignments until the maximum possible score of Q has been obtained:

$$Q_T = \frac{1}{v^+}\sum_{ij}\left(w_{ij}^+ - e_{ij}^+\right)\delta_{M_iM_j} - \frac{1}{v^+ + v^-}\sum_{ij}\left(w_{ij}^- - e_{ij}^-\right)\delta_{M_iM_j},$$

where v is the total weight of the network (sum of all negative and positive connections), w_{ij} is the weighted and signed connection between regions i and j, e_{ij} is the strength of a connection divided by the total weight of the network, and δ_{MiMj} is set to 1 when regions are in the same community and 0 otherwise. '+' and '–' super-scripts denote all positive and negative connections, respectively. The modularity of a given network is therefore a quantification of the extent to which the network may be subdivided into communities with stronger within-module than between-module connections.

For each epoch, we assessed the community assignment for each region 500 times and a consensus partition was identified using a fine-tuning algorithm from the Brain Connectivity Toolbox (BCT; http://www.brain-connectivity-toolbox.net/). We calculated all graph theoretical measures on un-thresholded, weighted and signed connectivity matrices[73]. The stability of the γ parameter was estimated by iteratively calculating the modularity across a range of γ values (0.5-2.5; mean Pearson's r = 0.859 +-0.01) on the time-averaged connectivity matrix for each subject – across iterations and subjects, a γ value of 1.0 was found to be the least variable, and hence was used for the resultant topological analyses.

Participation Coefficient

The participation coefficient, *PC*, quantifies the extent to which a region connects across all modules (i.e. between-module strength) and has previously been used to successfully characterize hubs within brain networks (e.g. see [75]). The PC for each region was calculated within each temporal window as,

$$\mathrm{PC} = 1 - \sum_{s=1}^{n_M} \left(\frac{\kappa_{isT}}{\kappa_{iT}} \right)^2$$

where k_{isT} is the strength of the positive connections of region *i* to regions in module *s* at time *T*, and k_{iT} is the sum of strengths of all positive connections of region *i* at time *T*. Negative connections were discarded prior to calculation. The participation coefficient of a region is therefore close to 1 if its connections are uniformly distributed among all the modules and 0 if all of its links are within its own module.

Principal Component Analysis

Pre-processed data from each subject was concatenated to form a multi-subject time series and a spatial PCA was performed on the resultant data[16]. The time series of each PC was then estimated by calculating the weighted mean of the group-level BOLD time series associated with each respective principal component. To aid inference, group-level tPC time series were calculated by taking the mean for each PC time series across all subjects. This and all subsequent code are freely available at http://github.com/macshine/bsi/.

Brain State Displacement and the Energy Landscape

To quantify the change in BOLD activity following phasic bursts of neuromodulation we calculated the BOLD mean-squared displacement (MSD). The MSD is a measure of the deviation in BOLD activity, $\boldsymbol{x}_t = [x_{1,t}, x_{2,t}, \ldots, x_{r,t}]$ for r parcels, with respect to the activity at the phasic onset, t_0. The MSD is calculated as the average change of each voxel

$$MSD = \langle \left| \boldsymbol{x}_{t+dt} - \boldsymbol{x}_{t_0} \right|^2 \rangle_r,$$

The MSD is calculated for different t_0 across dt TRs. The energy of each brain state displacement, $E(ds, dt)$, is then equal to the inverse probability, $\mathrm{P}(ds, dt)$, of its occurrence:

$$E(ds, dt) = \frac{1}{\mathrm{P}(ds, dt)}.$$

Meditation Dataset

Fourteen healthy right-handed non-smoking meditation practitioners (11 female; age 28-66) underwent Siemens 3T MRI scanning (T1: TR = 2600 msec, TE = 3.9 msec, TI = 900 msec, FOV = 24 cm, 256 x 256 matrix, voxel dimensions = 1 x 1 x 1 mm^3; T2*: weighted gradient-echo pulse sequence, TR = 1500 msec, TE = 30 msec, flip angle = 90 deg, FOV = 192 cm, 64 x 64 matrix, voxel dimensions = 3 x 3 x 4 mm^3). All participants signed a consent form approved by the Institutional Review Board at Emory University and the Atlanta Veterans Affairs Research and Development Committee as an indication of informed consent. Participants were asked to meditate for 20 min in the MRI scanner by maintaining focused attention on the breath and keeping the eyes closed. They were instructed to press a button whenever they realized their mind had wandered away from the breath, and then return their focus to the breath. The epoch of time immediately prior to the button press was thus the moment in time in which each individual recognized that their focus had deviated from their breath. This information was used to construct a finite impulses response model that mapped the 5 TRs prior-to and following each button press. We then modelled LC>BNM activity, low-dimensional dynamics and network topology around this epoch to construct a model of state-space reconfiguration as a function of intrinsic conscious awareness. Non-parametric, block-resampling null distributions were utilized for statistical testing ($p < 0.05$).

Supplementary Figures

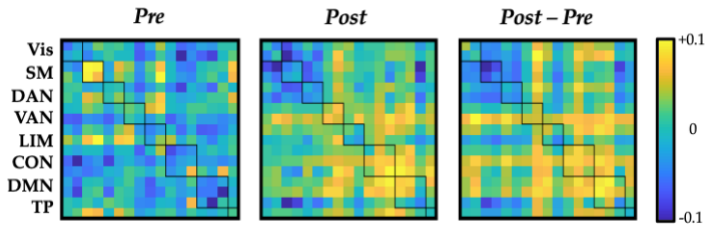

Figure S1. Time-varying correlations. Average correlation preceding (left) and following (middle) the zero-lagged τ_{LC-BNM} value, along with the difference between the two (right); squares represent eight pre-defined sub-networks: Vis – visual, SM – somatomotor, DAN – dorsal attention, VAN – ventral attention, LIM – limbic, CON – control, DMN – default and TP – temporal pole.

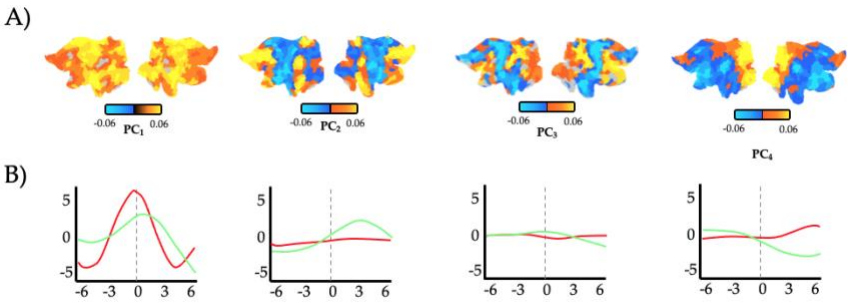

Figure S2. Low-dimensional peri-phasic time-series. A) un-thresholded flat map projections of $PC_{1\text{-}4}$; B) time-locked patterns of $PC_{1\text{-}4}$ with respect to $\tau_{LC\text{-}BNM}$ (red) and $\tau_{BNM\text{-}LC}$ (green).

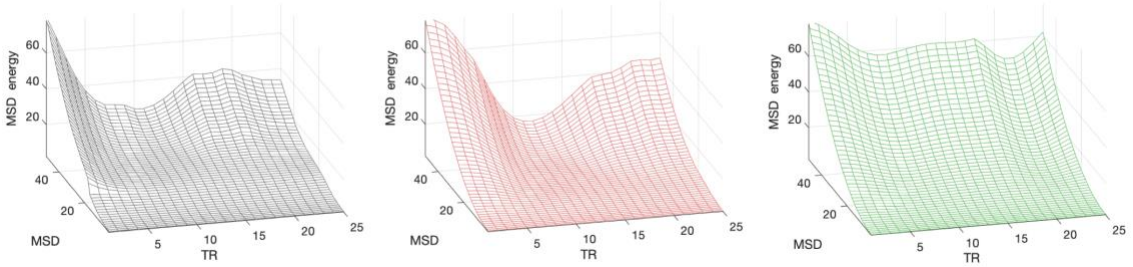

Figure S3. Energy landscape of baseline neural dynamics and under neuromodulatory tone. State transition energy as a function of MSD and TR of the baseline activity (black) and after phasic bursts in LC (red) and BNM (green)

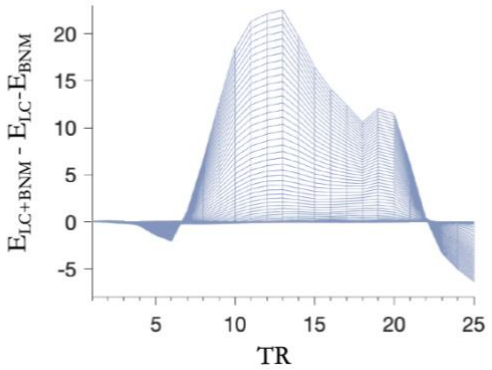

Figure S4 – LC and BNM move dynamics to differing regimes than unaroused activity and their simultaneous combination LC+BNM. The attractor landscape of simultaneous LC+BNM phasic bursts relative to their linear superposition, suggesting the simultaneous combination may allow the system to reach particularly unique brain-states that neither individually could reach.

References and Notes

1. Sporns, O. Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology* **23**, 162–171 (2013).
2. Shine, J. M. The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. *Progress in Neurobiology* 101951 (2020) doi:10.1016/j.pneurobio.2020.101951.
3. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nature Reviews Neuroscience* **13**, 336–349 (2012).
4. Shine, James M *et al.* Computational models link cellular mechanisms of neuromodulation to large-scale brain dynamics. *Nature Neuroscience* **Ahead of Print**,.
5. Brown, E. N., Purdon, P. L. & Van Dort, C. J. General Anesthesia and Altered States of Arousal: A Systems Neuroscience Analysis. *Annu. Rev. Neurosci.* **34**, 601–628 (2011).
6. Salinas, E. & Sejnowski, T. J. Book Review: Gain Modulation in the Central Nervous System: Where Behavior, Neurophysiology, and Computation Meet. *Neuroscientist* **7**, 430–440 (2001).
7. Sara, S. J. The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience* **10**, 211–223 (2009).
8. Hasselmo, M. E. & Sarter, M. Modes and models of forebrain cholinergic neuromodulation of cognition. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **36**, 52–73 (2011).
9. Samuels, E. R. & Szabadi, E. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II:

physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current neuropharmacology* **6**, 254–285 (2008).

10. Carter, M. E. *et al.* Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nature Neuroscience* **13**, 1526–1533 (2010).
11. Lin, S.-C., Brown, R. E., Hussain Shuler, M. G., Petersen, C. C. H. & Kepecs, A. Optogenetic Dissection of the Basal Forebrain Neuromodulatory Control of Cortical Activation, Plasticity, and Cognition. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**, 13896–13903 (2015).
12. Zaborszky, L. *et al.* Neurons in the basal forebrain project to the cortex in a complex topographic organization that reflects corticocortical connectivity patterns: an experimental study based on retrograde tracing and 3D reconstruction. *Cerebral cortex (New York, N.Y. : 1991)* **25**, 118–137 (2015).
13. Servan-Schreiber, D., Printz, H. & Cohen, J. A network model of catecholamine effects- Gain, signal-to-noise ratio, and behavior. *Science* **249**, 892–895 (1990).
14. Schmitz, T. W. & Duncan, J. Normalization and the Cholinergic Microcircuit: A Unified Basis for Attention. *Trends in Cognitive Sciences* **22**, 422–437 (2018).
15. Shine, J. M. Neuromodulatory Influences on Integration and Segregation in the Brain. *Trends in Cognitive Sciences* **23**, 572–583 (2019).
16. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience* **17**, 1500–1509 (2014).
17. Shine, J. M. *et al.* Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat Neurosci* **22**, 289–296 (2019).

18. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat Neurosci* (2020) doi:10.1038/s41593-019-0555-4.
19. Tkačik, G. *et al.* Thermodynamics and signatures of criticality in a network of neurons. *Proc Natl Acad Sci USA* **112**, 11508–11513 (2015).
20. Bialek, W. Perspectives on theory at the interface of physics and biology. *Rep. Prog. Phys.* **81**, 012601 (2018).
21. Watanabe, T. *et al.* A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nat Commun* **4**, 1370 (2013).
22. Ezaki, T., Watanabe, T., Ohzeki, M. & Masuda, N. Energy landscape analysis of neuroimaging data. *Phil. Trans. R. Soc. A.* **375**, 20160287 (2017).
23. Krzemiński, D. *et al.* Energy landscape of resting magnetoencephalography reveals fronto-parietal network impairments in epilepsy. *Network Neuroscience* **4**, 374–396 (2020).
24. Breakspear, M. Dynamic models of large-scale brain activity. *Nature Neuroscience* **20**, 340–352 (2017).
25. Mesulam, M. M. Cholinergic circuitry of the human nucleus basalis and its fate in Alzheimer's disease. *Journal of Comparative Neurology* **521**, 4124–4144 (2013).
26. Aston-Jones, G. & Cohen, J. D. An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience* **28**, 403–450 (2005).
27. Muller, L., Chavane, F., Reynolds, J. & Sejnowski, T. J. Cortical travelling waves: mechanisms and computational principles. *Nat Rev Neurosci* **19**, 255–268 (2018).

28. Shine, J. M. *et al.* The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* **92**, 544–554 (2016).

29. Shine, J. M., Aburn, M. J., Breakspear, M. & Poldrack, R. A. The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *Elife* **7**, e31130 (2018).

30. Robinson, R. Differential behavioral and biochemical effects of right and left hemispheric cerebral infarction in the rat. *Science* **205**, 707–710 (1979).

31. Tona, K.-D. *et al.* In vivo visualization of the locus coeruleus in humans: quantifying the test–retest reliability. *Brain Struct Funct* **222**, 4203–4217 (2017).

32. Kanamaru, T., Fujii, H. & Aihara, K. Deformation of attractor landscape via cholinergic presynaptic modulations: a computational study using a phase neuron model. *PloS one* **8**, e53854 (2013).

33. Munn, B. & Gong, P. Critical Dynamics of Natural Time-Varying Images. *Phys. Rev. Lett.* **121**, 058101 (2018).

34. Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M. & Valera, E. M. Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. *Cerebral cortex (New York, N.Y. : 1991)* **27**, 1831–1840 (2016).

35. Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E. & Barsalou, L. W. Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *NeuroImage* **59**, 750–760 (2012).

36. Sara, S. J. & Bouret, S. Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* **76**, 130–141 (2012).

37. Jahn, C. I., Varazzani, C., Sallet, J., Walton, M. E. & Bouret, S. Noradrenergic But Not Dopaminergic Neurons Signal Task State Changes and Predict Reengagement After a Failure. *Cerebral Cortex* bhaa089 (2020) doi:10.1093/cercor/bhaa089.
38. Sales, A. C., Friston, K. J., Jones, M. W., Pickering, A. E. & Moran, R. J. Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: an Active Inference model. *PLoS computational biology* **15**, e1006267 (2019).
39. Gotts, S. J., Gilmore, A. W. & Martin, A. Brain networks, dimensionality, and global signal averaging in resting-state fMRI: Hierarchical network structure results in low-dimensional spatiotemporal dynamics. *NeuroImage* **205**, 116289 (2020).
40. Shine, J. M. *et al.* The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei. *Neuron* **104**, 849-855.e3 (2019).
41. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
42. Müller, E. J., Munn, B. R. & Shine, J. M. Diffuse neural coupling mediates complex network dynamics through the formation of quasi-critical brain states. *Nat Commun* **11**, 6337 (2020).
43. de Gee, J. W. *et al.* Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. *eLife* **9**, e54014 (2020).
44. Pfeffer, T. *et al.* Catecholamines alter the intrinsic variability of cortical population activity and perception. *PLoS Biology* **16**, e2003453 (2018).
45. Kosciessa, J. Q., Lindenberger, U. & Garrett, D. D. *Thalamocortical excitability adjustments guide human perception under uncertainty*.

http://biorxiv.org/lookup/doi/10.1101/2020.06.22.165118 (2020) doi:10.1101/2020.06.22.165118.

46. Brezina, V. Beyond the wiring diagram: signalling through complex neuromodulator networks. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**, 2363–2374 (2010).
47. Aru, J., Suzuki, M. & Larkum, M. E. Cellular Mechanisms of Conscious Processing. *Trends in Cognitive Sciences* **24**, 814–825 (2020).
48. Suzuki, M. & Larkum, M. E. General Anesthesia Decouples Cortical Pyramidal Neurons. *Cell* **180**, 666-676.e13 (2020).
49. Labarrera, C. *et al.* Adrenergic Modulation Regulates the Dendritic Excitability of Layer 5 Pyramidal Neurons In Vivo. *Cell reports* **23**, 1034–1044 (2018).
50. Williams, S. R. & Fletcher, L. N. A Dendritic Substrate for the Cholinergic Control of Neocortical Output Neurons. *Neuron* **101**, 486-499.e4 (2019).
51. Bekar, L. K., Wei, H. S. & Nedergaard, M. The Locus Coeruleus-Norepinephrine Network Optimizes Coupling of Cerebral Blood Volume with Oxygen Demand. *J Cereb Blood Flow Metab* **32**, 2135–2145 (2012).
52. Lecrux, C. & Hamel, E. Neuronal networks and mediators of cortical neurovascular coupling responses in normal and altered brain states. *Phil. Trans. R. Soc. B* **371**, 20150350 (2016).
53. Moore, C. I. & Cao, R. The Hemo-Neural Hypothesis: On The Role of Blood Flow in Information Processing. *Journal of Neurophysiology* **99**, 2035–2047 (2008).
54. Chow, B. W. *et al.* Caveolae in CNS arterioles mediate neurovascular coupling. *Nature* **579**, 106–110 (2020).

55. Snider, S. B. *et al.* Disruption of the ascending arousal network in acute traumatic disorders of consciousness. *Neurology* **93**, e1281–e1287 (2019).

56. Robertson, I. H. A noradrenergic theory of cognitive reserve: implications for Alzheimer's disease. *Neurobiology of Aging* **34**, 298–308 (2013).

57. Rommelfanger, K. S. & Weinshenker, D. Norepinephrine: The redheaded stepchild of Parkinson's disease. *Biochemical pharmacology* **74**, 177–190 (2007).

58. Bohnen, N. I. & Albin, R. L. The cholinergic system and Parkinson disease. *Behavioural Brain Research* **221**, 564–573 (2011).

59. Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S. & Eckert, M. A. In vivo mapping of the human locus coeruleus. *NeuroImage* **47**, 1261–1267 (2009).

60. Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* **25**, 1325–1335 (2005).

61. Li, R. *et al.* A pulse of transient fMRI increases in subcortical arousal systems during transitions in attention. 63.

62. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping* **15**, 1–25 (2002).