Switching Operation Modes in the Neocortex via Cholinergic Neuromodulation

A Computational Model of Uncertainty, Learning and Inhibition

J.-Y. Puigbò, X. D. Arsiwalla, M. A. González-Ballester · P. F. M. J. Verschure

Abstract In order to deal with the uncertainty in the world, our brains need to be able to flexibly switch between the exploration of new sensory representations and exploitation of previously acquired ones. This requires to form accurate estimations of what is expected and how much it is. While modeling has allowed for the development of several ways to form predictions, how the brain could implement those is still under debate. Here, we recognize acetylcholine as one of the main neuromodulators driving learning based on uncertainty, promoting the exploration of new sensory representations. We identify its interactions with cortical inhibitory interneurons and derive a biophysically-grounded

J.-Y. Puigbò
SPECS, Institut de Bioenginyeria de Catalunya (IBEC), Barcelona, Spain
Barcelona Institute of Science and Technology, Barcelona, Spain
Dept. of Information and Comunication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

X. D. Arsiwalla
SPECS, Institut de Bioenginyeria de Catalunya (IBEC), Barcelona, Spain
Barcelona Institute of Science and Technology, Barcelona, Spain
Dept. of Information and Comunication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

M. A. González-Ballester
BCN MEDTech, Dept. of Information and Comunication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
Catalan Institution for Research and Advanced Studies, Barcelona, Spain

P. F. M. J. Verschure
SPECS, Institut de Bioenginyeria de Catalunya (IBEC), Barcelona, Spain
Barcelona Institute of Science and Technology, Barcelona, Spain
Catalan Institution for Research and Advanced Studies, Barcelona, Spain

computational model able to capture and learn from uncertainty. This model allows us to understand inhibition beyond gain control by suggesting that different interneuron sub-types either encode predictions or estimate their uncertainty, facilitating detection of unexpected cues. Moreover, we show how acetylcholine-like neuromodulation uniquely interacts with global and local sources of inhibition, disrupting perceptual certainty and promoting the rapid acquisition of new perceptual cues. Altogether, our model proposes that cortical acetylcholine favors sensory exploration over exploitation in a cortical microcircuit dedicated to estimating sensory uncertainty.

Keywords Neocortex · Inhibition · Learning · Uncertainty · Neuromodulation · Microcircuitry

1 Introduction

Cortical microcircuits are the canonical building blocks of computations in the neocortex [18,5,9]. These canonical circuits across the neocortex are highly plastic and adaptable. Hence, they display remarkable flexibility in expressing a repertoire of complex cortical computations, including those used in higher-order functions such as perception, attention and decision-making, in both, humans and animals. There is growing evidence that the broad organization of these functions across the neocortex has been preserved across mammals [39], and that the micro-circuitry across brain regions is remarkably similar. The latter is due to the anatomical composition of the neocortex. At the macroscopic scale (cerebral), the neocortex is a layered tissue comprising three to six distinct cortical layers that differ significantly in thickness, but are similar in cellular composition. At the mesoscopic scale (cellular), the neocortex is thought to consist of 80% excitatory principal cells and 20% inhibitory interneurons. Each of these cell types can be further classified with respect to morphology (basket cells, Martinotti cells), spiking dynamics (fast or regular spiking) or molecular interactions (responsive to somatostatin, parvalbumin, serotonin) into many sub-types that are generally preserved across different cortical areas. At the microscopic scale (molecular), cells in the neocortex have receptors for a number of neuromodulators (cholinergic, noradrealinergic, dopaminergic, serotonergic) that are comparable across areas regarding the type of neuromodulators, despite the marked differences across functional areas [52]. Hence, these microcircuits comprise the core work-force of cortical computations in human and animal brains.

How do these microcircuits express a diverse range of cortical functions? It has been observed in humans and demonstrated in primates, that following lesions which eliminate a perceptual modality (e.g., visual, auditory or tactile) or a region of perceptual space (e.g., involving an arm or an eye), neurons that would process that sensory information would rapidly be recruited to support surrounding sensory processes that would subsequently become more relevance for the animal's actions [12,19,63,42,33,49]. In a similar fashion, other animal studies have shown that the implantation of a third eye in a

particular species of frogs results in the recruitment of a portion of neurons from the optic tectum. In particular, those neurons will engage in the processing of incoming inputs from the new eye [14]. In a similar way, deprivation of sensory input, such as vision, results in a compensation for this loss by using the visual neocortex for processing auditory stimuli, as shown in cats [63]. These findings highlight crucial characteristics of microcircuits in the neocortex and maybe rest of the brain. First, despite having remarkably diverse abilities, different animals use these cortical circuits to process similar types of information. Moreover, while different types of information would seem to require radically different processing schemes, the neocortex preserves relatively similar layered, cellular and molecular structures in spite of different functions they are supposed to perform. Finally, different studies have shown that any area of the neocortex can indeed flexibly process different kinds of information when exposed to them. This suggests that its unique structural characteristics might be more descriptive of the modality in which it operates rather than a pre-defined function. The picture that emerges of the neocortex, is not one built of rigid functionally independent areas, but rather a plastic tissue involved in information processing across a wide range of sensory and non-sensory modalities. In this case, similarities across cortical regions, rather than the differences, would be useful for understanding complex mechanisms of information-processing across the neocortex.

The aforementioned flexibility in processing sensory information requires specific mechanisms of plasticity and adaptation, most importantly, synaptic plasticity and inhibition respectively. Synaptic plasticity provides a way to tune the parameters of a system. Through inhibition, these parameters adapt to different input features, like dynamic range, gain or sampling frequency [32]. This is, synaptic plasticity and inhibition together adapt or construct models of incoming sensory inputs, which refer to models of the external world as sampled by the agent. While adapting or updating existing models would be equivalent to preserving existing features of sensory inputs with minor modifications, constructing new models would imply overriding previous model features in order to address a new sensory scenario. This can be thought of as a form of exploitation-exploration trade-off, where at times preservation of a model is preferable and otherwise rapid incorporation of new features benefits future actions. This suggests a third important circuit mechanism of cortical microcircuits that enables rapid switching between exploratory modes and exploitatory modes, namely, neuromodulation. Behaviorally, the role of neuromodulation in switching between exploratory and exploitatory behaviors has been studied in animals [13]. An example of an exploitatory behavior is reward seeking, whereas avoidance of an aversive stimulus, constitutes an exploratory behavior. Computationally, reinforcement learning mechanisms have been used to model exploitatory behaviors [21]. Less is known about computational mechanisms of involving avoidance of aversive stimuli.

In this article, we investigate the role of neuromodulation in learning at the level of cortical microcircuits. Neuromodulation in the brain, unsurprisingly, is released in situations related to uncertainty. Dopamine is released when re-

ward prediction errors are found [68], Acetylcholine (ACh) is associated with cues predictive of expected aversive events or expected changes in task conditions [78, 54, 28], while Norepinephrine is associated with unexpected changes in conditions of an experiment [17]. In this work, we will focus on understanding the role of ACh in signaling and learning from expected changes or expected uncertainty [2], as a potential candidate to drive sensory exploration. Neuromodulation can be used as a mechanism to switch between exploitatory and exploratory learning. Following this, a range of different neuromodulators potentially signal different epochs, which are relevant for different types of learning. An example is learning to obtain a reward through a specific action, or learning to avoid future punishment. Having different neuromodulators involved in these two scenarios suggests distinct learning mechanisms.

Compared to Bayesian approaches seeking to model computational mechanisms of predictive coding in the brain [61, 5], in this article, we take a much more biologically grounded route to understanding mechanisms of cortical computation. Our work culminates in a biophysical dynamical systems model of how cortical microcircuits learn to estimate uncertainty in sensory stimuli. With Bayesian models there is no clear consensus on how they can be realized in a biologically detailed way. Moreover, to the best of our knowledge, current Bayesian approaches do not address how uncertainty can be estimating within the model itself. In this sense, there still remains an explanatory gap between Bayesian models of predictive coding and biophysically-realistic neural dynamics. Instead, here we look for computations arising from the biophysics of cortical microcircuits.

The outline of this paper is as follows. We first review the literature on biophysics of learning from surprising or uncertain events, focusing on the interplay of inhibitory networks and acetylcholine. We then derive a neural model to estimate uncertainty in the sensory domain. Finally, we apply biologically plausible constraints to its dynamics and discuss the role of inhibition and neuromodulation in the signaling of and learning from uncertainty. We propose that specific interneuron sub-types play an essential role in capturing sensory uncertainty in the neocortex and that neuromodulation (particularly via ACh) provides a switching mechanism in cortical dynamics to favor exploration of sensory cues over exploitation of existing models by modulation of an intricate disinhibitory network.

2 Biophysics of Learning and Uncertainty

The neural processes underlying learning mechanisms have been studied for decades. Through conditioning paradigms, we can probe different brain areas to understand their role in associating perception to action. One of the best-studied cases is eye-blink conditioning, where a rabbit receives an air puff in the eye preceded by an auditory stimulus. These experiments demonstrate a slow learning process that, after many exposures to the conditioning stimuli, translate into a well-timed eye-blink which avoids the aversive air puff.

In 1990, Bakin and Weinberger discovered that during this paradigm, the cerebellum-mediated learning of an anticipatory eye-blink response would be preceded by a much faster process in which the auditory conditioning stimulus was learned and associated to a startle or fear-related response [4]. This was understood as a learning process that was actually separated into two phases: a slow, reflex-specific learning phase relying on the cerebellum and driving a well-timed motor response and a fast, non-specific learning phase taking place in the neocortex in order to re-shape sensory representations to facilitate (though not exclusively) cerebellar learning [41,47]. The fact that two learning processes act in a parallel and dissociated but complementary fashion suggests that they might serve two different functional purposes. In this case, in particular, we could interpret that before learning a specific behavioral response, we first need to identify what cues are potentially relevant for succeeding in that task. The two-phase modeling of conditioning [31,58] proposes that the differentiation between separating learning in these two phases allows to rapidly acquire sensory evidence that will later be essential to perform effective actions. Therefore, a form of exploratory learning to capture new task-dependent contingencies in the sensory space would be required in some tasks to facilitate a later, slower exploitation phase to refine and perfect behavior and performance.

It is not surprising that further studies showed that this faster, non-specific form of learning was mediated by a specific neuromodulator, in this case, acetylcholine (ACh) [80,8,79]. Cortical ACh stems from the topologically organized cholinergic neurons in the Nucleus Basalis of Meynert (NBM). This neuromodulator was found to produce long-lasting synaptic changes in the auditory neocortex when the conditioning Stimulus (CS) was a sound [11], while it has been shown to actually refine cortical receptive fields of different sensory modalities through extensive evidence [6,20,67,83,62,50]. Many other neuromodulators have been shown to drive some form of learning: Dopamine [55], Noradrenaline [27] or even Oxcitocin [46] among others. What makes ACh different?

Cholinergic neurons in the NBM are topologically organized with different cortical areas. Higher levels of cortical acetylcholine have been found to be correlated with a number of brain processes, including increases in arousal, sleep-wake cycles, sensory information processing (visual, auditory, olfactory), memory, attention and other cognitive processes [48,1,53,29,24,56,51]. The improved performance in detection tasks becomes particularly relevant in this case. Sensory detection paradigms are usually composed of a noisy context or mask that is presented at every trial and a detectable sensory pattern that is hidden in that context. By regulating the ratio between noise and signal (SNR) experimenters have control on the subjects' performance, what allows to understand what elements influence such performance. Hence, it is obvious that if cortical ACh plays a role in regulating performance [28], it must have a way to either regulate the SNR in the neocortex or improve learning in situations of increased uncertainty or both, as proposed in [59]. While the possible interpretations of this observations are vast, it is undeniable that

ACh is driven by a mechanism that detects or predicts uncertainty and that its modulation of cortical dynamics improves SNR, signal detection and influence gating of information. Through what mechanisms the brain can learn or evaluate the uncertainty of sensory inputs or other internal signals is left unexplained. Moreover, which are the particular differences between ACh-mediated learning and learning driven by other neuromodulators is still unclear.

2.1 Biology of Acetylcholine and Inhibitory Interneurons

Cholinergic neurons in the nucleus Basalis of Myenert (NBM) are the main source of cortical Acetylcholine. ACh as a neuromodulator provides diffuse modulation and a broad innervation that reaches the vast majority of cortical areas of the brain. In contrast, glutamate and γ-Aminobutyric acid are local and specific to one or few synapses. This property allows it to affect globally the dynamics of the cortical circuitry instead of merely transmitting a piece of information. Acetylcholine interacts with both excitatory and inhibitory interneurons via two types of receptors: muscarinic (mAChR) and nicotinic ACh receptors (nAChR). While both receptors depolarize the neurons in the presence of acetylcholine, metabotropic mAChR have slower and long-lasting effects while ionotropic nAChR rapidly open calcium channels[26, 43]. Nonetheless, their overall impact in the network is markedly different. Muscarinic receptor agonists in the rat neocortex result in a global excitatory effect of both excitatory and inhibitory neurons [36]. In contrast, agonists of nicotinic receptors have a disinhibitory influence on the neural substrate, by preferentially targeting inhibitory neurons that hyperpolarize other inhibitory interneurons [81]. Altogether, acetylcholine affects the network level that depends on the balance and distribution of these two opposing cholinergic receptors [40]. The functional use of this segregation of receptors has not been identified yet, despite it seems that its influence on the inhibitory interneurons could be key for its understanding.

A fairly recent study showed that the cortical inhibitory interneuron network is composed of three types of interneurons distinct on their binding to three mutually exclusive proteins [65, 71]. These proteins are Parvalbumin, Somatostatin, and 5HT3A (Serotonin). In turn, these three molecularly distinct types of interneurons are also morphologically, dynamically and topologically distinct.

Parvalbumin responsive inhibitory interneurons (PVi+) tend to be identified as fast-spiking interneurons or basket cells [37, 70]. These cells typically have relatively small dendritic arbors that target one or few nearby cells [66]. These suggest a local reach of their inhibitory effect and has been associated with, e.g. feed-forward inhibition [69]. Somatostatin responsive inhibitory interneurons (SOMi+) have larger dendritic arbors and tend to have synapses across cortical layers. In this case, somatostatin interneurons seem to play a critical role in more external cortical layers. These layers are the place where most top-down and bottom-up interactions take place [45]. The dynamics of

these interneurons have been observed to be directly affected by locomotion, reducing surround suppression during movement via the external layers of V1 [3]. This modulation is known to be dependent on a disinhibitory network by VasoIntestinal Peptide interneurons (VIPi+) [25]. These observations suggest that they could play two crucial roles. First, extensive dendritic arbors would indicate a role on a more global form of gain control that integrates information from further areas. This is supported by the observations in reduced surround suppression driven by locomotion [3]. At the same time, their locations and interactions could suggest a role in receiving top-down information that could be more or less task-relevant, in the form of an attention modulating signal.

Finally, Serotonin responsive interneurons are usually co-selective with VasoIntestinal Peptide-selective interneurons. These reside principally in the upper layers of the neocortex, where they are almost the only neuron type. VIP cells play an essential role in cross-areal interactions. These cells are the ones signaling locomotion-related information to SOMi+ cells, being the driver of a disinhibitory effect [82,35] in cortical microcircuits. VIPi+ are involved in signaling top-down contextual information in order to gate [30] or modulate [3] sensory information in sensory cortical areas, while modulating executive control in prefrontal areas of the Neocortex [34]. Strikingly, the results of this last study are behaviorally equivalent to those in [28]: while optogenetic stimulation of VIP cells favors Go over No Go responses in a memory-guided behavior task in the first, stimulation of cholinergic neurons in the NBM promoted false alarms in a detection task in the last, suggesting a significant overlap in the electrophysiology of the stimulation.

Altogether, this evidence draws a cortical motif of inhibitory interactions that have seems to be essential for sensing, learning and acting under some degrees of uncertainty.

3 A Microcircuit Model for Learning Stimulus Uncertainty

3.1 Biophysical Description

While the interactions between the interneuron subtypes mentioned above could be reasonably variable, there is increasing evidence for the existence of a circuit motif [73,3,9]. This microcircuit is composed of an intricate network of primary excitatory neurons interconnected with neighboring PVi+ and SOMi+ interneurons. VIPi+ cells, in turn, mostly reside in the upper layers of the neocortex, mediating top-down interactions with SOMi+, by the influence of signals from other cortical areas. The evidence shows that manipulating some of this interneurons affects performance under uncertain environments, suggesting a latent role of this network in capturing and modulating the perceived uncertainty in the world.

While the plasticity in inhibitory networks is thought to hold the key to understanding perception and perceptual learning [23] via the regulation of the so-called excitatory-inhibitory balance, the debate is still open to identify what

the particular inhibitory and disinhibitory interactions within the Neocortex are. We here propose to start from a computational perspective and derive a model that captures the uncertainty of some input signal. Understanding the role of the different elements of such a model might help unveil the functions underlying cortical inhibition.

3.2 Mathematical Description

3.2.1 Derivation

In order to derive a model that captures sensory uncertainty we define a simple perceptual problem. We consider the task of estimating the light intensity on a sensor in the retina, assuming that the sensor draws its values from a random normal distribution X with mean μ_x and standard deviation σ_x, the probability of a specific observation becomes:

$$P(X = x \mid \mu_x, \sigma_x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x - \mu_x)^2}{2\sigma_x^2}} \tag{1}$$

We can rephrase this problem as the problem of finding the estimates of μ_x and σ_x that maximize $P(X = x \mid \mu_x, \sigma_x)$, or equivalently, maximizing its logarithm:

$$L = ln(P) = ln(\frac{1}{\sqrt{2\pi\sigma_x^2}}) - \frac{(x - \mu_x)^2}{2\sigma_x^2} \tag{2}$$

Considering Θ_x as the set of parameters μ_x, σ_x we want to find the estimates of Θ^*:

$$\Theta^* = \underset{\Theta_x}{argmax}(L) \tag{3}$$

In order to find the best estimate that maximizes the loss function L we can use gradient ascent. Gradient ascent is an iterative process that allows us to find the parameters (Θ^*) that maximize a function (L) by finding analytically the slope of L and iteratively climbing through the steepest path. Assuming that μ_x and σ_x are independent variables, we obtain the following set of dynamical equations that represent our system from the partial derivatives that define the slopes of L:

$$\nabla_{\mu_x} L = \frac{\partial L}{\partial \mu_x} = \frac{2(x - \mu)}{2\Sigma_x} = e = \frac{d\mu_x}{dt} \tag{4}$$

$$\nabla_{\Sigma_x} L = \frac{\partial L}{\partial \Sigma} = \frac{1}{2}(-\frac{1}{\Sigma} + \frac{(x - \mu)^2}{\Sigma^2}) = \frac{1}{2}(\frac{\epsilon e - 1}{\Sigma}) = \frac{d\Sigma}{dt} \tag{5}$$

Where $\Sigma = \sigma^2$. These derived dynamic equations require two extra variables to estimate the relative and absolute errors e and ϵ:

$$\frac{de}{dt} = (x - \mu) - \epsilon = (x - \mu) - \Sigma e \tag{6}$$

$$\frac{d\epsilon}{dt} = e\Sigma - \epsilon \tag{7}$$

which converge to the steady states of the system:

$$\frac{de}{dt} = 0 = (x - \mu) - \Sigma e \implies e_0 = \frac{x - \mu}{\Sigma} \tag{8}$$

$$\frac{d\epsilon}{dt} = 0 = e\Sigma - \epsilon \implies \epsilon_0 = e\Sigma \tag{9}$$

Namely the relative and absolute errors.

Representing this system as a network with nodes (neurons) and vertices (synapses) (see Fig. 2) we can identify that our system has a few important properties. First of all, the changes in our nodes depend (by definition) exclusively on inputs in the form of weighted sums. Moreover, the synapses or vertices also change using purely local information, i.e., the changes in a vertex depend solely on the neighboring nodes or itself. Finally, the synaptic changes derived using gradient ascent are equivalent to a Hebbian learning rule with or without homeostatic decay. Altogether, this network can be interpreted as a neural network where each node represents the firing rate of a particular neuron. While, analytically, this system converges to the desired solution (a good approximation of μ_x and σ_x as illustrated in Fig. 2 B and C), the fact that the error nodes can switch sign depending on the input and its prediction is not desirable when aiming to compare with a biologically plausible system. To address this concern, we constrain our model to have exclusively excitatory and exclusively inhibitory neurons, by using linear rectification:

$$f(z) = \begin{cases} z, & \text{if } z \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Because this will only consider one side of the error ($e_1 = f(x - wy)$), we need to mirror the circuit in order to compute both halves: e_1 and $e_2 = f(y - vx)$. This new circuit is illustrated in figure 2 D as the biologically constrained version of the circuit in figure 2 A.

4 Results

In order to see if the biologically constrained version of the network behaves as expected, we perform a series of tests. First, we check that the network parameters successfully estimate the mean and standard deviation of the network. Figure 2 shows that effectively, the network converges to a good estimation of the mean input as shown by the zero error obtained after training for a range of randomly selected initial conditions and parameters. Nonetheless, the variance is only partially estimated for large instances of σ_x. In the cases where σ_x is already small, the network doesn't fully converge, as illustrated in the first part of figure 3 D. To understand this phenomenon, we analyze the stability of our network. Studying the distribution of poles of the synamical system in the Real-Imaginary space allows us to analyze the stability in the sense of Lyapunov, as shown in figure 3 A. We can observe that the system is intrinsically

oscillatory (not surprising to think of a neural network given [44]) due to the two Imaginary poles close to zero in the real axis. To move the complex poles in along the Real axis, we modulate inversely Σ and e. This modulation results in a continuous transformation of the system from unstable or marginally stable dynamics to a fully stable regime (poles with exclusively negative Real), in which the system converges faster the further the negative poles are from 0. This modulation facilitates the partition of this dynamic space:

- one in which the relative error is too big and the system diverges (not shown), corresponding to the states with a pole with a positive Real part.
- a marginal zone where the relative error is large despite the variance too small to account for it, that leads to oscillatory dynamics with extremely slow converging or diverging dynamics
- a stable zone characterized by errors smaller than the variance or, equivalently, with errors close to zero and variance properly estimated

One can notice that, while the last case is the desired one, more stable when the error has been already minimized, the second corresponds to the case where relative errors are relatively larger than what the variance can account for. The last can be seen as the case when an outlier is detected or when the model is no longer able to capture the statistics of the input. Our model reacts to that by entering an oscillatory mode where the absolute, objective error (measured as $x - wy$) can't be reduced by the system and needs some external input to break this situation. Given that this abnormal state comes from the relative difference between Σ and e, we can either inject a transient increase in the conductivity of the synapse Σ or, equivalently, to inject current directly in the neuron e that enables the estimation of the variance. Figure 3 D, E show how this injection rapidly brings the state to the stable region that quickly converges where otherwise the dynamics would have converged much slower, if at all.

We can understand this current injection as a top-down bias to the estimation of uncertainty. In our model, while our y inhibitory cells feed forward information that can be used for prediction of x excitatory cells, ϵ inhibitory cells are more locally connected to estimate the variance of the system. In cortical anatomy, we can identify three main types of interneurons, PV, SOM and VIP. PV cells are identified as feed-forward interneurons that convey sensory information to regulate the excitatory input injected into the network (interpreted in our model as the relative error e). The second most noticeable type of inhibitory interneurons (SOM) are involved in gain control and are modulated by locomotion (in V1), acetylcholine (in other sensory areas) or other top-down signals to regulate the incorporation of new information to the network. We, therefore, propose that cholinergic modulation, via the intervention of VIP interneurons, transiently reduces the activity in SOM cells, which in turn brings the system temporarily to a non-stable regime that allows the exploration of sensory alternatives before the re-shaping of internal models when the effect fades off.

In summary, our computational model and its analysis have provided us with a minimal microcircuit to perform three main functions:

– To compute the uncertainty of an internal or external piece of information given its dynamics over time, in a biologically constrained neural network
– To use an estimate of uncertainty to modulate learning, thus improving learning speed as a trade-off with accuracy
– To provide a mechanism to switch between two regimes, one to rapidly learn when the system is unable to converge to a solution, and the opposite, to bring the system in an oscillatory state that can be beneficial for sensory exploration.

5 Conclusions and Discussion

The role of excitatory-inhibitory interactions has been debated thoroughly in the past. Nonetheless, this discussion has mostly considered inhibition as a homogeneous block. This block is sometimes responsible for performing some form of winner-take-all filtering or maintaining the balance in the output firing rates of excitatory cells. Our model in this article proposes that, if a network of neurons has to perform some form of probabilistic computation, two types of interneurons, with different functions and connections, need to be considered: The first type is required to compute the mean of the input, while the second is necessary to estimate its variance. The first can be seen as a gain control system, aiming at balancing excitatory input to detect inconsistencies. In fact, the reduced circuit that estimates only the mean of the activity of the input data matches models of excitatory-inhibitory balance, winner-take-all and decision making [72], even producing receptive fields similar to those observed in different sensory regions of the neocortex [61]. The second type of interneurons extends the first one by providing a mechanism to regulate learning through uncertainty. On one hand, it provides a faster convergence mechanism (not shown in this paper) by learning faster when uncertainty is increased. On the other hand, it provides a mechanism to regulate learning in a top-down fashion. According to our model, this means detection of incongruency in the world raises the brain's estimate of uncertainty, which in turn increases learning rates transiently without biasing the actual content to be learned. Our model extends current models of cortical function by proposing a role for different inhibitory neuron types.

Biophysical experiments have shown that two out of three biochemically segregated inhibitory populations are composed of around 60% to 90% of inhibition in most cortical layers [65,71]. While fast-spiking interneurons are typically associated with a gain-control kind of function, SOMi+, regular-spiking interneurons are found to be involved in a range of functions that require top-down or cross-area information, like surround-suppression inhibition during locomotion in V1 [3]. These functions can be compared with those of the neural units presented in our microcircuit model, where PVi+ corresponds to interneurons responsible for learning the mean of activity of the input and

SOMi+ as the ones involved in estimating variance of input. External input to the second type of interneurons would then modulate both, the perceived uncertainty and consequently the rate of synaptic change in the microcircuit.

To summarize, our model results in the following interpretation of cortical microcircuits: the recurrent microcircuitry of excitatory cells with two inhibitory populations inherently computes prediction errors and their uncertainty (through the variance). This can be seen as follows. Excitatory input from the thalamus comes to the principal excitatory cells of the neocortex. These neurons can be understood to be computing some form of prediction error via the inhibitory effect of parvalbumin cells. While PVi+ in different layers of the neocortex can be receiving excitatory input either from thalamocortical connections or from intracortical excitatory cells, they would be the candidates for providing predictions of the thalamocortical input. Finally, the broad dendritic range and long term dynamics of SOMi+ allows them to integrate information effectively over both space and time. This integration of information would mediate the estimation of uncertainty on the PVi+ predictions. SOMi+ inhibition would be stronger for greater estimated uncertainty, in turn modulating the perceived errors signaled by PCs. If this were the case, our model would suggest that top-down inhibition of the same SOMi+ (via VIPi+) should increase the sensitivity of the errors, re-enabling learning of uncertainty by switching to a more exploitatory mode. This disinhibitory role could be played by both, the neuromodulatory action of ACh or top-down modulation of layer I VIPi+ from other cortical areas.

While the interpretation of the brain performing some form of predictive coding has long been posed [61], we go one step further by proposing an explicit hypothesis of how a neural circuit could implement predictions in a dynamical system and extending previous models with an additional interneuron subtype involved in monitoring uncertainty. By targeting explicitly some interneuron subtypes, we should be able to identify if their plastic changes correspond with those we propose.

Our model presents a new lens to understand the function of cortical microcircuits and neuromodulation. While the most common neuromodulators (ACh, DA, Se, NE) have been studied for decades, little is known about how they influence our behavior. Looking at cortical function through our microcircuit model allows us to bridge the gap between behavior and biophysics by presenting inhibitory interactions as a predictive process. Taking ACh as an example, the links between the role of ACh in perception and decision making at the behavioral level, and the interactions of ACh with different elements of cortical substrates is still unclear. Our model proposes that a phasic decrease on cortical SOMi+ activity (probably driven by top-down VIP control, induced via ACh (see [59] for details), functionally amounts to a decrease in ϵ-driven regulation of the relative error, consequently signaling a smaller variance which translates into larger response to outliers, triggering false alarm responses while further promoting learning.

Finally, our model provides a mechanism to switch between operational modes. Modulating uncertainty can be seen as driving exploration or exploita-

tion by modulating how large an error will be perceived, relative to its expected variance. Going back to the examples of conditioning, the dopamine-dependent addiction circuit and the acetylcholine-dependent fear circuit can be seen as the axis between exploitation and exploration. While the first is involved in the repetition of particular behaviors and the formation of habits, ACh seems to be signaled to reshape and explore new potentially useful cues in the sensory domain. Our model suggests that, while exploitation and exploration could be seen as two ends of the same axis, two separated circuits enable each of the operation modes. Moreover, in a detection task with white noise masking a signal that needs to be detected, our circuit would aim at filtering the expected noise in search of a stable cue that helps at detecting the target. Normal levels of ACh should regulate the optimal threshold of noise cancellation to optimize perception. Increased ACh levels would also increase the probability to identify the random noise as an outlier, or equivalently, the signal. This can be seen as a form of exploration by reducing the tolerance to noise and enabling more erratic behavior under the presence of predicted uncertainty. Misses, on the contrary, induced in [3] by reducing ACh levels, could then be associated to a more conservative, exploitatory mode in which a more robust performance is preferred, therefore fully inhibiting the predicted noise.

The phenomenology behind Alzheimer's Disease, characterized by decreased levels of ACh due to loss of cholinergic innervation [15, 16] and Basal cholinergic neurons[76, 77], can be understood as an abnormally decreased ability to compute prediction errors, translating into a reduced ability to learn from those and therefore form new memories. Alternatively, Schizophrenia, correlated with generally with reduced amount of muscarinic and nicotinic acetylcholine receptors [60, 22], indirectly increasing the levels of extracellular ACh, can be understood as a decreased ability to gate unreliable information [57] , as it would be supported by the reported increases and reductions in sensory detection tasks, shown both in animals and humans [10, 74, 75].

Nonetheless, our model leaves some open questions. What drives the top-down modulation of SOMi+ or the modulation via ACh itself? These are still unanswered. While some papers suggest possible pathways, the picture is still fragmented and incomplete, and we are not much further from Yu, and Dayan's [2] hypothesis of ACh signaling expected uncertainty, without knowing the mechanism for representing that expectation. Moreover, the complexity of the layered structure of the Neocortex is not fully captured by our model. While our model only aims at explaining the role of the microcircuit *per se*, understanding how multiple instances of such microcircuits could be connected, forming different sorts of motifs could allow us to complete the picture of the predictive brain. Moreover, this motif is not found exclusively in the Neocortex. Evidence suggests that other ACh targeted regions with excitatory principal cells like the hippocampus, the amygdala and probably the olfactory bulb have the same kind of inhibitory microcircuits and equivalent interactions with ACh [7, 38, 64]. Further work should be done to identify how other neuromodulators interact with these microcircuits and affect behavior and cognitive function.

References

1. Aloisi, A., Casamenti, F., Scali, C., Pepeu, G., Carli, G.: Effects of novelty, pain and stress on hippocampal extracellular acetylcholine levels in male rats. Brain research **748**(1-2), 219–226 (1997)
2. Angela, J.Y., Dayan, P.: Uncertainty, neuromodulation, and attention. Neuron **46**(4), 681–692 (2005)
3. Ayaz, A., Saleem, A.B., Schölvinck, M.L., Carandini, M.: Locomotion controls spatial integration in mouse visual cortex. Current Biology **23**(10), 890–894 (2013)
4. Bakin, J.S., Weinberger, N.M.: Classical conditioning induces cs-specific receptive field plasticity in the auditory cortex of the guinea pig. Brain research **536**(1-2), 271–286 (1990)
5. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J.: Canonical microcircuits for predictive coding. Neuron **76**(4), 695–711 (2012)
6. Bear, M.F., Singer, W.: Modulation of visual cortical plasticity by acetylcholine and noradrenaline. Nature **320**(6058), 172 (1986)
7. Bendahmane, M., Ogg, M.C., Ennis, M., Fletcher, M.L.: Increased olfactory bulb acetylcholine bi-directionally modulates glomerular odor sensitivity. Scientific reports **6**, 25808 (2016)
8. Berry, S.D., Thompson, R.F.: Medial septal lesions retard classical conditioning of the nicitating membrane response in rabbits. Science **205**(4402), 209–211 (1979)
9. Blackwell, J.M., Geffen, M.N.: Progress and challenges for understanding the function of cortical microcircuits in auditory processing. Nature communications **8**(1), 2165 (2017)
10. Brown, K., Warburton, D.M.: Attenuation of stimulus sensitivity by scopolamine. Psychonomic Science **22**(5), 297–298 (1971)
11. Butt, A.E., Chavez, C.M., Flesher, M.M., Kinney-Hurd, B.L., Araujo, G.C., Miasnikov, A.A., Weinberger, N.M.: Association learning-dependent increases in acetylcholine release in the rat auditory cortex during auditory classical conditioning. Neurobiology of learning and memory **92**(3), 400–409 (2009)
12. Chino, Y., Kaas, J., Smith III, E., Langston, A., Cheng, H.: Rapid reorganization of cortical maps in adult cats following restricted deafferentation in retina. Vision research **32**(5), 789–796 (1992)
13. Cohen, J.D., McClure, S.M., Angela, J.Y.: Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. Philosophical Transactions of the Royal Society of London B: Biological Sciences **362**(1481), 933–942 (2007)
14. Constantine-Paton, M., Law, M.I.: Eye-specific termination bands in tecta of three-eyed frogs. Science **202**(4368), 639–641 (1978)
15. Coyle, J.T., Price, D.L., Delong, M.R.: Alzheimer's disease: a disorder of cortical cholinergic innervation. Science **219**(4589), 1184–1190 (1983)
16. Davies, P., Maloney, A.: Selective loss of central cholinergic neurons in alzheimer's disease. The Lancet **308**(8000), 1403 (1976)
17. Devauges, V., Sara, S.J.: Activation of the noradrenergic system facilitates an attentional shift in the rat. Behavioural brain research **39**(1), 19–28 (1990)
18. Douglas, R.J., Martin, K.: A functional microcircuit for cat visual cortex. The Journal of physiology **440**(1), 735–769 (1991)
19. Elbert, T., Flor, H., Birbaumer, N., Knecht, S., Hampson, S., Larbig, W.: Extensive reorganization of the somatosensory cortex in adult humans after nervous system injury. Neuroreport **5**(18), 2593–2597 (1994)
20. Everitt, B.J., Robbins, T.W.: Central cholinergic systems and cognition. Annual review of psychology **48**(1), 649–684 (1997)
21. Everitt, B.J., Robbins, T.W.: Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nature neuroscience **8**(11), 1481 (2005)
22. Freedman, R., Hall, M., Adler, L.E., Leonard, S.: Evidence in postmortem brain tissue for decreased numbers of hippocampal nicotinic receptors in schizophrenia. Biological psychiatry **38**(1), 22–33 (1995)
23. Froemke, R.C.: Plasticity of cortical excitatory-inhibitory balance. Annual review of neuroscience **38**, 195–219 (2015)

24. Froemke, R.C., Carcea, I., Barker, A.J., Yuan, K., Seybold, B.A., Martins, A.R.O., Zaika, N., Bernstein, H., Wachs, M., Levis, P.A., et al.: Long-term modification of cortical synapses improves sensory perception. Nature neuroscience **16**(1), 79 (2013)
25. Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., Stryker, M.P.: A cortical circuit for gain control by behavioral state. Cell **156**(6), 1139–1152 (2014)
26. Gilsbach, R., Hein, L.: Presynaptic metabotropic receptors for acetylcholine and adrenaline/noradrenaline. In: Pharmacology of neurotransmitter release, pp. 261–288. Springer (2008)
27. Glennon, E., Carcea, I., Martins, A.R.O., Multani, J., Shehu, I., Svirsky, M.A., Froemke, R.C.: Locus coeruleus activation accelerates perceptual learning. Brain research (2018)
28. Gritton, H.J., Howe, W.M., Mallory, C.S., Hetrick, V.L., Berke, J.D., Sarter, M.: Cortical cholinergic signaling controls the detection of cues. Proceedings of the National Academy of Sciences **113**(8), E1089–E1097 (2016)
29. Hasselmo, M.E.: Neuromodulation: acetylcholine and memory consolidation. Trends in cognitive sciences **3**(9), 351–359 (1999)
30. Hattori, R., Kuchibhotla, K.V., Froemke, R.C., Komiyama, T.: Functions and dysfunctions of neocortical inhibitory neuron subtypes. Nature neuroscience **20**(9), 1199 (2017)
31. Inderbitzin, M., Herreros-Alonso, I., Verschure, P.F.: An integrated computational model of the two phase theory of classical conditioning. In: Neural Networks (IJCNN), The 2010 International Joint Conference on, pp. 1–8. IEEE (2010)
32. Isaacson, J.S., Scanziani, M.: How inhibition shapes cortical activity. Neuron **72**(2), 231–243 (2011)
33. Kaas, J.H., Krubitzer, L.A., Chino, Y.M., Langston, A.L., Polley, E.H., Blair, N.: Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. Science **248**(4952), 229–231 (1990)
34. Kamigaki, T.: Prefrontal circuit organization for executive control. Neuroscience research (2018)
35. Karnani, M.M., Jackson, J., Ayzenshtat, I., Sichani, A.H., Manoocheri, K., Kim, S., Yuste, R.: Opening holes in the blanket of inhibition: localized lateral disinhibition by vip interneurons. Journal of Neuroscience **36**(12), 3471–3480 (2016)
36. Kawaguchi, Y.: Selective cholinergic modulation of cortical gabaergic cell subtypes. Journal of neurophysiology **78**(3), 1743–1747 (1997)
37. Kawaguchi, Y., Kubota, Y.: Neurochemical features and synaptic connections of large physiologically-identified gabaergic cells in the rat frontal cortex. Neuroscience **85**(3), 677–701 (1998)
38. Krabbe, S., Paradiso, E., d'Aquin, S., Bitterman, Y., Xu, C., Yonehara, K., Markovic, M., Gründemann, J., Ferraguti, F., Luthi, A.: Adaptive disinhibitory gating by vip interneurons permits associative learning. bioRxiv p. 443614 (2018)
39. Krubitzer, L.: The organization of neocortex in mammals: are species differences really so different? Trends in neurosciences **18**(9), 408–417 (1995)
40. Lawrence, J.J.: Cholinergic control of gaba release: emerging parallels between neocortex and hippocampus. Trends in neurosciences **31**(7), 317–327 (2008)
41. Lennartz, R.C., Weinberger, N.M.: Analysis of response systems in pavlovian conditioning reveals rapidly versus slowly acquired conditioned responses: Support for two factors, implications for behavior and neurobiology. Psychobiology **20**(2), 93–119 (1992)
42. Lomber, S.G., Meredith, M.A., Kral, A.: Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. Nature neuroscience **13**(11), 1421 (2010)
43. MacDermott, A.B., Role, L.W., Siegelbaum, S.A.: Presynaptic ionotropic receptors and the control of transmitter release. Annual review of neuroscience **22**(1), 443–485 (1999)
44. Mann, E.O., Paulsen, O.: Role of gabaergic inhibition in hippocampal network oscillations. Trends in neurosciences **30**(7), 343–349 (2007)
45. Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al.: Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. Journal of Comparative Neurology **522**(1), 225–259 (2014)
46. Marlin, B.J., Froemke, R.C.: O xytocin modulation of neural circuits for social behavior. Developmental neurobiology **77**(2), 169–189 (2017)

47. Martin, I., Levey, A.: The genesis of the classical conditioned response, vol. 8. Pergamon (1969)
48. McCormick, D.A., Bal, T.: Sleep and arousal: thalamocortical mechanisms. Annual review of neuroscience **20**(1), 185–215 (1997)
49. Merzenich, M.M., Nelson, R.J., Stryker, M.P., Cynader, M.S., Schoppmann, A., Zook, J.M.: Somatosensory cortical map changes following digit amputation in adult monkeys. Journal of comparative Neurology **224**(4), 591–605 (1984)
50. Metherate, R., Cox, C.L., Ashe, J.H.: Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine. Journal of Neuroscience **12**(12), 4701–4711 (1992)
51. Ogg, M.C., Ross, J.M., Bendahmane, M., Fletcher, M.L.: Olfactory bulb acetylcholine release dishabituates odor responses and reinstates odor investigation. Nature communications **9**(1), 1868 (2018)
52. Palomero-Gallagher, N., Zilles, K.: Cortical layers: Cyto-, myelo-, receptor-and synaptic architecture in human cortical areas. Neuroimage (2017)
53. Pepeu, G., Giovannini, M.G.: Changes in acetylcholine extracellular levels during cognitive processes. Learning & memory **11**(1), 21–27 (2004)
54. Phillips, J.M., McAlonan, K., Robb, W.G., Brown, V.J.: Cholinergic neurotransmission influences covert orientation of visuospatial attention in the rat. Psychopharmacology **150**(1), 112–116 (2000)
55. Pignatelli, M., Bonci, A.: Role of dopamine neurons in reward and aversion: a synaptic plasticity perspective. Neuron **86**(5), 1145–1157 (2015)
56. Pinto, L., Goard, M.J., Estandian, D., Xu, M., Kwan, A.C., Lee, S.H., Harrison, T.C., Feng, G., Dan, Y.: Fast modulation of visual perception by basal forebrain cholinergic neurons. Nature neuroscience **16**(12), 1857 (2013)
57. Potasiewicz, A., Nikiforuk, A., Hołuj, M., Popik, P.: Stimulation of nicotinic acetylcholine alpha7 receptors rescue schizophrenia-like cognitive impairments in rats. Journal of Psychopharmacology **31**(2), 260–271 (2017)
58. Puigbò, J.Y., Gonzalez-Ballester, M., Verschure, P.F.: Modeling the neural substrates of learning through conditioning: A two-phased model. IBM Journal of Research and Development **61**(2/3), 9–1 (2017)
59. Puigbò, J.Y., Maffei, G., Herreros, I., Ceresa, M., Ballester, M.G., Verschure, P.: Cholinergic behavior state-dependent mechanisms of neocortical gain control: a neurocomputational study. Molecular neurobiology **55**(1), 249–257 (2018)
60. Raedler, T.J., Knable, M.B., Jones, D.W., Urbina, R.A., Gorey, J.G., Lee, K.S., Egan, M.F., Coppola, R., Weinberger, D.R.: In vivo determination of muscarinic acetylcholine receptor availability in schizophrenia. American Journal of Psychiatry **160**(1), 118–127 (2003)
61. Rao, R.P., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience **2**(1), 79 (1999)
62. Rasmusson, D.: The role of acetylcholine in cortical synaptic plasticity. Behavioural brain research **115**(2), 205–218 (2000)
63. Rauschecker, J.P.: Compensatory plasticity and sensory substitution in the cerebral cortex. Trends in neurosciences **18**(1), 36–43 (1995)
64. Rhomberg, T., Rovira-Esteban, L., Vikór, A., Paradiso, E., Kremser, C., Nagy-Pál, P., Papp, O.I., Tasan, R., Erdélyi, F., Szabó, G., et al.: Vasoactive intestinal polypeptide-immunoreactive interneurons within circuits of the mouse basolateral amygdala. Journal of Neuroscience **38**(31), 6983–7003 (2018)
65. Rudy, B., Fishell, G., Lee, S., Hjerling-Leffler, J.: Three groups of interneurons account for nearly 100% of neocortical gabaergic neurons. Developmental neurobiology **71**(1), 45–61 (2011)
66. Runyan, C.A., Schummers, J., Van Wart, A., Kuhlman, S.J., Wilson, N.R., Huang, Z.J., Sur, M.: Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. Neuron **67**(5), 847–857 (2010)
67. Sarter, M., Hasselmo, M.E., Bruno, J.P., Givens, B.: Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. Brain Research Reviews **48**(1), 98–111 (2005)
68. Schultz, W., Dayan, P., Montague, P.R.: A neural substrate of prediction and reward. Science **275**(5306), 1593–1599 (1997)

69. Swadlow, H.A.: Fast-spike interneurons and feedforward inhibition in awake sensory neocortex. Cerebral Cortex **13**(1), 25–32 (2003)
70. Toledo-Rodriguez, M., Blumenfeld, B., Wu, C., Luo, J., Attali, B., Goodman, P., Markram, H.: Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. Cerebral Cortex **14**(12), 1310–1327 (2004)
71. Tremblay, R., Lee, S., Rudy, B.: Gabaergic interneurons in the neocortex: from cellular properties to circuits. Neuron **91**(2), 260–292 (2016)
72. Wang, X.J.: Neural dynamics and circuit mechanisms of decision-making. Current opinion in neurobiology **22**(6), 1039–1046 (2012)
73. Wang, X.J., Yang, G.R.: A disinhibitory circuit motif and flexible information routing in the brain. Current opinion in neurobiology **49**, 75–83 (2018)
74. Warburton, D., Brown, K.: The facilitation of discrimination performance by physostigmine sulphate. Psychopharmacologia **27**(3), 275–284 (1972)
75. Wesnes, K., Warburton, D.M.: Effects of scopolamine and nicotine on human rapid information processing performance. Psychopharmacology **82**(3), 147–150 (1984)
76. Whitehouse, P.J., Price, D.L., Clark, A.W., Coyle, J.T., DeLong, M.R.: Alzheimer disease: evidence for selective loss of cholinergic neurons in the nucleus basalis. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society **10**(2), 122–126 (1981)
77. Whitehouse, P.J., Price, D.L., Struble, R.G., Clark, A.W., Coyle, J.T., Delon, M.R.: Alzheimer's disease and senile dementia: loss of neurons in the basal forebrain. Science **215**(4537), 1237–1239 (1982)
78. Witte, E., Davidson, M., Marrocco, R.: Effects of altering brain cholinergic activity on covert orienting of attention: comparison of monkey and human performance. Psychopharmacology **132**(4), 324–334 (1997)
79. Woodruff-Pak, D.S., Li, Y.T., Kazmi, A., Kem, W.R.: Nicotinic cholinergic system involvement in eyeblink classical conditioning in rabbits. Behavioral neuroscience **108**(3), 486 (1994)
80. Woodruff-Pak, D.S., Li, Y.T., Kem, W.R.: A nicotinic agonist (gts-21), eyeblink classical conditioning and nicotinic receptor binding in rabbit brain. Brain research **645**(1-2), 309–317 (1994)
81. Yi, F., Ball, J., Stoll, K.E., Satpute, V.C., Mitchell, S.M., Pauli, J.L., Holloway, B.B., Johnston, A.D., Nathanson, N.M., Deisseroth, K., et al.: Direct excitation of parvalbumin-positive interneurons by m1 muscarinic acetylcholine receptors: roles in cellular excitability, inhibitory transmission and cognition. The Journal of physiology **592**(16), 3463–3494 (2014)
82. Zhang, S., Xu, M., Kamigaki, T., Do, J.P.H., Chang, W.C., Jenvay, S., Miyamichi, K., Luo, L., Dan, Y.: Long-range and local circuits for top-down modulation of visual cortex processing. Science **345**(6197), 660–665 (2014)
83. Zhang, Y., Hamilton, S.E., Nathanson, N.M., Yan, J.: Decreased input-specific plasticity of the auditory cortex in mice lacking m1 muscarinic acetylcholine receptors. Cerebral cortex **16**(9), 1258–1265 (2005)
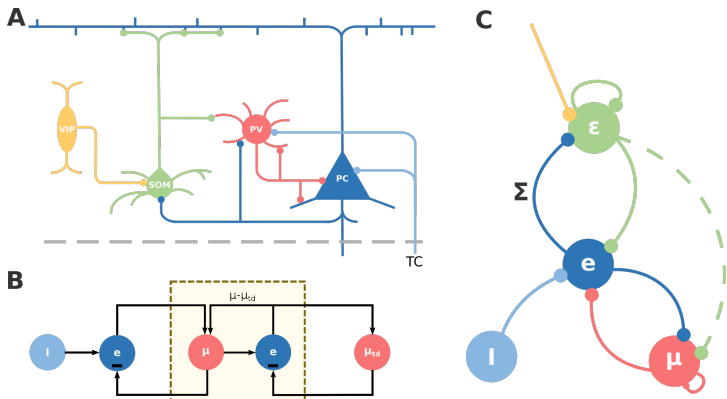
Fig. 1 Models of cortical inhibitory networks. **A** Model by [73] that includes the inhibitory network of three principal inhibitory cell types that compose the cortical inhibitory network presented by [65]. Recently, models cortical microcircuitry have included VIP cells as the key player for gating and modulating information from other cortical areas [9] as attentional dynamics driven by contextual information [3]. Notice that the classical microcircuit would summarize all inhibitory subtypes (VIP, SOM, PV) into a single one, recurrently connected to excitatory Pyramidal Cells (PCs) [18] **B** Predictive model of the visual neocortex that provides a functional interpretation of the cortical microcircuitry [61]. While these models provide a functional interpretation for the cortical microcircuits, the lack of sufficient biological constraints limits the scope of such models. **C** Model derived as in [61] to provide an estimation not just of the input but also of its statistical moments (mean and variance). While simpler in scope, the model can be constrained according to biology while identifying exclusive functions for each inhibitory interneuron subtype it includes (see derivation and results below).
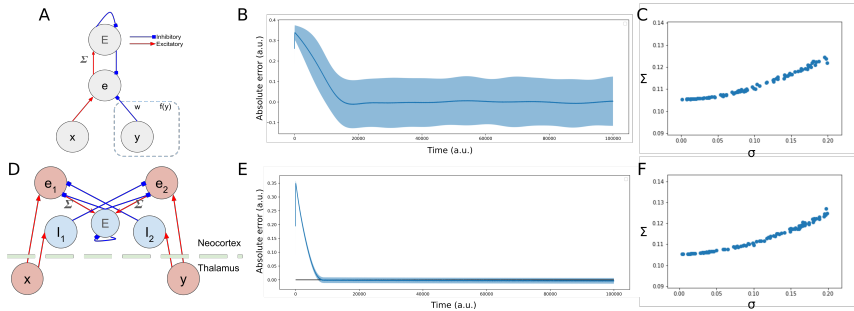
Fig. 2 A minimal neural circuit to estimate input statistics. **A** Graphical illustration of the dynamical system depicted in the text. **B** Absolute error computed as the mean difference between x and wy over time. Average over 500 trials with randomized network initializations. Shadow blue indicates standard deviation under the same conditions. **C** Estimation of variance (Σ) as compared to the standard deviation (σ) used to generate the random input x. **D** Illustration of a biologically constrained network where neurons can't switch sign and therefore have exclusively excitatory (red) or inhibitory (blue) influence on neighboring neurons. **E** and **F** are same as B and C for the biologically constrained model depicted in D.
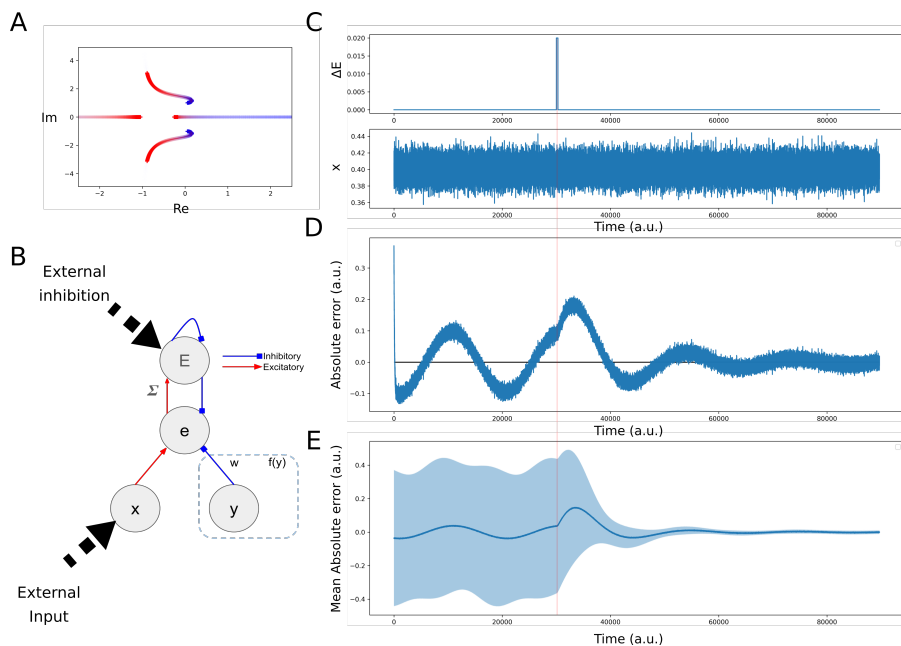
Fig. 3 Top-Down modulation of uncertainty re-enables learning after convergence. **A** illustration of the eigenvalues of our model in function of the relation between the relative error e and the estimated variance Σ. Red indicates greater values of Σ and blue greater values of e. **B** Illustration of the setup to simulate the top-down influence of uncertainty. While *external input* has been present in the form of a random normal variable that needs to be predicted, in this setup we include a top-down signal that illustrates the identification of an abnormal state in the network, i.e., there is a big error and a small predicted variance. In order to learn from this particular event, we inject external current in the units ϵ. **C** Graphical representation of the input signal x and the cholinergic equivalent input expressed as a $\Delta\epsilon$ in function of time. **D** Single example of the effect of a short top-down stimulation on the absolute error, after reaching an abnormal oscillatory state due to a mismatch between estimated relative error e and variance Σ. **E** Same as D averaged over 500 trials with different random initial conditions. Notice that the signals before stimulation have increased variability (blue shadow) and a brief stimulation is sufficient to bring the system back to a stable regime.