Selective modulation of cortical state during spatial attention

Tatiana A. Engel,1,2*† Nicholas A. Steinmetz,3* Marc A. Gieselmann,4 Alexander Thiele,4 Tirin Moore,2,3 Kwabena Boahen1

Neocortical activity is permeated with endogenously generated fluctuations, but how these dynamics affect goal-directed behavior remains a mystery. We found that ensemble neural activity in primate visual cortex spontaneously fluctuated between phases of vigorous (On) and faint (Off) activity. Spikes on 16 recorded channels were synchronously across cortical layers. These On-Off dynamics, reflecting global changes in cortical state, were also modulated at a local scale during selective attention. Moreover, the momentary phase of local ensemble activity predicted behavioral performance. Our results show that cortical state is controlled locally within a cortical map according to cognitive demands and reveal the impact of these local changes in cortical state on goaldirected behavior.

Endogenous fluctuations in neocortical spiking activity vary on a continuum between synchronized and desynchronized states, and the level of synchrony has been associated with the overall level of arousal (1, 2). During slow-wave sleep and anesthesia, ensemble neural activity exhibits slow synchronous transitions between periods of high activity and quiescence. In individual neurons, these transitions manifest as alternating Down (hyperpolarized) and Up (depolarized) phases of the membrane potential due to, respectively, an ebb and flow of synaptic activity (3–7). In awake animals, these slow synchronous transitions are less frequent, and thus, ensemble neural activity appears less synchronized than during anesthesia or slow-wave sleep (8–12). This relationship between arousal and cortical synchrony suggests that mechanisms controlling cortical state are brain-wide and unrelated to neural circuits involving the selective recruitment of local populations during goal-directed behavior. In particular, changes in cortical state should be orthogonal to the modulations of spiking activity observed locally within cortical maps during selective attention.

We discovered that spontaneous transitions between episodes of vigorous (On) and faint (Off) spiking occur synchronously across cortical layers in the visual cortex of behaving monkeys. We recorded ensemble spiking activity in area V4 of two rhesus macaques (G and B) with 16-channel linear array microelectrodes (Fig. 1C, left) arranged so that receptive fields (RFs) on all channels largely overlapped (Fig. 1A, fig. S1, and materials and methods 1). The On-Off transitions occurred synchronously throughout the cortical depth during fixation and in the absence of visual stimulation (Fig. 1, B and C). The On and Off episodes resembled the Up and Down phases commonly observed during anesthesia and slow-wave sleep (7–9) and were consistent with the large fluctuations in cortical membrane potentials recorded intracellularly in behaving monkeys (13).

To examine whether these On-Off fluctuations also occur during more demanding cognitive behaviors, we trained monkeys to perform a selective attention task. Monkeys were rewarded for detecting changes in a visual stimulus and indicating those changes with an antisaccade response (Fig. 1D). During each trial, a small central cue indicated the stimulus that was most likely to change orientation. The cued stimulus was thus the target of covert attention, whereas because of anticipation of antisaccadic response, the stimulus opposite to the cue was the target of overt attention (14). In spite of the difficulty of the task, monkeys performed well above chance, with 69 and 67% correct responses for monkeys G and B, respectively. While monkeys performed this task, we recorded from area V4 in 46 sessions (25 in monkey G and 21 in monkey B). As in the fixation task, we observed prominent On-Off transitions occurring synchronously across the cortical depth in both spontaneous and stimulus-driven activity, before and after the attention cue was presented, and evident in both single- and multunit activity (Fig. 1E, fig. S2, supplementary text 3.1). On episodes reliably followed stimulus onset on a majority of trials. However, subsequent On-Off transitions occurred irregularly within and across trials during the sustained response to the stable RF stimulus. Transitions were also irregular with respect to the attention cue’s onset.

To characterize the On-Off dynamics, we counted spikes in 10-ms time bins and used a Hidden Markov Model (HMM) as a statistically principled way to segment spike-count data into On and Off episodes (Fig. 2A and materials and methods 2.2) (15, 16). An HMM was fitted to 16-channel multunit activity, but all analyses based on the fitted HMM were performed on both single- and multunit activities, yielding highly consistent results. The HMM has a one-dimensional, latent variable representing an unobserved population state that switches between two phases, On and Off. Spikes on 16 recorded channels are assumed to be generated by inhomogeneous Poisson processes, with different mean rates during the On and Off phases. When an HMM is fitted to the spiking data, 34 parameters are estimated: firing rates in the On and Off phases for each of 16 channels and transition probabilities $p_{on}$ and $p_{off}$ for the entire ensemble (Fig. 2B). Using these parameters, we can then infer the most likely sequence of On and Off episodes that underlie the observed spike trains on a trial-by-trial basis (Fig. 2C). From visual inspection, On and Off episodes inferred by the HMM were closely aligned to the periods of vigorous and faint spiking. These On-Off transitions in spike rates were also phase-locked to low-frequency fluctuations in the local field potential (LFP) (fig. S3 and supplementary text 3.2). On average, the HMM captured about half of the maximal explainable variance in the data (fig. S4 and supplementary text 3.3). For most recording sessions (31 total, 67%), the two-phase HMM was the most parsimonious model among HMMs, with 1 or up to 8 possible phases (fig. S5). These 31 sessions were therefore used in subsequent analyses of On and Off episode durations.

For the remaining 15 (33%) sessions, a one-phase HMM was the most parsimonious model. Consistent with HMM assumptions, the durations of On and Off episodes were distributed exponentially, with the decay time constants $\tau_{on}$ and $\tau_{off}$ given by the average On and Off episode durations. For the remaining 15 (33%) sessions, the two-phase HMM was the most parsimonious model. Consistent with HMM assumptions, the durations of On and Off episodes were distributed exponentially, with the decay time constants $\tau_{on}$ and $\tau_{off}$ given by the average On and Off episode durations (Fig. 2D and fig. S6A). The averages across sessions of these average On and Off episode durations were, respectively, $\tau_{on} = 149 \pm 77$ ms and $\tau_{off} = 102 \pm 33$ ms for stimulus-driven activity and $\tau_{on} = 97 \pm 36$ ms and $\tau_{off} = 118 \pm 47$ ms for spontaneous activity (mean ± SD across 31 sessions) (fig. S8, B and C). We also analyzed laminar recordings from area V4 performed with a different type of linear electrode array, in two different behaving monkeys, and in a different laboratory. In this additional data set, On-Off transitions also occurred synchronously across cortical layers during spontaneous and stimulus-driven activity and were equally well described by the two-phase HMM (Fig. 2, E to G, fig. S7, and supplementary text 3.4).

To investigate the extent to which On-Off dynamics reflect arousal or selective attention, we analyzed data from the two monkeys performing the attention task. In rodents, cortical state dynamics closely covary with global arousal, as measured by pupil size, with dilation characterized by desynchronization of neural activity and constriction by an increase in low-frequency fluctuations (17, 18). Similarly, the pupil size was positively correlated with the average duration of On episodes on a trial-by-trial basis, during fixation and attention (fig. S8). Thus, the On-Off dynamics indeed reflected global changes of cortical state associated with arousal. Spatial attention, on the other hand, involves the selective...
recruitment of local neuronal populations encoding behaviorally relevant stimuli at one retinotopic location and the simultaneous suppression of populations encoding irrelevant stimuli in other retinotopic locations (19–21). We used three behavioral conditions to measure the local effects of both covert and overt attention on neuronal activity. In the covert attention condition, the cue directed the animal’s attention to the RF stimulus. In the overt attention condition, the cue directed attention to the stimulus opposite the RF and indicated that a saccade to the RF stimulus was likely to be required. In the control condition, the cue directed covert and overt attention to stimuli in directions orthogonal to the RF direction. The overall mean firing rate of V4 neurons was greater in covert and overt attention conditions relative to the control condition (Fig. 3A and fig. S9B), as has been previously reported (14). If these local effects of selective attention are indeed orthogonal to SCIENCE
Fig. 2. HMM of On-Off dynamics. (A) HMM schematic. In each time bin \( (t_1, t_2, t_3, \ldots) \), spike counts \( x_{1}, x_{2}, \ldots x_{16} \) on 16 channels are generated through inhomogeneous Poisson processes with mean rates \( r_{1}, r_{2}, \ldots r_{16} \) that are different between the On and Off phase. Transitions between unobserved On and Off phases are governed by probabilities \( p_{\text{on}} \) and \( p_{\text{off}} \). (B) Example HMM fit. Firing rates of 16 channels in the On and Off phase and transition probabilities (numbers above the curved arrows) estimated by the model. Error bars are 5th and 95th percentiles over 10 bootstrap samples. The HMM was fitted to multiunit activity within the time-window beginning 400 ms after the attention-cue onset and ending at the start of the blank period (Fig. 1D). (C) Example epoch of spiking activity segmented into On (green) and Off (pink) episodes by the HMM. (D) Distributions (black lines) of On (right) and Off (left) episode durations overlaid by exponential distributions (green and pink lines), with the decay time-constants set by HMM transition probabilities. (E to G) Same as (B) to (D), respectively, but for the additional laminar data set.

The On-Off dynamics were modulated by attention consistent with the third scenario. Firing rates during the On and Off phases were slightly, but significantly, enhanced during both types of attention (Wilcoxon signed rank test; covert, On-phase median modulation index (MI) = 0.008, \( P = 0.002 \), Off-phase median MI = 0.005, \( P = 0.046 \); overt, On-phase median MI = 0.014, \( P < 10^{-5} \), Off-phase median MI = 0.029, \( P < 10^{-5} \) (Fig. 3C and materials and methods 2.2.5). The average duration of Off episodes was significantly longer in the covert, but not the overt, attention condition compared with control conditions (Wilcoxon signed rank test; covert, median change in duration 4 ms, \( P = 0.004 \); overt, median change in duration -2 ms, \( P = 0.052 \) (Fig. 3D). However, the average duration of On episodes was significantly longer during both covert and overt attention as compared with controls (Wilcoxon signed rank test; covert, median change in duration 7 ms, \( P < 10^{-5} \); overt, median change in duration 13 ms, \( P < 10^{-5} \) (Fig. 3D). Correspondingly, the On-Off transition frequency was significantly lower during covert and overt attention compared with control conditions (Wilcoxon signed rank test; covert, median reduction in frequency 0.2 Hz, \( P = 0.001 \); overt, median reduction in frequency 0.2 Hz, \( P < 10^{-5} \) (fig. S9A).

In separate control analyses, we confirmed that the changes in On-Off dynamics were not an artifact of the attention-related increase in firing rates, or of the HMM’s assumption of discrete On-Off phases (figs. S9 and S10 and supplementary text 3.5 and 3.6). We also considered the influence of microsaccades. Unlike stimulus onset, which was reliably followed by On-episodes, only a small fraction of transitions was preceded by a microsaccade (fig. S11, A to D). Nonetheless, changes in frequency or direction of microsaccades could not account for observed increases in On-episode durations (fig. S11, E to G). Moreover, increases in On-episode durations were also observed on trials without microsaccades (fig. S11H and supplementary text 3.7).

Last, we asked whether On-Off dynamics, in addition to being modulated by attention, predicted behavioral performance. In our task, the probability of detecting a change was greater at
the cued location as compared with uncued locations (14). We investigated how the increase in detection probability at the cued location was related to On-Off dynamics (materials and methods 2.2.6). When the cued orientation change occurred in the RFs of recorded neurons, the probability of detecting that change was significantly greater when it occurred during an On-phase than during an Off-phase (median detection probability 64.8% during Off-phase, 78.3% during On-phase, difference 13.5%, P < 10\(^{-3}\), Wilcoxon signed rank test) (Fig. 4A). This difference in detection probability was evident beginning ~150 ms before the stimulus change (Fig. 4B), which is consistent with the average duration of On episodes. This effect was spatially selective; we found no difference in detection probability between On and Off phases when the cued change occurred outside the RFs of recorded neurons (median detection probability difference 0.4%, P = 0.943, Wilcoxon signed rank test) (Fig. 4A).

Spontaneous On-Off transitions, occurring synchronously throughout the cortical depth, were modulated locally within a cortical map during selective attention and predicted behavioral performance. These On-Off dynamics represent a substantial source of correlated variability classically observed in cortical responses (22), and many features of this correlated variability,

---

**Fig. 3. Effects of covert and overt attention on dynamics of On-Off transitions. (A)** Average firing rates of an example single-unit for covert (blue), overt (orange), and control trials (gray). Shaded area represents ± SEM. (B) Three hypothetical ways On-Off dynamics can interact with attention-related increase in average firing rate (blue line, attention; black dashed line, control). (C) Distribution of attention modulation index of single-unit firing rates in the On (right) and Off (left) phases during covert (blue) and overt attention (orange). (D) Distribution across recordings of the difference in average durations of On (right) and Off (left) episodes between covert attention and control (blue), and overt attention and control (orange) conditions. In (C) and (D), triangles indicate medians of the distributions; P values are for Wilcoxon signed rank test. (E) Cartoon of two sites in area V4 corresponding to attended (blue ellipse) and unattended (black ellipse) retinotopic locations (top) with corresponding On-Off dynamics (bottom).

---

**Fig. 4. On-Off population state predicts behavioral performance. (A)** Probability to detect an orientation change that occurred during the Off phase (x axis) versus a change that occurred during the On phase (y axis). Each point represents one recording session (blue, covert attention; gray, control condition). (Inset) Average difference between the On and Off detection probability. Error bars represent SD across recordings; asterisks indicate P < 10\(^{-3}\) for Wilcoxon signed rank test. (B)** Time course of the detection probability. At each time bin, the detection probability was calculated separately for trials on which the instantaneous population state was in the On phase (green line) and on which it was in the Off phase (pink line) at the corresponding time bin. Gray shading indicates significant difference in detection probability (two-sided paired t test, P < 0.05 corrected for multiple comparisons across all time bins).
such as spike-count correlations (23), can be understood as arising from the On-Off dynamics (fig. S12 and supplementary text 3.8.4). Recent models parsimoniously attribute changes in spike-count correlations during attention to fluctuations in shared modulatory signals (32), with smaller spike-count correlations accounted for by reduced fluctuations in these modulatory signals (32). The On-Off dynamics observed here could underlie the apparent trial-to-trial fluctuations in shared modulatory signals (32, 33) but can account for within-trial fluctuations as well (fig. S12 and supplementary text 3.8.5).

What mechanisms underlie the spatially and temporally precise control of cortical state during selective attention? Our results suggest that global temporally precise control of cortical state during attention to fluctuations in a spatially targeted manner (34, 35) also mediate the effects of selective attention (36) and influence circuits that control selective attention (37). On the other hand, cortico-cortical inputs appear to influence state changes in a spatially targeted manner (38, 39). Because diffuse neuromodulatory signals are interspersed with topographically precise projections throughout the cortex, local modulation of cortical state is likely to be widespread, extending to modalities beyond vision and serving many cognitive functions.

REFERENCES AND NOTES


PAIN RESEARCH

Gliogenic LTP spreads widely in nociceptive pathways

M. T. Kronschläger,∗ R. Drdla-Schutting,∗ M. Gassner, S. D. Honsek, H. L. Teuchmann, J. Sandkühler†

Learning and memory formation involve long-term potentiation (LTP) of synaptic strength. A fundamental feature of LTP induction in the brain is the need for coincident pre- and postsynaptic activity. This restricts LTP expression to activated synapses only (homosynaptic LTP) and leads to its input specificity. In the spinal cord, we discovered a fundamentally different form of LTP that is induced by glial cell activation and mediated by diffusible, extracellular messengers, including c-serine and tumor necrosis factor (TNF), and that travel long distances via the cerebrospinal fluid, thereby affecting susceptible synapses at remote sites. The properties of this gliogenic LTP resolve unexplained findings of memory traces in nociceptive pathways and may underlie forms of widespread pain hypersensitivity.

Activity-dependent, homosynaptic long-term potentiation (LTP) (1) at synapses in nociceptive pathways contributes to pain amplification (hyperalgesia) at the site of an injury or inflammation (2–5). Homosynaptic LTP can, however, not account for pain amplification at areas surrounding (secondary hyperalgesia) or remote from (widespread hyperalgesia) an injury. It also fails to explain pain hyperalgesia that is induced independently of neuronal activity in primary afferents—e.g., by the application of or the withdrawal from opioids (opioid-induced hyperalgesia) (6). Glial cells are believed to contribute to these forms of hyperalgesia and to LTP in nociceptive pathways (7–10). Induction of homosynaptic LTP can be accompanied by LTP in adjacent, inactive synapses converging onto the same neuron, especially early in development. The respective molecular signals for this heterosynaptic form of LTP are thought to be confined within the cytoplasm of the activated neuron, spreading tens of micrometers only (11). We have now tested the hypothesis that, in contrast to current beliefs, activation of glial cells is causative for the induction of LTP at spinal C-fiber synapses and that this gliogenic LTP constitutes a common denominator of homosynaptic and heterosynaptic LTP in the spinal cord. Our previous study revealed that selective activation of spinal microglia by fractalkine induces...
Selective modulation of cortical state during spatial attention
Tatiana A. Engel, Nicholas A. Steinmetz, Marc A. Gieselmann, Alexander Thiele, Tirin Moore and Kwabena Boahen (December 1, 2016)
Science 354 (6316), 1140-1144. [doi: 10.1126/science.aag1420]

Editor's Summary

Attention changes local brain activity

There is a well-known correlation between arousal and neuronal activity in the brain. However, it is unclear how these general effects are reflected on a local scale. Engel et al. recorded from higher visual areas in behaving monkeys and discovered a new principle of cortical state fluctuations. A special type of electrodes revealed that the state changes affected neuronal excitability across all layers of the neocortex. When the animals attended to a stimulus, the vigorous spiking states became longer and the faint spiking states became shorter. These states correlated with fluctuations in the local field potential. A sophisticated computational model of the state changes fitted a two-state model of neuronal responsiveness.

Science, this issue p. 1140