Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli

L. Mazzucato[1,2], G. La Camera[1,3,*], A. Fontanini[1,3,*]

[1] *Department of Neurobiology and Behavior, State University of New York at Stony Brook, Stony Brook, NY 11794*

[2] *Departments of Biology and Mathematics and Institute of Neuroscience, University of Oregon, Eugene, OR 97403*

[3] *Graduate Program in Neuroscience, State University of New York at Stony Brook, Stony Brook, NY 11794*

* Co-senior authors

Running Title: Metastable activity mediates expectation

Page Number: 14
Number of Figures: 4
Word count Abstract: 169
Word count Introduction: 456
Word count Discussion: 1248

Contact:
Alfredo Fontanini[a] and Giancarlo La Camera[b]
[a] Department of Neurobiology & Behavior, Life Sciences Building 516
State University of New York at Stony Brook, Stony Brook, NY 11794
alfredo.fontanini@stonybrook.edu
[b] Department of Neurobiology & Behavior, Life Sciences Building 513
State University of New York at Stony Brook, Stony Brook, NY 11794
giancarlo.lacamera@stonybrook.edu

Abstract

Sensory stimuli can be recognized more rapidly when they are expected. This phenomenon depends on expectation affecting the cortical processing of sensory information. However, virtually nothing is known on the mechanisms responsible for the effects of expectation on sensory networks. Here, we report a novel computational mechanism underlying the expectation-dependent acceleration of coding observed in the gustatory cortex (GC) of alert rats. We use a recurrent spiking network model with a clustered architecture capturing essential features of cortical activity, including the metastable activity observed in GC before and after gustatory stimulation. Relying both on network theory and computer simulations, we propose that expectation exerts its function by modulating the intrinsically generated dynamics preceding taste delivery. Our model, whose predictions are confirmed in the experimental data, demonstrates how the modulation of intrinsic metastable activity can shape sensory coding and mediate cognitive processes such as the expectation of relevant events. Altogether, these results provide a biologically plausible theory of expectation and ascribe a new functional role to intrinsically generated, metastable activity.

Introduction

Expectation exerts a strong influence on sensory processing. It improves stimulus detection, enhances discrimination between multiple stimuli and biases perception towards an anticipated stimulus[1-3]. These effects, demonstrated experimentally for various sensory modalities and in different species[2,4-7], can be attributed to changes in sensory processing occurring in primary sensory cortices. However, despite decades of investigations, little is known regarding how expectation shapes the cortical processing of sensory information.

While different forms of expectation likely rely on a variety of neural mechanisms, modulation of pre-stimulus activity is believed to be a common underlying feature[8-10]. Here, we investigate the link between pre-stimulus activity and the phenomenon of general expectation in a recent set of experiments performed in gustatory cortex (GC) of alert rats[6]. In those experiments, rats were trained to expect the intraoral delivery of one of four possible tastants following an anticipatory cue. The use of a single cue allowed the animal to predict the availability of gustatory stimuli, without forming expectations on which specific taste was being delivered. Cues predicting the general availability of taste modulated firing rates of GC neurons. Tastants delivered after the cue were encoded more rapidly than uncued tastants, and this improvement was phenomenologically attributed to the activity evoked by the preparatory cue. However,

the precise computational mechanism linking faster coding of taste and cue responses remains unknown.

Here we propose a mechanism whereby an anticipatory cue modulates the timescale of temporal dynamics in a recurrent population model of spiking neurons. In the model proposed here, neurons are organized in strongly connected clusters and produce sequences of metastable states similar to those observed during both pre-stimulus and evoked activity periods[11-17]. A metastable state is a vector of firing rates across simultaneously recorded neurons that can last for several hundred milliseconds before giving way to the next state in a sequence. The ubiquitous presence of state sequences in many cortical areas and behavioral contexts[18-24] has raised the issue of their role in sensory and cognitive processing. Here, we elucidate the central role played by pre-stimulus metastable states in processing forthcoming stimuli, and show how cue-induced modulations of state sequences drive anticipatory coding. Specifically, we show that an anticipatory cue affects sensory coding by decreasing the duration of metastable states and accelerating the pace of state sequences. This phenomenon, which results from a reduction in the energy barriers separating the metastable states, accelerates the onset of specific states coding for the presented stimulus, thus mediating the effects of general expectation. Analysis of the experimental data confirmed the predictions of the model.

Altogether, our results unveil a novel computational mechanism accounting for the acceleration of coding induced by expectation, and demonstrate for the first time the functional significance of modulating pre-stimulus, ongoing dynamics in cortical circuits.

Results

Anticipatory cue accelerates stimulus coding in a clustered population of neurons

To uncover the computational mechanism linking cue-evoked activity with coding speed, we modeled the gustatory cortex (GC) as a population of recurrently connected excitatory and inhibitory spiking neurons. Excitatory neurons are arranged in clusters[12,25,26] (Fig. 1a). Recurrent synaptic weights between neurons in the same cluster were potentiated compared to neurons in different clusters. Connections between excitatory and inhibitory neurons were random and homogeneous (i.e., non-clustered), as were connections between inhibitory neurons. Sensory inputs were modeled as depolarizing currents injected in randomly selected neurons. We used four sets of simulated stimuli, wired to produce gustatory responses like those observed in the experiments to sucrose, sodium chloride, citric acid, and quinine (see Supplementary Methods material for details on selectivity). In addition, we endowed the model with a set of anticipatory inputs designed to produce cue-responses analogous to those seen

experimentally in the case of general expectation. To simulate general expectation, we connected anticipatory inputs with random neuronal targets in the network. The peak value of the cue-induced current for each neuron was sampled from a normal distribution with zero mean and fixed variance (see Supplementary Results). This choice reflected the large heterogeneity of cue responses observed in the empirical data, where excited and inhibited neural responses occurred in similar proportions[10]. Fig 1b shows two representative cue-responsive neurons in the model: one inhibited by the cue and one excited by the cue. Given these conditions, we simulated the experimental paradigm adopted in awake-behaving rats to demonstrate the effects of general expectation[6,10]. In the original experiment, rats were trained to self-administer into an intra-oral cannula one of four possible tastants following an anticipatory cue. At random trials and time during the inter-trial interval, tastants were unexpectedly delivered in the absence of a cue. To match this experiment, the simulated paradigm interleaves two conditions: in expected trials, a stimulus (out of 4) is delivered at $t=0$ after an anticipatory cue (the same for all stimuli) delivered at $t=-0.5$s (Fig. 1b); in unexpected trials the same stimuli are presented in the absence of the cue. We then tested whether cue presentation affected stimulus coding. A multi-class classifier (see Online Methods) was used to assess the information about the stimuli encoded in the neural activity. Stimulus identity was encoded well in both conditions, reaching perfect accuracy after a few hundred milliseconds (Fig. 1c). However, comparing the time course of the decoding accuracy between conditions, we found that the increase in decoding accuracy was significantly faster in expected than in unexpected trials (Fig. 1c, pink and blue curves represent expected and unexpected conditions, respectively). Indeed, the onset time of a significant decoding occurred earlier in the expected vs. the unexpected condition (decoding latency was 0.13 ± 0.01 s [mean$\pm$s.e.m.] for expected compared to 0.21 ± 0.02 s for unexpected, across 20 independent sessions; p=0.002, signed-rank=14, d.o.f.=39; inset in Fig. 1c). Thus, in the model network, the interaction of cue response and activity evoked by the stimuli results in faster encoding of the stimuli themselves, mediating the expectation effect.

The clustered architecture of the model was chosen for its ability to reproduce fundamental features of the activity observed in GC[11,27]. To clarify the role of neural clusters in mediating expectation, we simulated the same experiments in a homogeneous network (i.e., without clusters) operating in the balanced asynchronous regime[25,26] (Fig. 1d, intra- and inter-cluster weights were set equal, all other network parameters and inputs were the same as for the clustered network). Even though single neurons' responses to the anticipatory cue were comparable to the ones observed in the clustered network (see Supplementary Results), stimulus encoding was not affected by cue presentation (Fig. 1f, pink and blue curves represent expected and unexpected conditions, respectively). In particular, the onset of a significant decoding was similar in

the two conditions (latency of significant decoding was 0.17 ± 0.01 s for expected and 0.16 ± 0.01 s for unexpected tastes averaged across 20 sessions; p=0.31, signed-rank=131, d.o.f.=39; inset in Fig. 1f).

The anticipatory activity observed in the clustered network was robust to variations in key parameters related to the sensory and anticipatory inputs, as well as network connectivity and architecture (Fig. S1 and Supplementary Results). Furthermore, acceleration of coding depended on the patterns of connectivity of anticipatory inputs, specifically on the fact that it increased the spatial spread in the afferent currents (Fig S1c). In a model where the cue recruited the recurrent inhibition (i.e., increasing the input currents to the inhibitory population), stimulus coding was decelerated (Fig S2), suggesting a potential mechanism mediating the effect of distractors.

Overall, these results demonstrate that a clustered network of spiking neurons can successfully reproduce the acceleration of sensory coding induced by expectation and that removing clustering impairs this function.

Anticipatory cue speeds up the network's dynamics

Having established that a clustered architecture mediates the effects of expectation on coding, we investigated the underlying mechanism.

Clustered networks spontaneously generate highly structured activity characterized by coordinated patterns of ensemble firing. This activity results from the network hopping between metastable states in which different combinations of clusters are simultaneously activated[11,14,15]. To understand how anticipatory inputs affected network dynamics, we analyzed the effects of cue presentation for a prolonged period of 5 seconds in the absence of stimuli (Fig. 2). Activating anticipatory inputs led to changes in network dynamics, with clusters turning on and off more frequently (Fig. 2a). We quantified this effect by showing that a cue-induced increase in input spatial spread (σ) led to a shortened cluster activation lifetime (left panel in Fig. 2b; Kruskal-Wallis one-way ANOVA: $p<10^{-17}$, $\chi^2(5)=91.2$), and a shorter cluster inter-activation interval (i.e., quiescent intervals between consecutive activations of the same cluster, right panel in Fig. 2b, kruskal-wallis one-way ANOVA: $p<10^{-18}$, $\chi^2(5)=98.6$).

Previous work demonstrated that metastable states of co-activated clusters result from attractor dynamics[11,14,15]. Hence, the shortening of cluster activations and inter-activation intervals observed in the model could be due to modifications in the network's attractor dynamics. To test this hypothesis, we performed a mean field theory analysis[28,29] of a simplified network with two clusters, comprising a reduced repertoire of configurations (or 'attractors'), which include two configurations in which either cluster is active, and the other inactive. For instance, consider the transition from a configuration with cluster

1 active while cluster 2 is inactive, to the opposite configuration with cluster 2 active and cluster 1 inactive. These configurations are shown in Fig. 2c as potential wells 'A' and 'B', respectively. Due to finite size effects, configurations A and B are metastable, and Kramer's theory predicts the network will switch between the two configurations with a finite transition probability $P(A \leftrightarrow B)$, inversely proportional to the height Δ of the barrier separating them[15,30]. The presence of an anticipatory cue decreases the height of the barrier separating the two potential wells (Fig. 2c). This effect is proportional to the cue-induced spread in the afferent currents (Fig. 2c, lighter shades represent larger spatial spread σ).

This analysis shows that the anticipatory cue increases the spontaneous transition rates between the network's metastable configurations by reducing the energy barrier necessary to hop among configurations. In the following we uncover an important consequence of this phenomenon for sensory processing.

Anticipatory cue induces faster onset of taste-coding states

The cue-induced modulation of attractor dynamics led us to formulate a hypothesis for the mechanism underlying the acceleration of coding. The activation of anticipatory inputs prior to sensory stimulation may allow the network to enter more easily configurations encoding stimuli while exiting more easily non-coding configurations. Fig. 3a shows population rasters in response to the same stimulus presented in the absence of a cue or after a cue. The red spikes represent activity in taste-selective clusters and show a faster activation latency in response to the stimulus preceded by the cue compared to the uncued condition. A systematic analysis revealed that in the cued condition, the clusters activated by the subsequent stimulus had a significantly faster activation latency than in the uncued condition (Fig. 3b, 0.22 ± 0.01 s (mean±s.e.m.) during cued compared to 0.32 ± 0.01 s for uncued stimuli; $p<10^{-5}$, rank sum test R(39)=232).

We elucidated this effect using mean field theory. In the simplified two-cluster network of Fig. 3c (the same network as in Fig. 2c), the configuration where the taste-selective cluster is active ("coding state") and the nonselective cluster is active ("non-coding state") have initially the same potential energy, in the absence of stimulation (local minima of the black line in Fig. 3c), separated by an energy barrier whose height is reduced by the anticipatory cue (dashed vs. full line). When the taste stimulus is presented, it activates the stimulus-selective cluster, so that the coding state will now sit in a deeper well (lighter lines) compared to non-coding state. Stronger stimuli (lighter shades in Fig. 3c) increase the difference between the wells' depths breaking their initial symmetry, so that now a transition from the non-coding to the coding state is more likely than a transition from the coding to the non-coding state[30]. The faster transition

rate into coding configurations translates into *faster* coding, on average, of the stimuli encoded by those states.

We tested the model prediction on the data from Samuelsen et al. (Fig 4)[6]. To compare the data to the model simulations, we randomly sampled ensembles of model neurons whose sizes matched the empirical dataset ones. Since we only have access to a subset of neurons in the experiments, rather than the full network configuration, we segmented the ensemble activity for both model and data in sequences of metastable states via a Hidden Markov Model (HMM) analysis (see Methods). Previous work demonstrated that HMM states, i.e. patterns of coordinated ensemble firing activity, can be treated as reduced observations of metastable network configurations[11]. Indeed, activation of taste-coding configurations for a particular stimulus results in HMM states containing information about that stimulus (i.e., taste-coding HMM states). If the hypothesis originating from the model is correct, transitions from non-coding HMM states to taste-coding HMM states should be faster in the presence of the cue compared to uncued trials. We indeed found faster transitions to HMM coding states in cued trials for both model and data (Fig. 4a and 4c, respectively; color-coded horizontal bars overlay coding states). The latency of coding states was significantly faster during cued compared to uncued trials in both the model (Fig. 4b, mean latency of the first coding state was 0.32 ± 0.02 s for expected vs 0.38 ± 0.01 s for unexpected trials; rank sum test $R(39)=319$, $p=0.014$) and the empirical data (Fig. 4d: 0.46 ± 0.02 s for expected vs 0.56 ± 0.03 s for unexpected trials; rank sum test $R(37)=385$, $p=0.026$).

Altogether, these results demonstrate that anticipatory inputs speed up sensory coding by reducing the energy barriers separating metastable states.

Discussion

Expectations modulate perception and sensory processing. Expected stimuli are recognized more accurately and rapidly than unexpected ones. In the gustatory cortex, acceleration of taste coding has been related to changes in firing activity evoked by cues predicting the general availability of tastants[6]. However, the computational mechanisms linking pre-stimulus activity with changes in the latency of sensory coding are still unknown. Here we propose a novel mechanism that explains the effects of expectation through the modulation of the dynamics intrinsically generated by the cortex. Our results represent also the first theoretical evidence of a functional role for the erratic, intrinsically generated activity that is ubiquitously observed in cortical circuits[11,18-21,24,31-34].

The proposed mechanism requires a recurrent spiking network where excitatory neurons are arranged in clusters, which has been demonstrated to capture essential features of the dynamics of neural activity in sensory circuits[11,16]. In such a model,

network activity during both spontaneous and stimulus-evoked periods unfolds through state sequences, each state representing a metastable network attractor. In response to an anticipatory cue, the pace of state sequences speeds up, accelerated by a higher transition probability among states. The latter is caused by lowering the potential barrier separating metastable states in the attractor landscape. This anticipates the offset of states not conveying taste information and the onset of states containing the most information about the delivered stimulus ('coding states'), causing the faster decoding observed by Samuelsen *et al*[6] (see Fig. 1c).

Notably, this novel mechanism for anticipation is unrelated to increases in network excitability which would lead to unidirectional changes in taste-evoked firing rates. It relies instead on an increase in the spatial spread of the afferent currents to the sensory network brought about by the anticipatory cue. This increase in the input's variance is observed experimentally after training[10], and is therefore the consequence of having learned the anticipatory meaning of the cue. The acceleration of the dynamics of state sequences predicted by the model was also confirmed in the data from ensembles of simultaneously recorded neurons in awake-behaving rats.

These results show for the first time a precise and explanatory link between the intrinsic dynamics of neural activity in a sensory circuit and a specific cognitive process, that of general expectation[6].

Clustered connectivity and metastable states

A key feature of our model is the clustered architecture of the excitatory population. Removing excitatory clusters eliminates the cue-induced anticipatory effect (Fig. 1d-f). The choice of such a network architecture reflects the existence of assemblies of functionally correlated neurons in GC and other cortical areas, and is further justified by growing evidence from electrophysiological and imaging experiments suggesting that pyramidal neurons in various cortical circuits may be arranged in clusters endowed with strong local connectivity[35,36].

Theoretical work in recurrent networks shows that a clustered architecture can produce stable patterns of population activity called attractors[12]. Noise (either externally[14,37] or internally generated[11,15]) may destabilize those states, driving the emergence of temporal dynamics based on the progression through metastable states. Network models with clustered architecture provide a parsimonious explanation for the state sequences that have been observed ubiquitously in alert mammalian cortex, during both task engagement[17,18,38,39] and inter-trial periods.[11,32,33] In addition, this type of models accounts for various physiological observations such as stimulus-induced reduction of trial-to-trial variability[11,14,15,40], neural dimensionality[27], and firing rate multistability[11] (see also[41,42]).

The results presented here demonstrate another fundamental function performed by clustered networks: the modulation of coding latency. Changes in the depth of attractor wells, induced by the anticipatory cue (which in turn may depend on the activation of top-down and neuromodulatory afferents[6,43]), can accelerate or slow down network dynamics. The acceleration resulting from shallower wells leads to a reshaping of ongoing activity and to a quicker recruitment of states capable of coding sensory information.

Functional role of heterogeneity in cue responses

As stated in the previous section, the presence of clusters is a necessary ingredient to obtain a faster latency of coding. Here we discuss the second necessary ingredient, i.e., the presence of heterogeneous neural responses to the anticipatory cue (Fig. 1b).

Responses to anticipatory cues have been extensively studied in cortical and subcortical areas in alert rodents[6,10,44,45]. Cues evoke heterogeneous patterns of activity, either exciting or inhibiting single neurons. The proportion of cue responses and their heterogeneity develops with associative learning,[10,45] suggesting a fundamental function of these patterns.

However, the presence of both excited and inhibited cue responses poses a challenge to simple models of neuromodulation. The presence of cue-evoked suppression of firing[10] suggests that cues do not improve coding by simply increasing the excitability of cortical areas. Additional mechanisms and complex patterns of connectivity may be required to explain the suppression effects induced by the cue. Here we provide a parsimonious explanation of how heterogeneous responses can improve coding without postulating any specific pattern of connectivity other than i) random projections from thalamic and anticipatory afferents and ii) the clustered organization of the intra-cortical circuitry. Notice that the latter contains wide distributions of synaptic weights and can be understood as the consequence of Hebb-like re-organization of the circuitry during training[46,47].

By incorporating excitated and inhibited cue responses in our model, cues-induced modulations do not affect the mean input to the network, but only change its spatial spread across neurons. As a result, the anticipatory cue leaves average firing rates unchanged in the clustered network (Fig. S3), and only modulates the network temporal dynamics. Our model thus provides a mechanism whereby increasing the spatial variance of top-down inputs has, paradoxically, a beneficial effect on sensory coding.

Cortical timescales, state transitions, and cognitive function

In populations of spiking neurons, a clustered architecture can generate reverberating activity and sequences of metastable states. Transitions from state to state can be typically caused by external inputs[11,15]. For instance, in frontal cortices, sequences of states are related to specific epochs within a task, with transitions evoked by behavioral events[18,19,22]. In sensory cortex, progressions through state sequences can be triggered by sensory stimuli and reflect the dynamics of sensory processing[23,38]. State sequences with a slow timescale have been observed also in the absence of any external stimulation, promoted by intrinsic fluctuations in neural activity[11,34]. However, the potential functional role, if any, of this type of ongoing activity has remained unexplored.

Recent work has started to uncover the link between ensemble dynamics and sensory and cognitive processes. State transitions in various cortical areas have been linked to decision making[37,48], choice representation[22], rule-switching behavior[24], and the level of task difficulty[23]. However, no theoretical or mechanistic explanations have been given for these phenomena.

Here we provide a mechanistic link between state sequences and expectation, by showing that intrinsically generated sequences can be accelerated, or slowed down, thus affecting sensory coding. Moreover, we show that the interaction between external stimuli and intrinsic dynamics does not result in the simple triggering of state transitions, but rather in the modulation of the intrinsic transition probabilities. The modulation of intrinsic activity can dial the duration of states, producing either shorter or longer timescales. A shorter timescale leads to faster state sequences and coding anticipation after stimulus presentation (Fig. 1 and 4). Other external perturbations may induce different effects: for example, recruiting the network's inhibitory population slows down the timescale, leading to a slower coding (Fig. S2).

The interplay between intrinsic dynamics and anticipatory influences presented here is a novel mechanism for generating diverse timescales, and has rich computational consequences. We demonstrated its function in increasing coding speed, but its role in mediating cognition is likely to be broader and calls for further explorations.

Author Contributions

LM, GLC, and AF designed the project and wrote the manuscript; LM performed model simulations and theoretical analyses; LM analyzed the empirical data.

Acknowledgements

This work was supported by a National Institute of Deafness and Other Communication Disorders Grant K25-DC013557 (LM), by the Swartz Foundation

Award 66438 (LM), by National Institute of Deafness and Other Communication Disorders Grants NIDCD R01DC012543 and R01DC015234 (AF), and partly by a National Science Foundation Grant IIS-1161852 (GLC). The authors would like to thank Drs. S. Fusi, A. Maffei, G. Mongillo, and C. van Vreeswijk for useful discussions.

Competing Financial Interests

The authors declare no competing financial interests.

References

- 1 Gilbert, C. D. & Sigman, M. Brain states: top-down influences in sensory processing. *Neuron* **54**, 677-696 (2007).
- 2 Jaramillo, S. & Zador, A. M. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature neuroscience* **14**, 246-251, doi:10.1038/nn.2688 (2011).
- 3 Engel, A. K., Fries, P. & Singer, W. Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience* **2**, 704-716 (2001).
- 4 Nobre, A. C., Correa, A. & Coull, J. T. The hazards of time. *Current opinion in neurobiology* **17**, 465-470 (2007).
- 5 Doherty, J. R., Rao, A., Mesulam, M. M. & Nobre, A. C. Synergistic effect of combined temporal and spatial expectations on visual attention. *Journal of Neuroscience* **25**, 8259-8266 (2005).
- 6 Samuelsen, C. L., Gardner, M. P. & Fontanini, A. Effects of cue-triggered expectation on cortical processing of taste. *Neuron* **74**, 410-422, doi:10.1016/j.neuron.2012.02.031 (2012).
- 7 Niwa, M., Johnson, J. S., O'Connor, K. N. & Sutter, M. L. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *Journal of Neuroscience* **32**, 9323-9334 (2012).
- 8 Yoshida, T. & Katz, D. B. Control of prestimulus activity related to improved sensory coding within a discrimination task. *J Neurosci* **31**, 4101-4112, doi:10.1523/JNEUROSCI.4380-10.2011 (2011).
- 9 Gardner, M. P. & Fontanini, A. Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 13000-13017, doi:10.1523/jneurosci.1820-14.2014 (2014).
- 10 Vincis, R. & Fontanini, A. Associative learning changes cross-modal representations in the gustatory cortex. *eLife* **5**, doi:10.7554/eLife.16420 (2016).
- 11 Mazzucato, L., Fontanini, A. & La Camera, G. Dynamics of Multistable States during Ongoing and Evoked Cortical Activity. *J Neurosci* **35**, 8214-8231, doi:10.1523/JNEUROSCI.4819-14.2015 (2015).

- 12 Amit, D. J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)* **7**, 237-252 (1997).
- 13 Renart, A., Moreno-Bote, R., Wang, X. J. & Parga, N. Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Comput* **19**, 1-46, doi:10.1162/neco.2007.19.1.1 (2007).
- 14 Deco, G. & Hugues, E. Neural network mechanisms underlying stimulus driven variability reduction. *PLoS computational biology* **8**, e1002395, doi:10.1371/journal.pcbi.1002395 (2012).
- 15 Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience* **15**, 1498-1505, doi:10.1038/nn.3220 (2012).
- 16 Mazzucato, L., Fontanini, A. & La Camera, G. Stimuli reduce the dimensionality of cortical activity. *Front Sys Neurosci* **10**, 11 (2016).
- 17 Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62-68, doi:10.1038/nature10918 (2012).
- 18 Abeles, M. *et al.* Cortical activity flips among quasi-stationary states. *Proc Natl Acad Sci USA* **92**, 8616-8620 (1995).
- 19 Seidemann, E., Meilijson, I., Abeles, M., Bergman, H. & Vaadia, E. Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **16**, 752-768 (1996).
- 20 Arieli, A., Shoham, D., Hildesheim, R. & Grinvald, A. Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *Journal of neurophysiology* **73**, 2072-2093 (1995).
- 21 Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science (New York, N.Y.)* **273**, 1868-1871 (1996).
- 22 Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nature neuroscience* **19**, 973-980, doi:10.1038/nn.4320 (2016).
- 23 Ponce-Alvarez, A., Nacher, V., Luna, R., Riehle, A. & Romo, R. Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32**, 11956-11969, doi:10.1523/jneurosci.6176-11.2012 (2012).
- 24 Durstewitz, D., Vittoz, N. M., Floresco, S. B. & Seamans, J. K. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* **66**, 438-448, doi:10.1016/j.neuron.2010.03.029 (2010).
- 25 van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science (New York, N.Y.)* **274**, 1724-1726 (1996).
- 26 Renart, A. *et al.* The asynchronous state in cortical circuits. *Science (New York, N.Y.)* **327**, 587-590, doi:10.1126/science.1179850 (2010).

- 27 Mazzucato, L., Fontanini, A. & La Camera, G. Stimuli Reduce the Dimensionality of Cortical Activity. *Front Syst Neurosci* **10**, 11, doi:10.3389/fnsys.2016.00011 (2016).
- 28 Mascaro, M. & Amit, D. J. Effective neural response function for collective population states. *Network (Bristol, England)* **10**, 351-373 (1999).
- 29 Mattia, M. *et al.* Heterogeneous attractor cell assemblies for motor planning in premotor cortex. *Journal of Neuroscience* **33**, 11155-11168 (2013).
- 30 Hänggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Reviews of modern physics* **62**, 251 (1990).
- 31 Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954-956, doi:10.1038/nature02078 (2003).
- 32 Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322-1327, doi:10.1126/science.1159775 (2008).
- 33 Euston, D. R., Tatsuno, M. & McNaughton, B. L. Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science* **318**, 1147-1150, doi:10.1126/science.1148979 (2007).
- 34 Luczak, A., Bartho, P. & Harris, K. D. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**, 413-425, doi:10.1016/j.neuron.2009.03.014 (2009).
- 35 Kiani, R. *et al.* Natural Grouping of Neural Responses Reveals Spatially Segregated Clusters in Prearcuate Cortex. *Neuron*, doi:10.1016/j.neuron.2015.02.014 (2015).
- 36 Miller, J. E., Ayzenshtat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4053-4061, doi:10.1073/pnas.1406077111 (2014).
- 37 Miller, P. & Katz, D. B. Stochastic transitions between neural states in taste processing and decision-making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **30**, 2559-2570, doi:10.1523/jneurosci.3047-09.2010 (2010).
- 38 Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P. & Katz, D. B. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18772-18777, doi:0705546104 [pii]
- 10.1073/pnas.0705546104 (2007).
- 39 Runyan, C. A., Piasini, E., Panzeri, S. & Harvey, C. D. Distinct timescales of population coding across cortex. *Nature*, doi:10.1038/nature23020 (2017).
- 40 Churchland, M. M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience* **13**, 369-378, doi:nn.2501 [pii]
- 10.1038/nn.2501 (2010).
- 41 Murray, J. M. & Escola, G. S. Learning multiple variable-speed sequences in striatum via cortical tutoring. *eLife* **6**, doi:10.7554/eLife.26084 (2017).
- 42 Stern, M., Bolding, K., Abbott, L. & Franks, K. A Transformation From Latency To Ensemble Coding In A Model Of Piriform Cortex. *bioRxiv*, doi:10.1101/118364 (2017).

- 43 Linster, C. & Fontanini, A. Functional neuromodulation of chemosensation in vertebrates. *Current opinion in neurobiology* **29**, 82-87, doi:10.1016/j.conb.2014.05.010 (2014).
- 44 Liu, H. & Fontanini, A. State Dependency of Chemosensory Coding in the Gustatory Thalamus (VPMpc) of Alert Rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**, 15479-15491, doi:10.1523/jneurosci.0839-15.2015 (2015).
- 45 Grewe, B. F. *et al.* Neural ensemble dynamics underlying a long-term associative memory. *Nature* **543**, 670-675, doi:10.1038/nature21682 (2017).
- 46 Zenke, F., Agnes, E. J. & Gerstner, W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nat Commun* **6**, 6922, doi:10.1038/ncomms7922 (2015).
- 47 Litwin-Kumar, A. & Doiron, B. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nat Commun* **5**, 5319, doi:10.1038/ncomms6319 (2014).
- 48 Engel, T. A. *et al.* Selective modulation of cortical state during spatial attention. *Science (New York, N.Y.)* **354**, 1140-1144 (2016).

Methods

[2702 words]

Behavioral training and electrophysiology. Adult female Long–Evans rats were used in the experiment.[1] Movable bundles of 16 microwires attached to a "mini-microdrive" were implanted bilaterally in the gustatory cortex and intraoral cannulae (IOC) were inserted bilaterally and cemented. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Stony Brook University and complied with university, state, and federal regulations on the care and use of laboratory animals (for more details, see Supplementary Methods and Ref. [1]). After postsurgical recovery, rats were trained to self-administer fluids through IOCs by pressing a lever under head-restraint within 3s presentation of an auditory cue ('expected trials'; a 75 dB pure tone at a frequency of 5 kHz). The interval at which lever pressing delivered water was gradually increased to 40 ± 3 s (inter-trial period). Early presses were discouraged by the addition of a 2s delay of the cue. During training and experimental sessions, additional tastants were automatically delivered through the IOC at random times near the middle of the inter-trial interval and in the absence of the anticipatory cue ('unexpected' trials). The following tastants were delivered: 100 mM NaCl, 100 mM sucrose, 100 mM citric acid, and 1 mM quinine HCl. Water (50 μl) was delivered to rinse the mouth clean through a second IOC, 5s after the delivery of each tastant.

Multiple single-unit action potentials were amplified, bandpass filtered, and digitally recorded. Single units were isolated using a template algorithm, clustering techniques, and examination of inter-spike interval plots (Offline Sorter, Plexon). Starting from a pool of 299 single neurons in 37 sessions, neurons with peak firing rate lower than 1 Hz (defined as silent) were excluded from further analysis, as well as neurons with a large peak around the 6-10 Hz in the spike power spectrum, which were classified as somatosensory[2-4]. Only ensembles with 5 or more simultaneously recorded neurons were included in the rest of the analyses. Ongoing and evoked activity were defined as occurring in the 5 seconds-long interval before or after taste delivery, respectively.

Ensemble states detection. A Hidden Markov Model (HMM) analysis was used to detect ensemble states in both the empirical data and model simulations. Here, we give a brief description of the method used and we refer the reader to Refs. [5-8] for more detailed information.

The HMM assumes that an ensemble of N simultaneously recorded neurons is in one of M hidden states at each given time bin. States are firing rate vectors $r_i(m)$, where $i = 1, \ldots, N$ is the neuron index and $m = 1, \ldots, M$ identifies the state. In each state, neurons were assumed to discharge as stationary Poisson processes (Poisson-HMM) conditional on the state firing rate. Trials were segmented in 2 ms bins, and the value of either 1 (spike) or 0 (no spike) was assigned to each bin for each given neuron; if more than one

neuron fired in a given bin (a rare event), a single spike was randomly assigned to one of the firing neurons (Bernoulli approximation). A single HMM was used to fit all trials in each recording session, resulting in the emission probabilities $r_i(m)$ and in a set of transition probabilities between the states. Emission and transition probabilities were calculated with the Baum-Welch algorithm[9] with a fixed number of hidden states M, yielding a maximum likelihood estimate of the parameters given the observed spike trains. Since the model log-likelihood LL increases with M, we repeated the HMM fits for increasing values of M until we hit a minimum of the Bayesian Information Criterion (BIC, see below and Ref.[9]). For each M, the LL used in the BIC was the sum over 10 independent HMM fits with random initial guesses for emission and transition probabilities. This step was needed since the Baum-Welch algorithm only guarantees reaching a local rather than global maximum of the likelihood. The model with the lowest BIC (having M^* states) was selected as the winning model, where $BIC = -2LL + [M(M-1) + MN] \cdot \ln T$, T being the number of observations in each session (= number of trials x number of bins per trials). Finally, the winning HMM model was used to "decode" the states from the data according to their posterior probability given the data. During decoding, only those states with probability exceeding 80% in at least 25 consecutive 2ms-bins were retained (henceforth denoted simply as "states")[6,8]. This procedure eliminates states that appear only very transiently and with low probability, also reducing the chance of overfitting. A median of 6 states per ensemble was found, varying from 3 to 9 across ensembles.

Coding states. In each condition (i.e., expected vs. unexpected), the frequencies of occurrence of a given state across taste stimuli were compared with a test of proportions (chi-square, $p<0.001$ with Bonferroni correction to account for multiple states). When a significant difference was found across stimuli, a post-hoc Marascuilo test was performed. A state whose frequency of occurrence was significantly higher in the presence of one taste stimulus compared to all other tastes was deemed a 'coding state' for that stimulus (Fig. 4).

Spiking network model. We modeled the local neural circuit as a recurrent network of N leaky-integrate-and-fire (LIF) neurons, with a fraction $n_E = 80\%$ of excitatory (E) and $n_I = 20\%$ of inhibitory (I) neurons.[10] Connectivity was random with probability $p_{EE} = 0.2$ for E to E connections and $p_{EI} = p_{IE} = p_{II} = 0.5$ otherwise. Synaptic weights J_{ij} from pre-synaptic neuron $j = E, I$ to post-synaptic neuron $i = E, I$ scaled as $J_{ij} = j_{ij}/\sqrt{N}$, with j_{ij} drawn from normal distributions with mean $j_{\alpha\beta}$ (for $\alpha, \beta = E, I$) and 1% SD. Networks of different architectures were considered: *i)* networks with segregated clusters (referred to as "clustered network," parameters Tables 1 and 2), and *ii)* networks with overlapping clusters (see Suppl. Table 1 and Suppl. Methods for details),

iii) homogeneous networks (Suppl. Table 2). In the clustered network, E neurons were arranged in Q clusters with $N_c = 100$ neurons per clusters on average (1% SD), the remaining fraction n_{bg} of E neurons belonging to an unstructured "background" population. In the clustered network, neurons belonging to the same cluster had intra-cluster synaptic weights potentiated by a factor J_+; synaptic weights between neurons belonging to different clusters were depressed by a factor $J_- = 1 - \gamma f(J_+ - 1) < 1$ with $\gamma = 0.5$; $f = (1 - n_{bg})/Q$ is the average number of neurons in each cluster.[10] When changing the network size N, all synaptic weights J_{ij} were scaled by $\sqrt{N}$, the intra-cluster potentiation values were J_+=5, 10, 20, 30, 40 for $N = 1, 2, 4, 6, 8 \times 10^3$ neurons, respectively, and cluster size remained unchanged (see also Table 1); all other parameters were kept fixed. In the homogeneous network, there were no clusters ($J_+ = J_- = 1$).

Model neuron dynamics. Below threshold the LIF neuron membrane potential evolved in time as

$$\frac{d}{dt}V = -\frac{V}{\tau_m} + I_{rec} + I_{ext} ,$$

where τ_m is the membrane time constant and the input currents I are a sum of a recurrent contribution I_{rec} coming from the other network neurons and an external current $I_{ext} = I_0 + I_{stim} + I_{cue}$ (units of Volt/s). Here, I_0 is a constant term representing the input from other brain areas; I_{stim} and I_{cue} represent the incoming stimuli and cue, respectively (see *Stimulation protocols* below). When V hits threshold V_{thr}, a spike is emitted and V is then clamped to the rest value V_{reset} for a refractory period τ_{ref}. Thresholds were chosen so that the homogeneous network neurons fired at rates $r_E = 5$ spks/s and $r_I = 7$ spks/s. The recurrent contribution to the postsynaptic current to the *i*-th neuron was a low-pass filter of the incoming spike trains

$$\tau_{syn}\frac{d}{dt}I_{rec} = -I_{rec} + \sum_{j=1}^{N} J_{ij} \sum_k \delta(t - t_k^j) ,$$

where τ_{syn} is the synaptic time constant; J_{ij} is the recurrent synaptic weights from presynaptic neuron *j* to postsynaptic neuron *i*, and t_k^j is the *k*-th spike time from the *j*-th presynaptic neuron. The constant external current was $I_0 = N_{ext} p_{i0} J_{i0} \nu_{ext}$, with $N_{ext} = n_E N$, $p_{i0} = 0.2$, $J_{i0} = j_{i0}/\sqrt{N}$ with j_{E0} for excitatory and j_{I0} for inhibitory neurons (see Table 1), and $r_{ext} = 7$ spks/s. For a detailed mean field theory analysis of the clustered network and a comparison between simulations and mean field theory during ongoing and stimulus-evoked periods we refer the reader to the Suppl. Methods and Refs.[8,11].

Stimulation protocols. Stimuli were modeled as time-varying afferent currents targeting 50% of neurons in stimulus-selective clusters $I_{stim}(t) = I_0 \cdot r_{stim}(t)$, where $r_{stim}(t)$ was expressed as a fraction of the baseline external current I_0. Each cluster had a 50% probability of being selective to a given stimulus, thus different stimuli targeted overlapping sets of clusters. The anticipatory cue, targeting a random 50% subset of E neurons, was modeled as a double exponential with rise and decay times of 0.2 s and 1 s, respectively, unless otherwise specified; its peak value for each selective neuron was sampled from a normal distribution with zero mean and standard deviation σ (expressed as fraction of the baseline current I_0; σ=20% unless otherwise specified). The cue didn't change the mean afferent current but only its spatial (quenched) variance across neurons.

In the unexpected condition, stimulus onset at $t = 0$ was followed by a linear increase $r_{stim}(t)$ in the afferent current to stimulus-selective neurons reaching a value r_{max} at $t =$ 1 s ($r_{max} = 20\%$, unless otherwise specified). In the expected condition, stimuli were preceded by the anticipatory cue $r_{cue}(t)$ with onset at $t = -0.5$ s before stimulus presentation.

Network simulations. All data analyses, model simulations and mean field theory calculations were performed using custom software written in MATLAB (MathWorks), and C. Simulation comprised 20 realizations of each network (each one representing a different experimental session), with 20 trials per stimulus in each of the 2 conditions (unexpected and expected); or 40 trials per session in the condition with "cue-on" and no stimuli (Fig. 2). Dynamical equations for the LIF neurons were integrated with the Euler method with 0.1 ms step. All simulation scripts are available from the authors upon request.

Mean field theory. Mean field theory was used in a simplified network with 2 excitatory clusters (parameters in Table 2) using the population density approach[12-14]: the input to each neuron was completely characterized by the infinitesimal mean μ_α and variance σ_α^2 of the post-synaptic current (see Supplementary Methods for their expressions). The network fixed points satisfied the $Q + 2$ self-consistent mean field equations[10]

$$r_\alpha = F_\alpha(\mu_\alpha(\boldsymbol{r}), \sigma_\alpha^2(\boldsymbol{r})) \,, \tag{1}$$

where $\boldsymbol{r} = [r_1, \dots, r_Q, r_E^{bg}, r_I]$ is the population firing rate vector (boldface represents vectors). F_α is the current-to-rate function for population α, which varied depending on the population and the condition. In the absence of the anticipatory cue, the LIF current-to-rate function F_α^0 was used

$$F_\alpha^0(\mu_\alpha,\sigma_\alpha) = \left(\tau_{ref} + \tau_{m,\alpha}\sqrt{\pi}\int_{H_{eff,\alpha}}^{\Theta_{eff,\alpha}} e^{u^2}\,[1+\mathrm{erf}(u)]\right)^{-1},$$

where $\Theta_{eff,\alpha} = \frac{V_{thr,\alpha}-\mu_\alpha}{\sigma_\alpha} + ak_\alpha$, $H_{eff,\alpha} = \frac{V_{reset,\alpha}-\mu_\alpha}{\sigma_\alpha} + ak_\alpha$. Here, $k_\alpha = \sqrt{\tau_{syn,\alpha}/\tau_{m,\alpha}}$, $a = \frac{|\zeta(1/2)|}{\sqrt{2}} \sim 1.03$.[15,16] In the presence of the anticipatory cue, a modified current-to-rate function F_α^{cue} was used to capture the cue-induced Gaussian quenched noise in the afferent currents to the cue-selective populations ($\alpha = 1, \dots, Q$):

$$F_\alpha^{cue}(\mu_\alpha,\sigma_\alpha) = \int Dz\, F_\alpha^0(\mu_\alpha + z\,\sigma\,\mu_{ext},\sigma_\alpha),$$

where $Dz = dz \exp\left(-\frac{z^2}{2}\right)/\sqrt{2\pi}$ is a Gaussian measure with zero mean and unit variance, $\mu_{ext} = I_0$ is the baseline afferent current and σ is the anticipatory cue's SD as fraction of μ_{ext} . Fixed points $\boldsymbol{r}^*$ of equation (1) were found with Newton's method, and represented attractors when the stability matrix

$$S_{\alpha\beta} = \frac{1}{\tau_{s,\alpha}}\left(\frac{\partial F_\alpha(\mu_\alpha(\boldsymbol{r}),\sigma_\alpha^2(\boldsymbol{r}))}{\partial r_\beta} - \frac{\partial F_\alpha(\mu_\alpha(\boldsymbol{r}),\sigma_\alpha^2(\boldsymbol{r}))}{\partial \sigma_\alpha^2}\frac{\partial \sigma_\alpha^2}{\partial r_\beta} - \delta_{\alpha\beta}\right), \tag{2}$$

evaluated at the fixed point $\boldsymbol{r}^*$ was negative definite. Stability was defined with respect to an approximate linearized dynamics of the mean m_α and SD s_α of the input currents[17]

$$\begin{aligned} \tau_{syn,\alpha}\frac{dm_\alpha}{dt} &= -\,m_\alpha + \mu_\alpha(\boldsymbol{r}) \,, \\ \frac{\tau_{syn,\alpha}}{2}\frac{ds_\alpha^2}{dt} &= -\,s_\alpha^2 + \sigma_\alpha^2(\boldsymbol{r}) \,, \\ r_\alpha(t) &= F_\alpha(m_\alpha(\boldsymbol{r}), s_\alpha^2(\boldsymbol{r})) \,, \end{aligned} \tag{3}$$

where μ_α and σ_α^2 are the stationary values given in the Suppl. Methods.

Effective mean field theory for the reduced network. The mean field theory equations (1) for the *P*=*Q*+2 populations may be reduced to a set of effective equations governing the dynamics of a smaller subset of *q*<*P* of populations, henceforth referred to as populations *in focus*[18]. The reduction is achieved by integrating out the remaining *P-q out-of-focus* populations. This procedure was used to estimate the energy barrier separating the two network attractors in Figg. 2 and 3. Given a fixed set of values $\tilde{\boldsymbol{r}} = [\tilde{r}_1, \dots, \tilde{r}_q]$ for the in-focus populations, one solves the mean field equations for *P-q* out-of-focus populations

$$r_\beta(\tilde{r}_1, \dots, \tilde{r}_q) = F_\beta[\mu_\beta(\tilde{r}_1, \dots, \tilde{r}_q, r_{q+1}, \dots, r_P), \sigma_\beta^2(\tilde{r}_1, \dots, \tilde{r}_q, r_{q+1}, \dots, r_P)] \,,$$

for $\beta = q+1, \dots, P$ to obtain the stable fixed point $\boldsymbol{r}'(\tilde{\boldsymbol{r}}) = [r'_{q+1}(\tilde{\boldsymbol{r}}), \dots r'_P(\tilde{\boldsymbol{r}})]$ of the out-of-focus populations as functions of the in-focus firing rates $\tilde{\boldsymbol{r}}$. Stability of the solution $\boldsymbol{r}'(\tilde{\boldsymbol{r}})$ is computed with respect to the stability matrix (2) of the reduced system of *P-q* out-of-focus populations. Substituting the values $\boldsymbol{r}'(\tilde{\boldsymbol{r}})$ into the fixed-point equations for the q populations in focus yields a new set of equations relating input rates $\tilde{\boldsymbol{r}}$ to "output" rates $\boldsymbol{r}^{out}$:

$$r_\alpha^{out}(\tilde{\boldsymbol{r}}) = F_\alpha[\mu_\alpha(\tilde{\boldsymbol{r}}, \boldsymbol{r}'(\tilde{\boldsymbol{r}})), \sigma_\alpha^2(\tilde{\boldsymbol{r}}, \boldsymbol{r}'(\tilde{\boldsymbol{r}}))] \,,$$

for $\alpha = 1, \dots, q$. The input $\tilde{\boldsymbol{r}}$ and output $\boldsymbol{r}^{out}$ firing rates of the in-focus populations will be different, except at a fixed point of the full system where they coincide. The correspondence between input and output rates of in-focus populations defines the effective current-to-rate transfer functions

$$r_\alpha^{out}(\tilde{\boldsymbol{r}}) = F_\alpha^{eff}[\mu_\alpha(\tilde{\boldsymbol{r}}), \sigma_\alpha^2(\tilde{\boldsymbol{r}})] \,, \tag{4}$$

for $\alpha = 1, \dots, q$ in-focus populations at the point $\tilde{\boldsymbol{r}}$. The fixed points $r_\alpha^{out}(\tilde{\boldsymbol{r}}^*) = \tilde{r}_\alpha^*$ of the in-focus equations (4) are fixed points of the entire system. In case the out-of-focus populations attain multiple attractors for a given value of $\tilde{\boldsymbol{r}}$, the set of effective transfer functions F_α^{eff} is labeled by the chosen attractor, although in our analysis of the two-clustered network only one attractor was present for a given value of $\tilde{\boldsymbol{r}}$.

Energy potential. In a network with Q=2 clusters, one can integrate out the background and inhibitory (out-of-focus) populations to obtain the effective transfer functions for the two (in-focus) populations representing the two clusters, with firing rates $\tilde{\boldsymbol{r}} = [\tilde{r}_1, \tilde{r}_2]$ (equation (4) for q=2). Network dynamics can be visualized on a two-dimensional plane $[\tilde{r}_1, \tilde{r}_2]$. The mismatch between input and output firing rates of the in-focus populations yields an effective force $f_\alpha = r_\alpha^{out}(\tilde{r}) - \tilde{r}_\alpha$, for $\alpha = 1, 2$, according to (4). A fixed point of the full system is given by $\boldsymbol{f} = [f_1, f_2] = 0$. The two-cluster network has two attractors ('A' and 'B' in Fig. 2c) and a saddle point C (local maximum in Fig. 2c). We estimated the height Δ of the potential energy barrier on the trajectory from A to B through C as the line integral of the force from the first attractor A to C

$$\Delta = -\int_A^C \boldsymbol{f} \cdot \boldsymbol{ds} \,,$$

where $\boldsymbol{ds}$ is the line element on the trajectory through points of minimal absolute value of the force, $|\boldsymbol{f}|$. Other barrier heights were estimated in a similar way.

Population decoding. The amount of stimulus-related information carried by spike trains was assessed through a decoding analysis[19]. A multiclass classifier was constructed from Q neurons sampled from the population (one neuron from each of the Q clusters for clustered networks, or Q random excitatory neurons for homogeneous networks).

Spike counts from all trials of n_{stim} taste stimuli in each condition (expected vs. unexpected) were split into training and test sets for cross-validation. A "template" was created for each stimulus, condition and time bin (200 ms, sliding over in 50 ms steps) as the population PSTH over the training set, containing the trial-averaged spike counts of each neuron in each bin (the same number of trials across stimuli and conditions were used). Population spike counts for each test trial were classified according to the smallest Euclidean distance from the templates across 10 training sets ('bagging' or bootstrap aggregating procedure[20]). Specifically, from each training set L, we created bootstrapped training sets L_b, for $b = 1, .., B$=10, by sampling with replacement from L. In each bin, each test trial was then classified B times using the B classifiers, obtaining B different votes, and the most frequent vote was chosen as the bagged classification of the test trial. Cross-validated decoding accuracy in a given bin was defined as the fraction of correctly classified test trials in that bin.

Significance of decoding accuracy was established via a permutation test: 1000 shuffled datasets were created by randomly permuting stimulus labels among trials, and a 'shuffled distribution' of 1000 decoding accuracies was obtained. In each bin, decoding accuracy of the original dataset was deemed significant if it exceeded the upper bound, $\alpha_{0.05}$, of the 95% confidence interval of the shuffled accuracy distribution in that bin (this included a Bonferroni correction for multiple bins, so that $\alpha_{0.05} = 1 - 0.05/N_b$, with N_b the number of bins). Decoding latency (insets in Figs. 1c and 1f) was estimated as the earliest bin with significant decoding accuracy.

Cluster dynamics. For cluster dynamics analyses (lifetime, inter-activation interval, latency; Figg. 2 and 3), cluster spike count vectors r_i (for $i = 1, \ldots, Q$) in 5 ms bins were obtained by averaging spike counts of neurons belonging to a given cluster. A cluster was deemed active if its firing rate exceeded 10 spks/s. This threshold was chosen to lie between the inactive and active clusters' firing rates, which were obtained from a mean field theory solution of the network[8].

Symbol	Description	Value
j_{EE}	Mean E→E synaptic weight × $\sqrt{N}$	1.1 mV
j_{EI}	Mean I→E synaptic weight × $\sqrt{N}$	5.0 mV
j_{IE}	Mean E→I synaptic weight × $\sqrt{N}$	1.4 mV
j_{II}	Mean I→I synaptic weight × $\sqrt{N}$	6.7 mV
j_{E0}	Mean afferent synaptic weights to E neurons × $\sqrt{N}$	5.8 mV
j_{I0}	Mean afferent synaptic weights to I neurons × $\sqrt{N}$	5.2 mV
r_{ext}^{E}	Average afferent rate to E neurons (baseline)	5 spks/s
r_{ext}^{I}	Average afferent rate to I neurons (baseline)	7 spks/s
V_{thr}^{E}	E neuron threshold potential	3.9 mV
V_{thr}^{I}	I neuron threshold potential	4.0 mV
V_{reset}	E and I neurons reset potential	0 mV

τ_m	E and I membrane time constant	20 ms
τ_{ref}	Absolute refractory period	5 ms
τ_{syn}	E and I synaptic time constant	4 ms
n_{bg}	Background neurons fraction	10%
N_c	Average cluster size	100 neurons

Table 1: Parameters for the clustered network with N LIF neurons. The intra-cluster potentiation parameter values were J_+= 5, 10, 20, 30, 40 for networks with $N = 1, 2, 4, 6, 8 \times 10^3$ neurons, respectively.

Symbol	Description	Value
j_{EE}	Mean E→E synaptic weight × $\sqrt{N}$	0.8 mV
j_{EI}	Mean I→E synaptic weight × $\sqrt{N}$	10.6 mV
j_{IE}	Mean E→I synaptic weight × $\sqrt{N}$	2.5 mV
j_{II}	Mean I→I synaptic weight × $\sqrt{N}$	9.7 mV
j_{E0}	Mean afferent synaptic weights to E neurons × $\sqrt{N}$	14.5 mV
j_{I0}	Mean afferent synaptic weights to I neurons × $\sqrt{N}$	12.9 mV
J_+	Potentiated intra-cluster E→E weights factor	9
V_{thr}^E	E neuron threshold potential	4.6 mV
V_{thr}^I	I neuron threshold potential	8.7 mV
n_{bg}	Background neurons fraction	65%

Table 2: Parameters for the simplified 2-cluster network with N=800 LIF neurons (the remaining parameters were as in Table 1).

References

- 1 Samuelsen, C. L., Gardner, M. P. & Fontanini, A. Effects of cue-triggered expectation on cortical processing of taste. *Neuron* **74**, 410-422, doi:10.1016/j.neuron.2012.02.031 (2012).
- 2 Katz, D. B., Simon, S. A. & Nicolelis, M. A. Dynamic and multimodal responses of gustatory cortical neurons in awake rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **21**, 4478-4489 (2001).
- 3 Jezzini, A., Mazzucato, L., La Camera, G. & Fontanini, A. Processing of hedonic and chemosensory features of taste in medial prefrontal and insular networks. *J Neurosci* **33**, 18966-18978, doi:10.1523/JNEUROSCI.2974-13.2013 (2013).
- 4 Horst, N. K. & Laubach, M. Reward-related activity in the medial prefrontal cortex is driven by consumption. *Front Neurosci* **7**, 56, doi:10.3389/fnins.2013.00056 (2013).

- 5 Abeles, M. *et al.* Cortical activity flips among quasi-stationary states. *Proc Natl Acad Sci USA* **92**, 8616-8620 (1995).
- 6 Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P. & Katz, D. B. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18772-18777 (2007).
- 7 Ponce-Alvarez, A., Nacher, V., Luna, R., Riehle, A. & Romo, R. Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32**, 11956-11969, doi:10.1523/jneurosci.6176-11.2012 (2012).
- 8 Mazzucato, L., Fontanini, A. & La Camera, G. Dynamics of Multistable States during Ongoing and Evoked Cortical Activity. *J Neurosci* **35**, 8214-8231, doi:10.1523/JNEUROSCI.4819-14.2015 (2015).
- 9 Zucchini, W. & MacDonald, I. L. *Hidden Markov models for time series: an introduction using R*. (CRC Press, 2009).
- 10 Amit, D. J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)* **7**, 237-252 (1997).
- 11 Mazzucato, L., Fontanini, A. & La Camera, G. Stimuli Reduce the Dimensionality of Cortical Activity. *Frontiers in systems neuroscience* **10**, 11, doi:10.3389/fnsys.2016.00011 (2016).
- 12 Tuckwell, H. C. *Introduction to theoretical neurobiology*. Vol. 2 (Cambridge University Press, 1988).
- 13 Lansky, P. & Sato, S. The stochastic diffusion models of nerve membrane depolarization and interspike interval generation. *J Peripher Nerv Syst* **4**, 27-42 (1999).
- 14 Richardson, M. J. The effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons. *Physical review. E, Statistical, nonlinear, and soft matter physics* **69**, 051,918 (2004).
- 15 Brunel, N. & Sergi, S. Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics. *J Theor Biol* **195**, 87-95, doi:10.1006/jtbi.1998.0782 (1998).
- 16 Fourcaud, N. & Brunel, N. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural computation* **14**, 2057-2110, doi:10.1162/089976602320264015 (2002).
- 17 La Camera, G., Rauch, A., Luscher, H. R., Senn, W. & Fusi, S. Minimal models of adapted neuronal response to in vivo-like input currents. *Neural Comput* **16**, 2101-2124 (2004).
- 18 Mascaro, M. & Amit, D. J. Effective neural response function for collective population states. *Network (Bristol, England)* **10**, 351-373 (1999).
- 19 Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585-590, doi:10.1038/nature12160 (2013).
- 20 Breiman, L. Bagging predictors. *Machine learning* **24**, 123-140 (1996).

Figures

Fig. 1

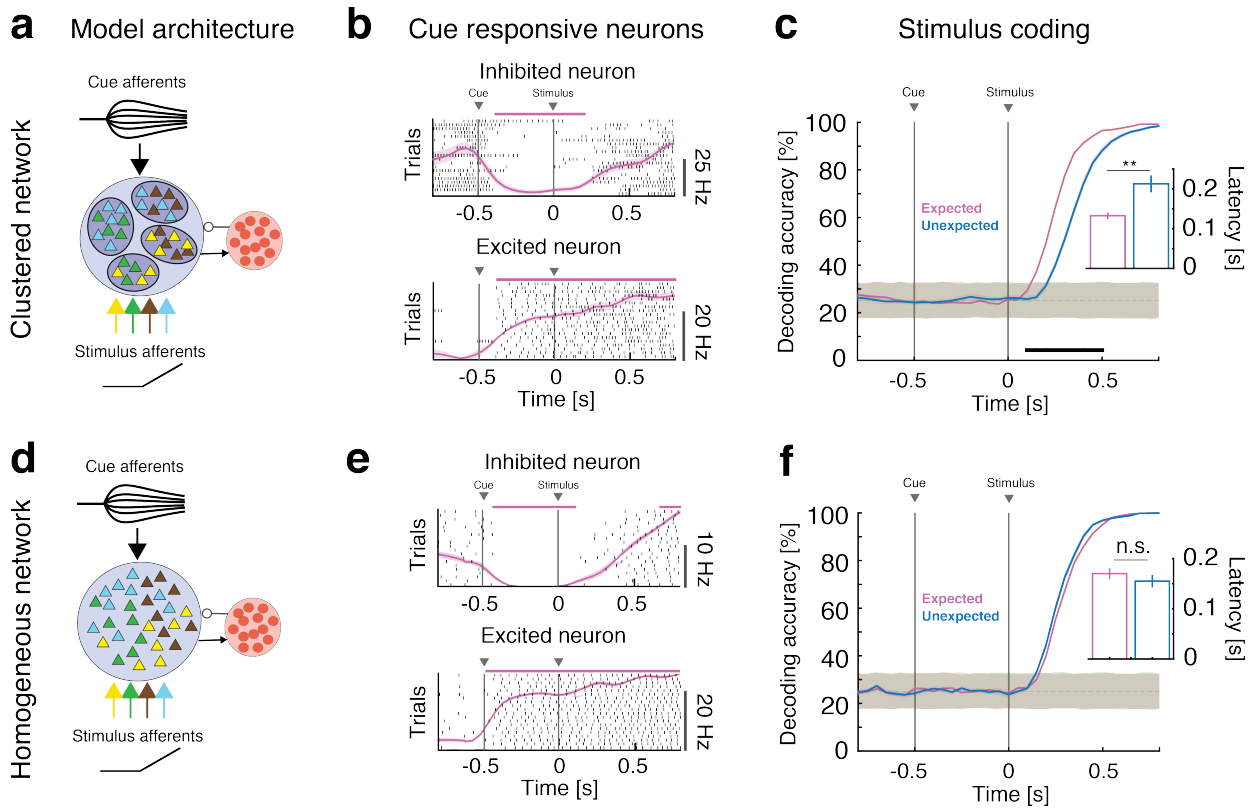

Fig. 1: **Anticipatory activity requires clustered network architecture.** Effects of anticipatory cue on stimulus coding in clustered (**a-c**) and homogeneous (**d-f**) networks. Schematics of clustered network architecture and stimulation. A recurrent network of inhibitory (red circles) and excitatory neurons (triangles, arranged in clusters in **a**, darker blue shades representing stronger intra-cluster recurrent connectivity; or with homogeneous connectivity in **d**) receives bottom-up sensory stimuli targeting random, overlapping subsets of clusters (selectivity to 4 stimuli is color-coded), and one top-down anticipatory cue inducing a spread in the afferent currents to excitatory neurons. In both clustered and homogeneous networks, Representative single neuron responses to cue and stimulus in expected trials in the clustered (**b**) and homogeneous (**e**) networks (black vertical bars represent action potentials; mean±s.e.m. of PSTH overlaid in pink) are significantly different from baseline (pink horizontal bars, $p<0.05$, t-test with multiple bin correction) and can be either excited (top panels) or inhibited (bottom panels). Time course of cross-validated decoding accuracy increases faster during expected (pink, shaded areas represent s.e.m.) than unexpected (blue) trials (black bar:

$p<0.05$ with multiple bin correction; gray shaded areas represent 95% shuffled C.I.) in clustered networks (**c**), but not in homogeneous networks (**f**). Latency of a significant decoding is faster in expected than in unexpected trials (mean±s.e.m.; **=$p<0.01$, t-test) in clustered (inset in **c**) but no in homogeneous networks (inset in **f**).

Fig. 2

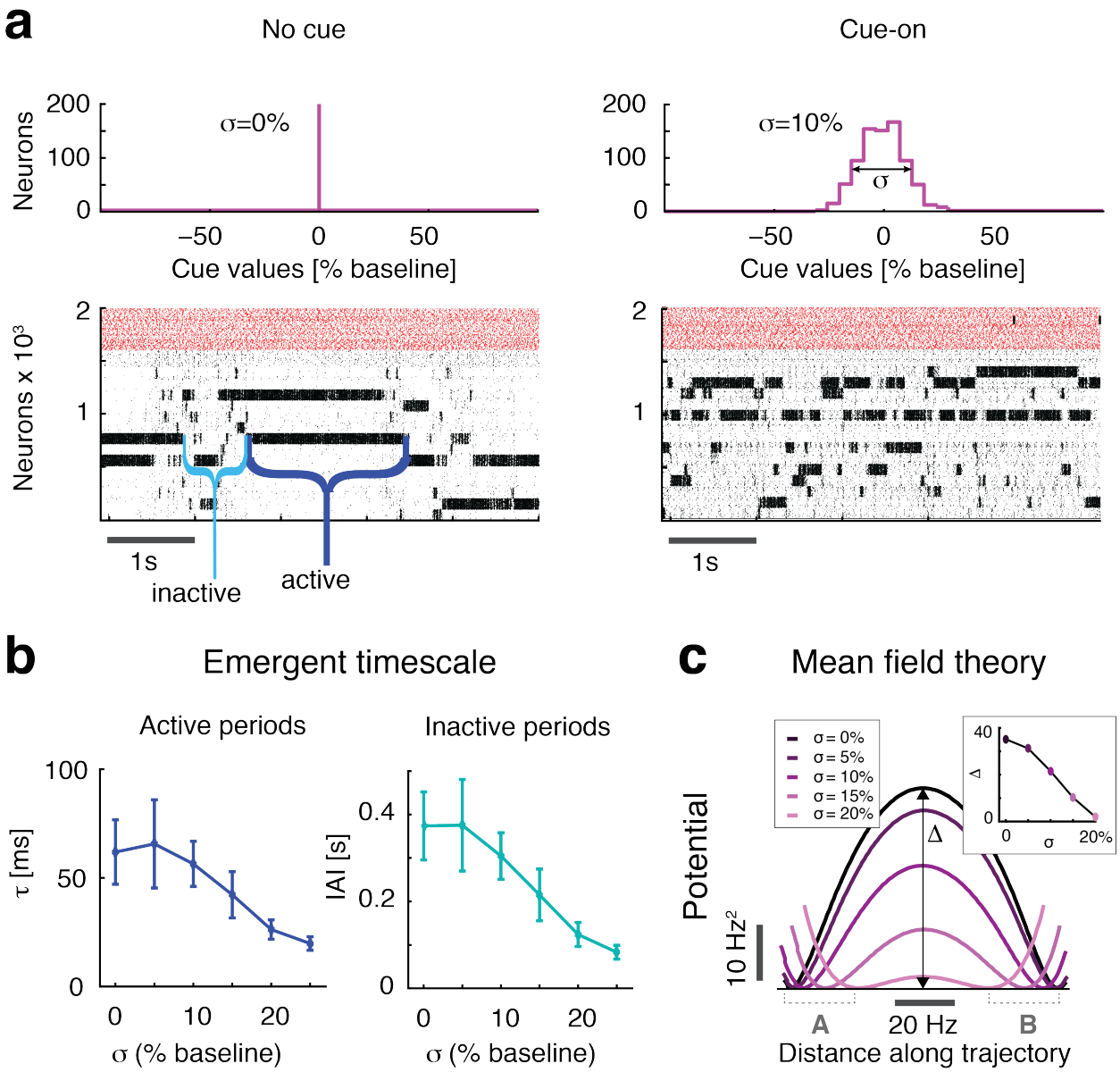

Fig. 2: **Anticipatory cue speeds up network dynamics**. **a**: In trials where the cue was presented in the absence of stimuli, the dynamics of cluster activation and deactivation accelerated proportionally to the increase in afferent currents' spread σ induced by the cue (top: distribution of cue peak values across neurons: left, no cue; right, σ = 10% in units of baseline current; bottom: representative trials in each condition, network neurons arranged according to cluster membership; black and red dots represent excitatory and inhibitory neurons' action potentials, respectively). **b**: The average

cluster activation lifetime (left) and inter-activation interval (right) significantly decrease when increasing σ. **c**: Activity in a simplified two-cluster network is described by two states (potential wells 'A' and 'B') where either cluster is active alone. In presence of the cue, but in the absence of taste stimuli, the two potential wells are symmetric (curves represent energy potentials calculated via mean field theory). The anticipatory cue shrinks the height Δ of the potential barrier on the trajectory connecting the two states (inset, lighter pink shades denote stronger cue, i.e., larger σ), making transitions between states more likely.

Fig. 3

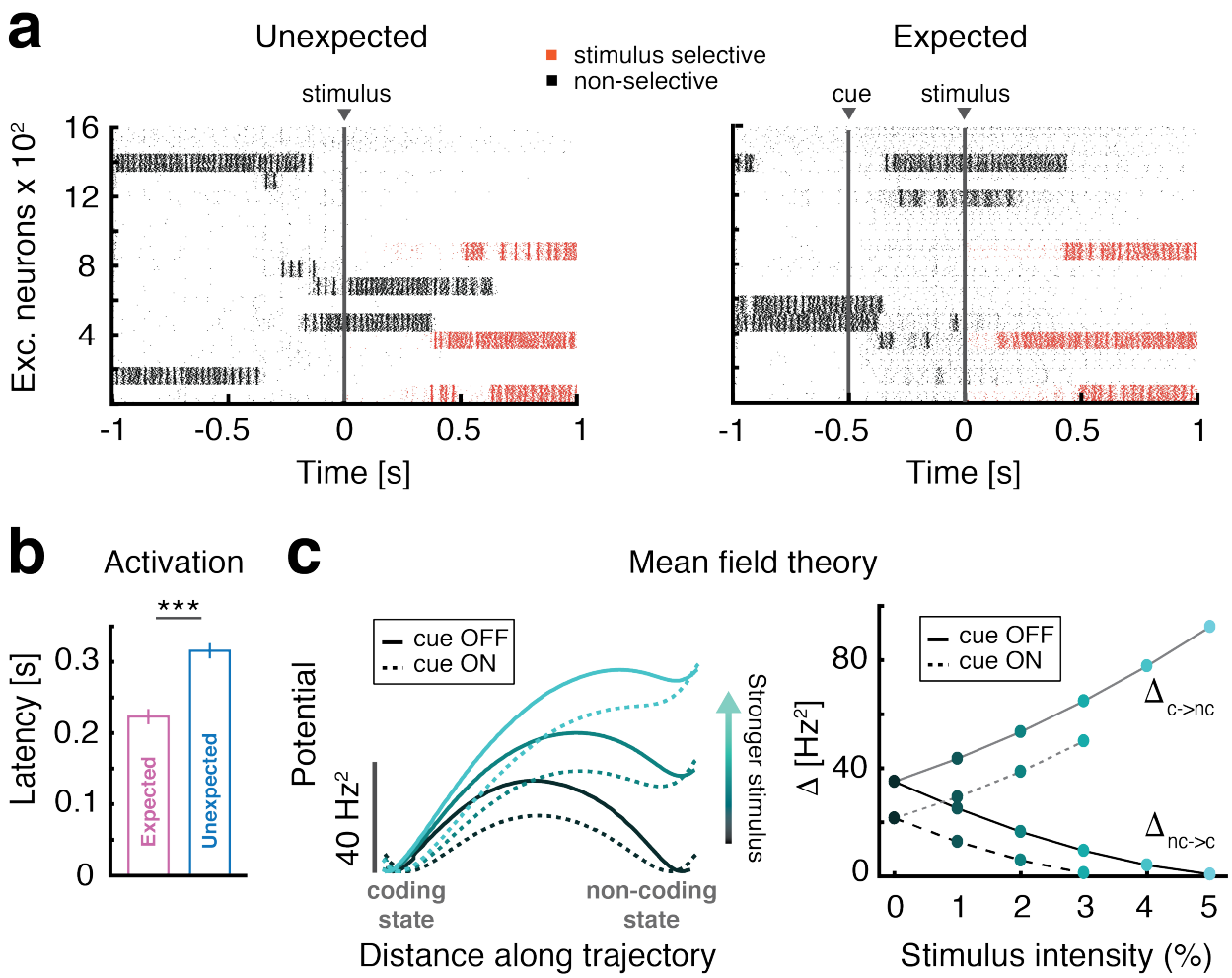

Fig. 3: **Anticipatory cue induces faster onset of stimulus-coding states. a:** Representative trials in the unexpected (left) and expected (right) conditions in response to the same stimulus at *t=0*. Stimulus-selective clusters activate earlier in response to the stimulus when the cue precedes the stimulus (excitatory neurons arranged according to cluster membership; red dots, action potentials emitted by neurons in selective clusters). **b**: Activation latency of selective clusters after stimulus presentation is anticipated during expected (pink) compared to unexpected (blue) trials (mean±s.e.m., ****p<0.001*). **c**: In a simplified network with two states, after stimulus presentation the stimulus-coding state (left well in left panel) is more likely to occur than the non-coding state (right well). Stronger stimuli (lighter shades of cyan) decrease the barrier height Δ (right panel) on the trajectory from the non-coding to the coding state. In expected trials (dashed lines), the barrier Δ is smaller than in unexpected ones (full lines), leading to a

faster transition probability from non-coding to coding states compared to unexpected trials (for stimulus ≥ 4% the barrier vanishes leaving just the coding state).

Fig. 4

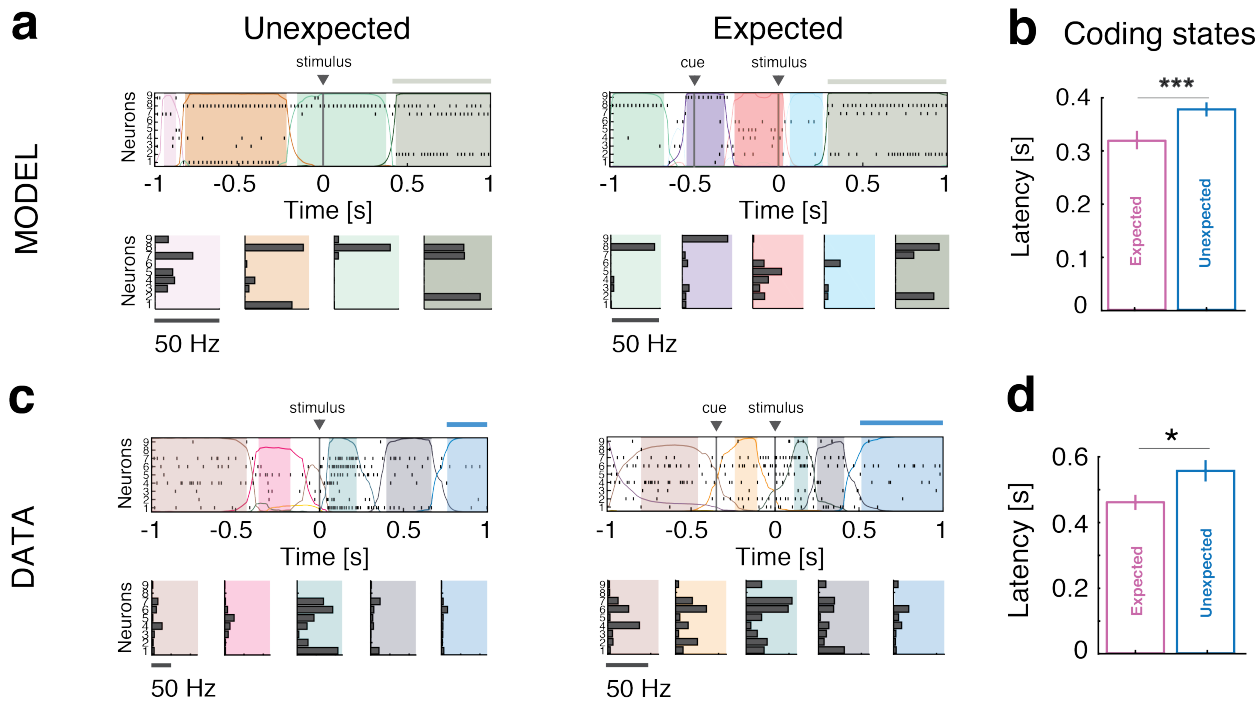

Fig. 4: **Anticipation of coding states: model vs. data**. Representative trials from one ensemble of 9 simultaneously recorded neurons (action potentials, black vertical lines) from model simulations (**a**) and from the empirical data (**c**), during unexpected (left) and expected (right) condition. Top panels: States are extracted via a HMM analysis (colored areas indicate intervals where the probability of a state exceeds 80%; colored curves represent states probabilities). Bottom panels: firing rate vector for each state. Stimulus-coding states (overlaid colored thick bar in **a**, **c**, top panels) have faster onset latency during expected compared to unexpected trials both in the model (**b**) and in the data (**d**) (mean±s.e.m. , *$p<0.05$, ***$p<0.001$).

Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli

Luca Mazzucato[1,2], Giancarlo La Camera[1,3,*], and Alfredo Fontanini[1,3,*]

[1]Department of Neurobiology and Behavior, State University of New York at Stony Brook, Stony Brook, NY 11794
[2]Department of Biology and Mathematics and Institute of Neuroscience, University of Oregon, Eugene, OR 97403
[3]Graduate Program in Neuroscience, State University of New York at Stony Brook, Stony Brook, NY 11794
[*]co-corresponding authors

Supplementary Results

Single neuron responses in network models

In the clustered network model of Fig. 1a–c, a large fraction of excitatory neurons was stimulus-responsive ($79 \pm 6\%$ across 20 simulated networks), and cue responsive ($61 \pm 7\%$, among which $49 \pm 8\%$ were excited and $51 \pm 8\%$ inhibited, mean±SD, see Fig. 1b); $52 \pm 8\%$ of neurons responded to both stimulus and cue. Single neurons' responses thus reflected the large heterogeneity of cue responses observed in the empirical data [1, 2].

In the homogeneous network of Fig. 1d–f, single neurons' response properties were comparable to the ones observed in the clustered network. A large fraction of neurons were stimulus-responsive ($69 \pm 3\%$ across 20 simulated networks), or cue responsive ($47 \pm 2\%$, among which $45 \pm 1\%$ were excited and $55 \pm 1\%$ inhibited, see Fig. 1e); $30 \pm 2\%$ of neurons responded to both stimulus and cue.

Robustness of anticipatory activity

To test the robustness of anticipatory activity in the clustered network, we systematically varied key parameters related to the sensory and anticipatory inputs, as well as network connectivity and architecture (Fig. S1). Increasing stimulus intensity led to a faster encoding of the stimulus in both conditions (Fig. S1a). Stimuli were always decoded faster when preceded by the cue (Fig. S1a; two–way ANOVA with factors 'stimulus intensity' (p$< 10^{-18}$, $F(4) = 30.4$) and 'condition' (= expected vs. unexpected; p$< 10^{-15}$, $F(1) = 79.8$)). The amount of anticipatory activity induced by the cue depended on stimulus intensity only weakly (p(interaction)$= 0.05$, $F(4) = 2.4$). Moreover, we found that anticipatory activity was maintained when we increased the number of presented stimuli, while keeping fixed the probability that each cluster was selective to a given stimulus at 50% (Fig. S1b; two-way ANOVA with factors 'number of stimuli' ($p = 0.57$, $F(3) = 0.67$) and 'condition' ($p < 10^{-7}$, $F(1) = 35.3$; p(interaction)$= 0.66$, $F(3) = 0.54$).

Anticipatory activity was robust to variations in cue parameters. Anticipatory activity was present for values of cue–induced spread σ in the afferent currents in the range $\sigma > 10\%$ (Fig. S1c, two–way ANOVA with factors 'cue intensity' ($p = 0.09$, $F(7) = 1.8$) and 'condition' ($p < 10^{-7}$, $F(1) = 31$); p(interaction)$= 0.07$, $F(7) = 1.9$). Anticipatory activity was robust to variations in the time course of cue-evoked currents (Fig. S1e, two–way ANOVA with factors 'time course' ($p = 0.03$, $F(4) = 2.8$) and 'condition' ($p < 10^{-15}$, $F(1) = 72.7$; p(interaction)$= 0.12$, $F(4) = 1.8$).

We then tested robustness of anticipatory activity to variations in network size N and architecture. We estimated the decoding accuracy in ensembles of equal size (20 neurons) sampled from networks of increasing size. When scaling network size we kept fixed both cluster size and the probability that each cluster was selective to a given stimulus (50%). Network synaptic weights scaled as in Table 1 and 2 in the Main Text. Cue-induced anticipation was always present albeit more pronounced in larger networks (Fig. S1d, two–way ANOVA with

factors 'network size' ($p < 10^{-20}$, $F(3) = 49.5$), 'condition' ($p < 10^{-16}$, $F(1) = 90$; p(interaction)= 10^{-10}, $F(3) = 20$).

Finally, we lifted the assumption of segregated clusters, and found that anticipatory activity was present also in networks with overlapping cluster membership (Fig. S1f) which endows the network with a more biologically plausible architecture [8].

We did not find any anticipatory coding in the homogeneous network model as a result of the same manipulations of Fig. S1a-e (not shown).

Supplementary Methods

Mean field theory

The clustered network comprised $Q+2$ neural populations: the first Q populations representing the Q excitatory clusters; the $(Q+1)$–th population representing the "background" unstructured excitatory population; and $(Q+2)$–th population representing the homogeneously connected inhibitory population. In the population density approach, the input to each neuron in population α was completely characterized by the infinitesimal mean μ_α and variance σ_α^2 of the post-synaptic current. For the excitatory clustered population (see Methods and Table 1 in the main text for parameter values), the mean and variance of the infinitesimal moments of the input current read

$$\begin{aligned}
\mu_\alpha &= \tau_{m,E}\sqrt{N}\Big(\frac{n_E f}{Q}\left(p_{EE}J_+ j_{EE} r_\alpha + \sum_{\beta=1}^{Q-1} p_{EE}J_- j_{EE} r_\beta\right) + n_E(1-f)p_{EE}J_- j_{EE} r_E^{(bg)} \\
&\quad - n_I p_{EI} j_{EI} r_I + n_E p_{E0} j_{E0} r_{\text{ext}}\Big)\,, \\
\sigma_\alpha^2 &= \tau_{m,E}\Big(\frac{n_E f}{Q}\left(p_{EE}(J_+ j_{EE})^2(1+\delta^2) r_\alpha + \sum_{\beta=1}^{Q-1} p_{EE}(J_- j_{EE})^2(1+\delta^2) r_\beta\right) \\
&\quad + n_E(1-f)p_{EE}(J_- j_{EE})^2(1+\delta^2) r_E^{(bg)} + n_I p_{EI} j_{EI}^2(1+\delta^2) r_I\Big)\,,
\end{aligned}$$

where $\alpha = 1,\ldots,Q$. $r_E^{(bg)}$ is the firing rate of the background E population, $n_E = 4/5$ and $n_I = 1/5$ are the fraction of excitatory and inhibitory neurons in the network, respectively, and $\delta = 1\%$ is the SD of the distribution of synaptic weights. Afferent current and variance to a neuron belonging to the background E population read

$$\begin{aligned}
\mu_E^{(bg)} &= \tau_{m,E}\sqrt{N}\Big(\frac{n_E f}{Q}\sum_{\beta=1}^{Q} p_{EE}J_- j_{EE} r_\beta + n_E(1-f)p_{EE} j_{EE} r_E^{(bg)} \\
&\quad - n_I p_{EI} j_{EI} r_I + n_E p_{E0} j_{E0} r_{\text{ext}}\Big)\,, \\
(\sigma_E^{(bg)})^2 &= \tau_{m,E}\Big(\frac{n_E f}{Q}\sum_{\beta=1}^{Q} p_{EE}(J_- j_{EE})^2(1+\delta^2) r_\beta + n_E(1-f)p_{EE} j_{EE}^2(1+\delta^2) r_E^{(bg)} \\
&\quad + n_I p_{EI} j_{EI}^2(1+\delta^2) r_I\Big)\,,
\end{aligned}$$

and to the homogeneous inhibitory population read

$$\begin{aligned}
\mu_I &= \tau_{m,I}\sqrt{N}\Big(\frac{n_E f}{Q}\sum_{\beta=1}^{Q} p_{IE} j_{IE} r_\beta + n_E(1-f)p_{IE} j_{IE} r_E^{(bg)} \\
&\quad - n_I p_{II} j_{II} r_I + n_E p_{I0} j_{I0} r_{\text{ext}}\Big)\,, \\
\sigma_I^2 &= \tau_{m,I}\Big(\frac{n_E f}{Q}\sum_{\beta=1}^{Q} p_{IE} j_{IE}^2(1+\delta^2) r_\beta + n_E(1-f)p_{IE} j_{IE}^2(1+\delta^2) r_E^{(bg)} \\
&\quad + n_I p_{II} j_{II}^2(1+\delta^2) r_I\Big). \qquad (0.1)
\end{aligned}$$

Symbol	Description	Value
j_{EE}	Mean E→E weights $\times\sqrt{N}$.	1.4mV
j_{EI}	Mean I→E weight $\times\sqrt{N}$.	5.0mV
j_{IE}	Mean E→I weight $\times\sqrt{N}$.	2.5mV
j_{II}	Mean I→I weight $\times\sqrt{N}$.	6.1mV
j_{E0}	Mean afferent synaptic weights to E neurons $\times\sqrt{N}$.	7.3mV
j_{I0}	Mean afferent synaptic weights to I neurons $\times\sqrt{N}$.	6.5mV
J_+	Potentiated intra-cluster E→E weights factor.	10.5
r_{ext}^E	Average afferent rate to E neurons (baseline).	5 spks/s
r_{ext}^I	Average afferent rate to I neurons (baseline).	7 spks/s
V_E^{thr}	E neuron threshold potential.	3.6mV
V_I^{thr}	I neuron threshold potential.	5.7mV

Table. S1. Parameters for the network with overlapping cluster membership with $N = 2000$ LIF neurons (details in Supplementary Methods; remaining parameters as in Table 1).

Network with overlapping clusters

In the spiking network with overlapping clusters [8], each neuron (network size: $N = 2000$) had a probability $f = 0.06$ of belonging to one of the $Q = 14$ clusters: the fraction of neurons in a population shared by a set of k clusters (a population of 'rank' $[k]$) was $f^k(1-f)^{Q-k}$; the fraction of neurons in the unstructured background population (rank $[0]$) was $(1-f)^Q$. Synaptic weights $J_{[i][j]}$ between pre- and post-synaptic neurons in excitatory populations of ranks $[j]$ and $[i]$ are given by

$$J_{[i][j]} = p_{EE}\left(\epsilon_{[i][j]}\xi_+ + (1-\epsilon_{[i][j]})\xi_-\right) \ ,$$

where p_{EE} is the connection probability, and $\epsilon_{[i][j]}$ and $1-\epsilon_{[i][j]}$ are the potentiation and depression probabilities, respectively. Weights were potentiated or depressed according to a stochastic learning rule [8, 9] that, in the limit of slow learning, leads to the asymptotic value

$$\epsilon_{[i][j]} = \frac{P_{[i][j]}}{P_{[i][j]} + \rho f D_{[i][j]}} \ ,$$

where $P_{[i][j]} = \sum_{k=1}^{Q} \eta_{[i]}^k \eta_{[j]}^k$ is the number of clusters in common between populations $[i]$ and $[j]$ ($\eta_{[i]}^k = 1$ if cluster k contains neuron i and $\eta_{[i]}^k = 0$ otherwise, while $D_{[i][j]} = \sum_{k=1}^{Q} \eta_{[i]}^k(1-\eta_{[j]}^k)$, and $\rho = 2.75$. ξ_+ and ξ_- are random variables sampled from normal distributions with means J_+ and J_- and variances $\delta^2 J_+^2$ and $\delta^2 J_-^2$, respectively ($\delta = 0.01$). The E to I, I to E, and I to I connection probability was the same as for the homogeneous and clustered networks described (Table 1 in the main text). Full details of the theory are found in [9]; the mean field analysis of the network with overlapping clusters is developed in [8]. The parameters of the network are reported in Table S1.

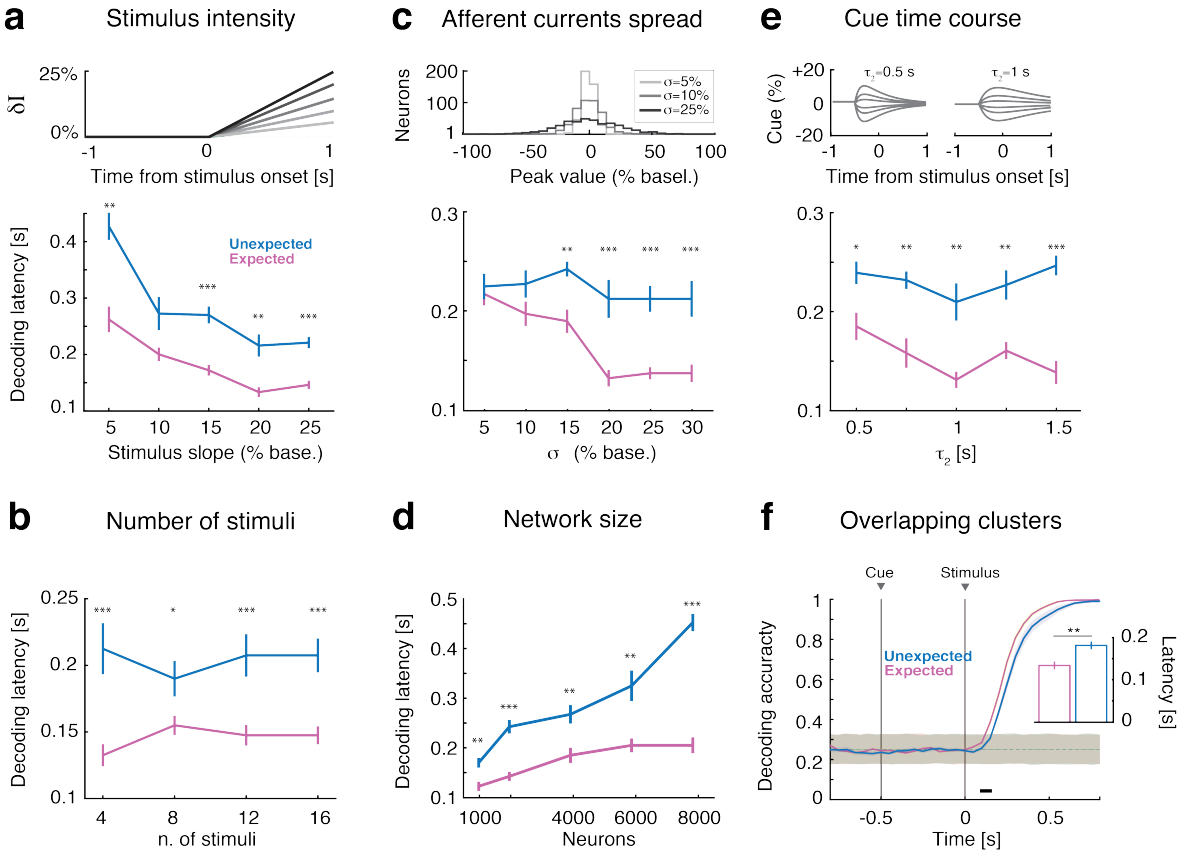

Fig. S1. **Robustness of anticipatory activity in clustered networks**. **a**: Latency of a significant decoding increased with stimulus intensity in both conditions (top, darker shades represent stronger stimuli; bottom, decoding latency, mean±s.e.m.), and it is faster in expected (pink) than unexpected trials (blue, $^{*} = p < 0.05$, $^{**} = p < 0.01$, $^{***} = p < 0.001$, t-test). Decoding latency does not depend on the number of presented stimuli (**b**). Increasing the cue-induced spread in the afferent currents σ (**c**, top: distribution of peak afferent currents across neurons) leads to more pronounced anticipatory activity (bottom, latency in unexpected (blue) and expected (pink) trials). **d**: Anticipatory activity is present for a large range of network sizes; $J_{+} = 5, 10, 20, 30, 40$ for size $N = 1, 2, 4, 6, 8 \times 10^{3}$. **e**: Anticipatory activity is present for a large range of cue time courses (top: double exponential cue profile with rise and decay times $[\tau_1, \tau_2] = g \times [0.1, 0.5]$ s, for g in the range from 1 to 3; bottom: decoding latency during unexpected (blue) and unexpected (pink) trials. **f**: Anticipatory activity is present in a network with overlapping clusters (see Supplementary Methods).

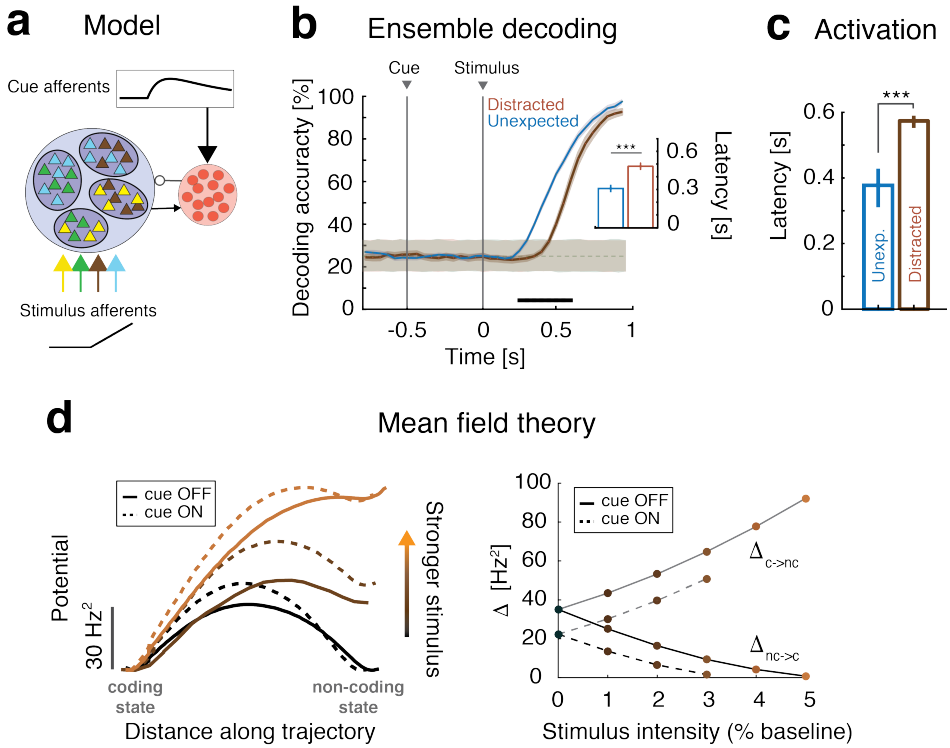

Fig. S2. **A "distracting" cue recruiting recurrent inhibition slows down stimulus coding**. **a**: Schematics of clustered network architecture and stimulation (notations as in Fig. 1a); the cue increased the mean afferent current to the inhibitory population. **b**: Time course of cross–validated population decoding during distracted (brown) trials is slower than during unexpected (blue) trials (notations as in Fig. 1c); decoding latency is delayed in distracted trials (inset, means.e.m., $^{***} = p < 0.001$). **c**: Activation latency of stimulus-selective clusters after stimulus presentation is delayed during distracted trials (latency, mean$\pm$s.e.m.). **d**: In the simplified 2-cluster network (notations as in Fig. 3c, lighter brown denotes stronger stimuli), the transition probability from the non-coding (right well) to the coding state (left well) increased with larger stimuli. In distracted trials (dashed curves) the barrier height Δ from the non-coding to the coding state is larger than in unexpected trials (full curves), leading to slow down of coding in the distracted condition.).

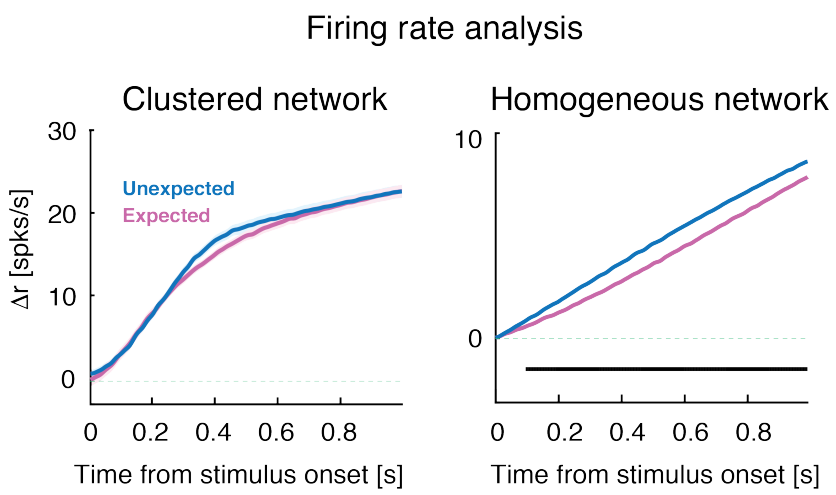

Fig. S3. Time course of the firing rate difference Δr between stimulus-selective and nonselective clusters in the expected (pink curve, shaded area represents s.e.m.) and unexpected (blue) condition. In the clustered network (left, see Fig. 1a), the difference is not significant. In the homogeneous network (right, see Fig. 1d) the difference is larger in the unexpected condition compared to the expected one (black horizontal bar denote significant difference; t-test, p$<$ 0.05 with multiple bin correction).

References

[1] C. L. Samuelsen, M. P. Gardner, and A. Fontanini, *Effects of cue-triggered expectation on cortical processing of taste*, *Neuron* **74** (2012), no. 2 410–422.

[2] R. Vincis and A. Fontanini, *Associative learning changes cross-modal representations in the gustatory cortex*, *Elife* **5** (2016) e16420.

[3] A. Fontanini and D. B. Katz, *State-dependent modulation of time-varying gustatory responses*, *J Neurophysiol* **96** (2006), no. 6 3183–93.

[4] A. Fontanini and D. B. Katz, *7 to 12 hz activity in rat gustatory cortex reflects disengagement from a fluid self-administration task*, *J Neurophysiol* **93** (2005), no. 5 2832–40.

[5] M. Phillips and R. Norgren, *A rapid method for permanent implantation of an intraoral fistula in rats*, *Behavioral Research, Methods, and Instrumentation* (1970), no. 2 124.

[6] D. B. Katz, S. A. Simon, and M. A. Nicolelis, *Dynamic and multimodal responses of gustatory cortical neurons in awake rats*, *J Neurosci* **21** (2001), no. 12 4478–89.

[7] N. K. Horst and M. Laubach, *Reward-related activity in the medial prefrontal cortex is driven by consumption*, *Front Neurosci* **7** (2013) 56.

[8] E. Curti, G. Mongillo, G. La Camera, and D. J. Amit, *Mean field and capacity in realistic networks of spiking neurons storing sparsely coded random memories*, *Neural computation* **16** (2004), no. 12 2597–2637.

[9] N. Brunel, F. Carusi, and S. Fusi, *Slow stochastic hebbian learning of classes of stimuli in a recurrent neural network*, *Network: Computation in Neural Systems* **9** (1998), no. 1 123–152.