

Review

Normalization and the Cholinergic Microcircuit: A Unified Basis for Attention

Taylor W. Schmitz^{1,*} and John Duncan^{2,3}

Attention alters three key properties of population neural activity – firing rate, rate variability, and shared variability between neurons. All three properties are well explained by a single canonical computation – normalization – that acts across hierarchically integrated brain systems. Combining data from rodents and nonhuman primates, we argue that cortical cholinergic modulation originating from the basal forebrain closely mimics the effects of directed attention on these three properties of population neural activity. Cholinergic modulation of the cortical microcircuit underlying normalization may represent a key biological basis for the rapid and flexible changes in population neuronal coding that are required by directed attention.

The Noise-Control Systems of the Brain

Neurons are noisy. Even to successive presentations of the same stimuli, the responses of neurons will vary. Researchers typically discard this variability by averaging the responses of neurons over many trials. However, our brains do not have this luxury. From one moment to the next, we make sense of our environments in real time. This is particularly relevant to attention, in which a behaviorally relevant stimulus is prioritized over multiple different competing stimuli. If the neuronal responses to stimuli vary unpredictably, how does attention compensate for this uncertainty? In this review we draw on several lines of research in non-human animals, which, collectively, are starting to zero in on the core computations and neurobiological basis of the noise-control systems of the brain and their role in attention.

In the sensory cortex, neuronal populations are organized according to maps of selectivity for different stimulus features. Because neuronal populations with this functional architecture are inherently noisy, some component of their responses to stimuli will not be informative. We begin by discussing monkey electrophysiological work indicating that attention alters three population neuronal response patterns – firing rate, rate variability, and correlated variability – to enhance particular stimuli while compensating for noise. We focus first on the visual cortex, where these three patterns are currently best characterized, and discuss how attention-driven changes optimize the amount of information carried in the population neural code. We then discuss the emerging view that these neural signatures of attention are not coincidental, and instead reflect modulation of a single computation: normalization. Moreover, because normalization is likely to be a canonical computation – emerging in any cortical population where competitive interactions must overcome noise – we argue that attention likely utilizes normalization at multiple stages of the cortical hierarchy beyond the visual cortex, up to and including the highest levels of association cortex.

This raises a core question: what brain system is capable of rapidly and reversibly modulating normalization in local neural populations, at multiple stages of cortical processing? In the second part of the review we discuss cutting-edge optogenetics and biosensor research – predominantly

Highlights

Multi-electrode electrophysiology work indicates widespread effects of attention on neural firing rates, firing rate variability or noise, and correlated noise between neurons.

These changes are well explained by a single computation – normalization – optimizing population neuronal coding similarly throughout the cortical hierarchy.

The effects of attention are closely mirrored by cholinergically driven changes in population response patterns, implying that both modulate normalization.

Modulation of cholinergic drive may be a key neural mechanism of directed attention.

¹Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, 3801 rue University, Montreal, QC, H3A 2B4, Canada

²Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, UK

³Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

*Correspondence: taylor.schmitz@mcgill.ca (T.W. Schmitz).

in rodent models – which has provided strong evidence that **acetylcholine** (ACh; see [Glossary](#)) closely mimics the effects of directed attention on normalization. Specifically, we discuss work showing that the release of ACh onto specific cortical cells – via stimulation of **cholinergic neurons** of the **basal forebrain** system – rapidly alters the firing rate, rate variability, and correlated variability of population neural responses. As with experimental manipulations of directed attention, this cholinergically driven cortical circuit optimizes information in the population neural code. We then discuss complementary evidence showing that the basal forebrain system is itself ‘wired’ to target multiple levels of the cortical hierarchy – and thus could constitute a unified neural system for distributed attention. Finally, we discuss future directions motivated by this theoretical framework.

The Noisy Functional Architecture of Neuronal Populations

In the sensory cortex, neurons are organized according to feature selectivity maps, such that each neuron will respond maximally only to specific stimulus features – its preferred feature space. In the visual cortex, this feature space could include a particular orientation or color [1–4], or location in the visual field [5–8], or usually both. In the auditory cortex, a neuron might prefer a particular frequency [9]; in the somatosensory cortex, a specific texture [10]. However, what a scientist can derive psychophysically about the preferred feature of a neuron is an oversimplification of its actual preferred feature space. Even primary visual neurons usually do not prefer values along a single feature dimension, but prefer conjunctions of features such as orientation, spatial frequency, and direction of motion [11].

If each neuron of the sensory cortex exhibits a preference for a unique feature space, this will give rise to immensely diverse neuronal populations. This diversity equips the brain with the ability to represent the features of virtually any stimulus. However, it also creates a problem because any stimulus will activate many different neurons, with different preferred feature spaces. If the responses of neurons contained little to no stimulus-independent noise, or if the noise for a neuron was independent of its neighbors, the average response over a sufficiently large population would provide reliable (and discriminable) estimates of the responses of the most-informative neurons to different stimuli. Unfortunately, this is not the case. Neurons exhibit noise in the form of stimulus-independent fluctuations, and, most problematic of all, this noise is correlated between neurons [12].

When attending to one of multiple competing stimuli, the responses of the most informative neurons (those with preferred feature spaces that best represent the attended stimulus) will compete with responses from less informative neurons against a background of shared uncertainty. While selectively biasing the responses of the most informative neurons over these competing responses [13], the brain also requires a mechanism that compensates for the underlying noise. Below we describe evidence linking this mechanism to three neural signatures and a single core computation.

Selective Attention and Noise in the Sensory Cortices

One way to enhance the response to a selected stimulus is simply to increase the signal of the neurons that are most informative of the stimulus. This is accomplished by increasing the spike rate of neurons whose feature preferences best match the stimulus [3,14]. This form of neural modulation is one of the most widely studied properties of directed visual attention [2,4,6,15]. Consider the activity of an individual neuron when its preferred feature is either inside or outside the focus of attention. In the experiment depicted in [Figure 1A](#), for instance, the focus of attention of the monkey is directed to either of the two simultaneously presented stimuli in left and right visual hemifields, while recordings are acquired from two visual cortical neurons

Glossary

Acetylcholine (ACh): a neurochemical released by cholinergic neurons that can act both as a neuromodulator (slow and diffuse modulation affecting many neurons) and a neurotransmitter (fast and localized transmission of signals across the synapse of individual neurons).

Basal forebrain: a small collection of nuclei situated in front of and below the striatum, which contain virtually all the cortically projecting cholinergic neurons of the brain.

Cholinergic neuron: a neuron that manufactures, transports, and releases ACh.

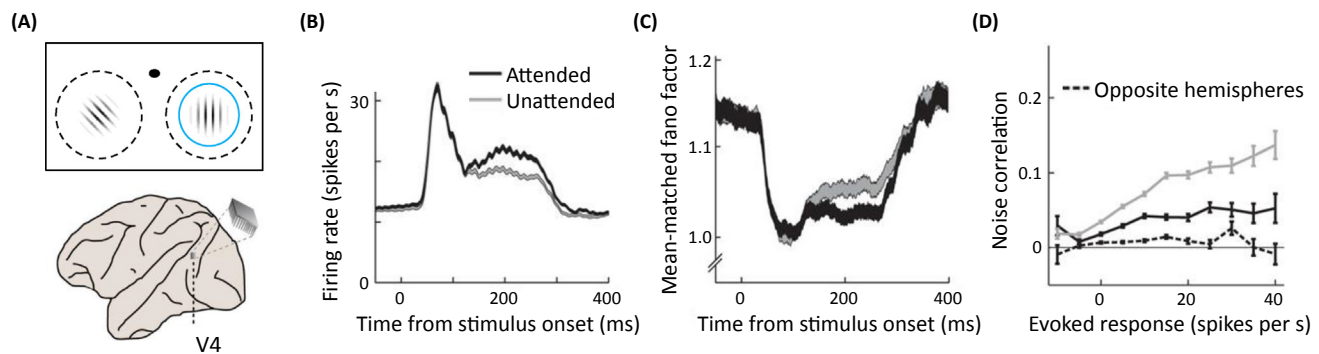
Divisive normalization: a type of computation in which the measured estimate of the excitatory response of a neuron to a stimulus is divided by the sum of its own excitatory response and those of nearby neurons.

Fano factor: the trial-to-trial variance in the spike-count of a neuron divided by its spike-count mean.

Linear classifier: a means of statistical classification used to determine which class or group a value belongs to based on its characteristics. In electrophysiology, we might determine which condition (e.g., attended versus unattended) the response pattern of a neuron belongs to based on firing rate.

Noise correlation: a measure of the degree to which trial-to-trial variance in spike count is shared by a pair of neurons.

Signal correlation: a measure of the degree to which the mean responses to different stimuli are shared by a pair of neurons.



Trends in Cognitive Sciences

Figure 1. Directed Visual Attention Alters Signal-to-Noise Ratio (SNR) and Noise Correlation of Population Neural Response. (A) Macaques performed an orientation-change detection task. An example target stimulus is displayed. Spatial attention was manipulated by cuing different target locations. Monkeys fixated on the upper central black dot. The locus of covert attention is depicted by the blue circle (not visible). The receptive field locations of two neurons are depicted by the black circles (not visible). (B) Single-unit recordings revealed that when attention was directed to the preferred location of a neuron (black lines), firing rates (signal) increased, and (C) trial-to-trial variability (noise) decreased relative to inattention (gray lines). The Fano factor is the variance of the spike counts divided by their mean. Attention-driven decreases in Fano factor persisted even when controlling for differences in the mean of the spike counts on different trials. (D) For pairs of neurons with overlapping receptive fields within the same hemisphere, attention decreased noise correlation (unbroken black line) relative to inattention (gray line). Pairs of neurons in the opposite hemisphere (broken black line) exhibited values of r_{noise} close to zero. (A) Adapted, with permission, from [34]. (B–D) Adapted, with permission, from [17].

preferring either left or right hemifield locations. Under such circumstances, attention to the preferred feature of a neuron (e.g., right hemifield location) increases its mean firing rate on a given trial compared to inattention (Figure 1B). This feature-similarity gain mechanism [3] amplifies the responsiveness of neurons that are most informative of the stimulus, and thus the stimulus representation.

However, by itself, an attention-driven increase in firing rate for an individual neuron is only informative if it is also reliable. If the spike count of a neuron to the same stimulus varies widely from one trial to the next relative to its mean response over trials – producing a high **Fano factor** – we will be less certain of what stimulus occurred on any given trial. In addition to its effect on spike rates, attention also directly alters the variability of neuronal firing rates from trial to trial. These effects on variance are not simply a byproduct of coincidental changes in firing rate. By isolating neurons that exhibit negligible change in their firing rate before stimulus onset versus after stimulus onset, or during attention versus inattention, researchers have demonstrated that reductions in Fano factor persist even in these ‘mean-matched’ neurons [16–18]. Across trials, this joint influence of attention on firing rate and trial-to-trial variability improves the signal-to-noise ratio (SNR) of the response output of the neuron (Figure 1B,C).

Despite the influence of attention on SNR, correlated variability remains as possibly the main limiting factor on how accurately responses in a neural population reflect a specific stimulus. Even if an individual neuron exhibits low levels of trial-to-trial variability, correlations in the stimulus-independent spiking activity between neurons can lead to major distortions in the population-averaged responses to a stimulus. Unlike private variability, or neuronal noise that is independent among neurons, correlated variability cannot be averaged out from population neuronal responses. These **noise correlations**, or r_{noise} , are typically small, positive, and strongest among neurons closely situated to one another and with similar feature preferences, in other words they exhibit a topographic organization typical of sensory tuning [19,20].

How does attention compensate for correlated variability? In recent multi-electrode electrophysiology work [17,19,21,22] a novel mechanistic basis of attention has emerged which may

provide an answer to this question. In addition to improving the SNR of individual neurons, attention alters the structure of noise correlation in neuronal populations (Figure 1D). From an information processing perspective, this finding has enormous implications because noise correlations can have a large influence on population sensitivity – the amount of information carried by a population neural code [12]. Decoding techniques such as **linear classifiers** can be used to estimate the influence of noise correlations on population sensitivity. A linear classifier quantifies the discriminability between neuronal responses to two different types of stimulus, for example, different orientations. If noise correlations in the data reduce discriminability – in other words classification performance goes down – we can infer that they adversely influence population sensitivity. Indeed, for populations of neurons, computational modeling and electrophysiology work indicates that attention-driven changes in noise correlations can potentially exert large influences on population sensitivity [12,23], although this remains an active area of study and debate [24–26].

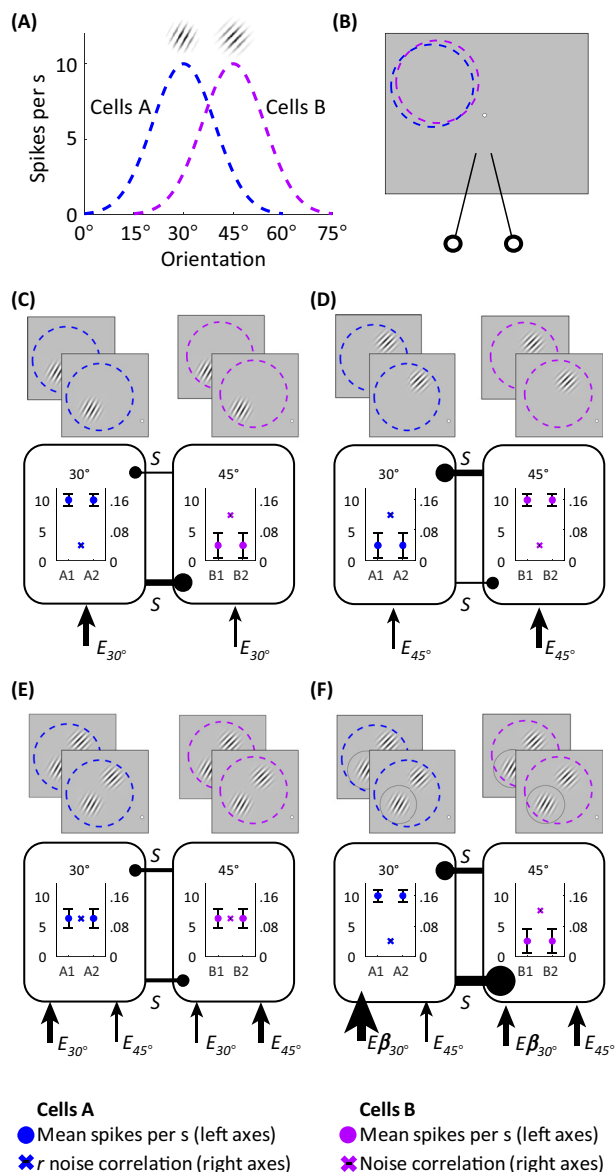
These discoveries reveal potent mechanisms of attention expressed not only in the SNR of the response patterns of individual neurons but also in the correlated variability among populations of neurons with similar feature preferences. Do these three neural signatures reflect coincidental mechanisms or a single core computation? In the next section we discuss recent combined computational and electrophysiology work indicating that normalization provides a candidate explanatory model for the effects of attention on firing rate, rate variability, and correlated variability.

A Canonical Computation for Stimulus Interaction and Noise Reduction

Normalization of neural responses is a canonical cortical computation that is used to model phenomena ranging from responses to light in the retina, to stimulus competition in the primary visual cortex, to the effects of attention in multiple areas of the visual hierarchy [27]. In normalization models, the activity of a given neuron is scaled by net activity across a larger, surrounding pool. For example, if neurons within the pool are mutually inhibitory, then the activity of any given neuron is forced down as total activity goes up. In recent years, increasing evidence has shown that attention can be well understood as a modulation of intrinsic cortical normalization [28,29].

What normalization means is that the response output of a single neuron is not simply determined by the match between its preferred feature space and the excitatory input from a particular stimulus. In our natural environments, multitudes of concurrent stimuli compete for sensory access. To accurately model the response of any given neuron to a stimulus in its receptive field, we must also take into account potential normalization, or suppressive effects, from concurrent stimuli. Consider two neuronal populations A and B (Figure 2), and, for simplicity, limit the members of each population to two neurons (A1 and A2, B1 and B2). The neurons of populations A and B prefer partially overlapping feature spaces. This overlap is evident in the overlap between their tuning curves to orientation stimuli in Figure 2A and overlapping receptive field locations in Figure 2B. Populations A and B are also mutually suppressive: stimulus-driven excitatory input (*E*) of one population will induce suppressive input (*S*) to the other (Figure 2C,D).

When both the preferred and non-preferred stimulus of a given neuron are presented inside its receptive field (Figure 2E), experimental observations reveal that its mean response output does not reflect the summation of the excitatory input from each stimulus [28,30]. Instead, the mean response of a neuron is suppressed compared to instances in which its preferred stimulus is presented alone (Figure 2C,D). Essentially, by its combination of suppressive *E* and *S* input,



Trends in Cognitive Sciences

Figure 2. Interactions between Normalization and Visual Attention. (A) Two hypothetical visual cortical neurons A (blue) and B (purple), each with feature preferences for different orientations (y axes, firing rate; x axes, orientations in degrees; insets are example optimal stimuli). The response distribution of each neuron for a range of orientations is plotted as a hypothetical tuning curve. Their tuning curves overlap, indicating that they have a signal correlation (r_{signal}) greater than zero. (B) The receptive fields of neurons A (blue) and B (purple) also overlap at similar visual field locations relative to fixation (central dot). Overt attention is directed to fixation. (C–F) (Top panels) Repetitions of stimulus configurations with an unattended 30° orientation (C), an unattended 45° orientation (D), unattended concurrent orientations (E), and covertly attended 30° orientation (F). In each case, stimuli are depicted from the perspective of the receptive fields for neurons A (blue) and B (purple). (C–F) (Bottom panels) Two hypothetical neuronal populations, each consisting of two neurons with feature preferences identical to either neurons A (A1,A2, blue) or neurons B (B1,B2, purple). Normalization determines the balance between excitatory E (arrows) and suppressive S (terminals) modulation, generating the observed tuning curves in (A). This balance is expressed by three population-response patterns: mean firing rates over trials (blue and purple circles, left-hand y axes are spikes per second), trial-to-trial firing rate variability (bars are SD), and correlated variability between neurons (blue and purple crosses, right-hand y axes are r_{noise}). (C,D) The balance between E and S input is offset because the stimulus better matches the feature preferences of one population over the other. In the better-matched population, mean firing rates increase, trial-to-trial variability decreases, and r_{noise} decreases. The opposite is true for the poorer-matched population. (E) Both E and S inputs are balanced by concurrent unattended presentation of the preferred and non-preferred stimuli of a population. This induces a compromise between the two populations whereby mean firing rates, trial-to-trial variability, and r_{noise} all exhibit an intermediate change in profile. (F) The identical stimulus configuration in (E), but with attention covertly directed to the 30° orientation (the 'attentional field' is depicted by a black circle). The balance between E and S inputs is offset by attention, modeled as a multiplicative gain on its E inputs ($E\beta$). The circuit approaches the state it would adopt for the attended stimulus presented alone.

each stimulus tends to drive activity to the value it would have if that stimulus were presented alone. With two simultaneous stimuli, the resulting activity level is a compromise. These findings are well modeled by dividing the excitatory inputs E of a neuron by the suppressive input S , with the latter being computed as the sum of activity across all neurons within a surrounding 'suppressive field' [31–33]. This **divisive normalization** provides a simple and powerful computational basis for predicting neuronal firing rates under diverse naturalistic sensory conditions [27].

Normalization models also have implications for trial-to-trial variability and correlated variability of population neuronal responses. Returning to our example in Figure 2, consider the influence

of reciprocal S modulation between populations A and B. The trial-to-trial variability of the responses of an individual neuron to a stimulus is influenced by the variability of its suppressive input. When the variability of S is decreased over repetitions of a stimulus (e.g., population A in Figure 2C), so too is the variability in modulation of the responses of a given neuron to that stimulus [21]. Recent combined computational and electrophysiological work also indicates that S inputs are a source of correlated noise [34]. Based on these analyses, S inputs originating from one population (A) are shared by multiple neurons in another population (e.g., B1 and B2). A stimulus-driven increase in S input from population A to B (Figure 2C) will drive up r_{noise} between neurons B1 and B2. The decrease in reciprocal S inputs to population A, conversely, drives down r_{noise} in population A.

Computationally, divisive normalization accommodates attention as a single free parameter (often denoted by the symbol β) which acts as a multiplicative gain on E inputs. Figure 2F depicts this interaction between directed attention β and the balance between E and S inputs that determine normalization. Under visual stimulation conditions identical to Figure 2E, directed attention to one of the two stimuli induces a multiplicative gain on its excitatory inputs $E\beta$. This multiplicative gain amplifies the effect of the attended input on network activity, including both its excitatory and suppressive effects, driving the network towards the state it would adopt for this input occurring alone [28,29]. As is also apparent between Figure 2F and 2C, reductions in the noise component of population neuronal responses – trial-to-trial variability and correlated variability – owing to modulation by either a stimulus or attention can be modeled in the same way: an offset in the balance of reciprocal S modulation between populations. A second noise-reducing influence might originate from the biasing signal attention β itself, which in addition to amplifying E inputs might also stabilize their shared fluctuations, thereby further decreasing correlated variability and trial-to-trial variability in the modulated population [21,35–38].

The examples in Figure 2 highlight two important properties of normalization. First, normalization in itself is not dependent on directed attention. The E and S inputs of individual visual cortical neurons will be influenced by any stimuli in their receptive and suppressive fields, inducing normalization even when those stimuli are outside the focus of attention. Second, normalization provides the underlying computational strategy for directed attention. As a biasing signal, attention optimizes population coding in large part through modulation of the E and S inputs in competing neuronal populations.

A third important property of normalization is that it is likely to be a canonical computation of cortical circuits that is embedded throughout the cortical hierarchy [27]. Normalization is relatively straightforward to model in striate and extrastriate cortices, where multi-electrode array data can be used to infer the feature preferences of multiple different neurons. As we move higher in the cortical hierarchy, however, the preferred feature spaces of neurons become more difficult to infer. This poses a challenge to modeling the E and S inputs for a given neuron. Do the inputs arise from sensory stimuli, cognitive operations, or some mixture of both?

Selective Attention and Noise in the Association Cortices

Like neurons in the sensory cortex, neurons in higher-order association cortex have feature preferences that lie along multiple intersecting dimensions of a feature space [39]. Each intersection can be construed as the integration of values along each dimension. In the primary visual cortex, this intersection might be a particular orientation given a specific spatial frequency and a particular motion direction [11]. As we move up the hierarchy, this intersection will be defined by increasing numbers of feature dimensions, resulting in exceptionally high-dimensional feature

spaces in areas such as the frontal cortex. What gives rise to this increase in dimensionality? The answer most likely rests on the number of connections that neurons receive at different stages of the cortical hierarchy, which is substantially higher and more diverse in the prefrontal compared to sensory cortices [40]. If a neuron integrates a higher number of short- and long-range inputs from different brain areas, its preferred feature space will be higher-dimensional.

Like populations of neurons in the sensory cortex, populations of neurons in higher-order association cortex have diverse and overlapping preferred features, albeit in a higher-dimensional space. To bias competition in association cortex, attention is therefore faced with the same noisy functional architecture as in the sensory cortex. This would imply that directed attention might also evoke changes in firing rate, rate variability, and correlated variability to optimize population coding in association cortex. Monkey electrophysiology work supports this prediction. When a particular stimulus feature is held in working memory, the SNR of individual lateral prefrontal neurons improves – firing rate increases, while trial-to-trial variability decreases [41,42]. Similar patterns are observed in other areas of higher-order cortex such as posterior parietal cortex [43] and frontal eye fields [44]. Multi-electrode studies have also demonstrated that working memory maintenance alters correlated variability between lateral prefrontal neurons [45–47]. As in the sensory cortex during visual attention, correlated variability can influence population neural coding. Removing noise correlation between prefrontal cortex neurons improves decoding performance for different stimuli held in working memory [45,47].

Does normalization account for the observed changes in firing rate, rate variability, and correlated variability in association cortex? If so, how do we estimate the E and S inputs for neurons in association cortex? Emerging electrophysiology work has provided some clues. Among multisensory neurons responsive to both visual and vestibular input, Ohshiro and colleagues have shown that a non-preferred input in one modality can suppress the response of a neuron when paired with a preferred input in the other modality [39,48]. This form of cross-modal suppression is consistent with normalization at the level of multisensory integration (Figure 3). A multisensory neuron might receive maximal excitatory input from a particular

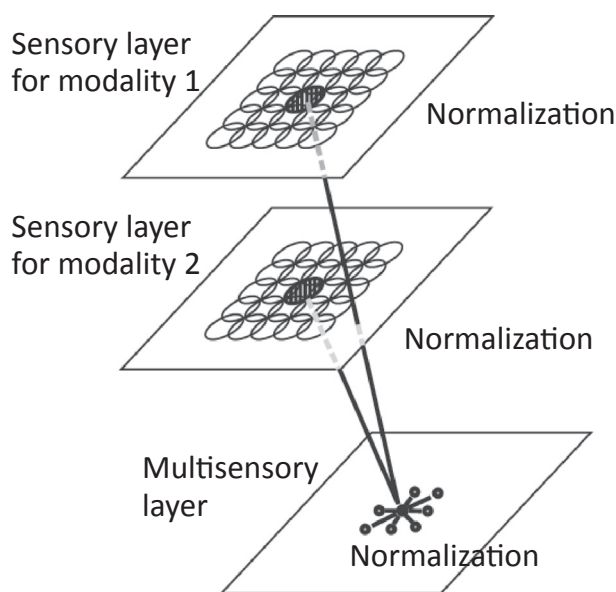


Figure 3. A Schematic of Hierarchical Integration, Attention, and Normalization. At each level of the cortical hierarchy, proceeding from the two distinct unisensory layers (visual or vestibular) to the multisensory layer (visual and vestibular), normalization represents the computational strategy for determining neuronal responses from the E and S inputs of competing populations. In the multisensory layer, neurons respond to a particular conjunction of visual and vestibular cues. Populations responsive to different conjunctions are mutually suppressive, and attention biases normalization in favor of neurons representing the selected conjunction (central node). In the unisensory layers, this strategy is repeated for the population best representing either the attended visual or vestibular feature (black circles), and those of mutually suppressive populations with overlapping feature preferences (white circles). Adapted, with permission, from [48].

Trends in Cognitive Sciences

conjunction of visual and vestibular features, which is then normalized with respect to the mutually suppressive input from other multisensory neurons preferring distinct but overlapping visual–vestibular feature conjunctions. Lending further support to the hierarchical integration of normalization, other electrophysiology work has shown that normalization can account for both stimulus- and attention-modulated changes in correlated variability not only between neurons within the same area but also between neurons in different brain areas, for example, visual cortical V1 and MT [49–51].

Moving to the highest levels of integration in the hierarchy, such as the prefrontal cortex, our ability to accurately model the *E* and *S* inputs that determine normalization and its expression in population neuronal activity becomes increasingly constrained by our ability to accurately infer the high-dimensional feature preferences of individual neurons. For instance, prefrontal neurons might exhibit mixed selectivity for conjunctions of external stimuli and internal task states – if the stimulus is *X*, and the rule is *Y*, then respond *Z* [52–54]. Nevertheless, the challenge of accurately characterizing high-dimensional feature preferences – where the *E* and *S* inputs for a given neuron might be defined by the integration of multiple sensory modalities and internal states arising from ongoing cognitive operations and memory – represents a necessary first step to accurately predicting how directed attention alters firing rate, rate variability, and correlated variability in these regions.

Extrapolating from the framework depicted in Figure 3, attention is unified by a common computation across the cortical hierarchy. When biased competition establishes a winning cell population at a higher level of integration, for example, in the prefrontal cortex, this in turn may give competitive support to diverse input neurons at lower levels of integration [13,55–57]. Normalization optimizes coding for the relevant signal, at each level of integration, by biasing the competitive interactions among local populations and reducing noise. Accordingly, a brain system is required that can reach multiple areas of the cortex to modulate normalization simultaneously, selectively, and at multiple different timescales. In the next section we discuss recent work suggesting that a key player may be the cholinergic basal forebrain system.

The Cholinergic Basal Forebrain Is a Unitary Biological System of Attention

Since the discovery of acetylcholine by Sir Henry Dale in 1914, our understanding of the role of this biochemical in the central nervous system has gone through several major revisions. Once thought to be a sluggish and diffusely acting neuromodulator involved in controlling general states of arousal or alertness, recent animal research has shown that ACh can also act as a neurotransmitter, rapidly signaling at the synapse of individual neurons [58–60]. One of the most exciting developments supporting this changing view indicates that cholinergic drive, like directed attention, can rapidly and reversibly alter the firing rate, rate variability, and correlated variability of cortical neurons.

Much of this evidence comes from rodent models in which optogenetic or electrical stimulation is applied to cholinergic neurons within the basal forebrain – which contains all the cortically projecting cholinergic neurons of the brain [61] – to experimentally control endogenous release of ACh onto distal cortical neurons [62–65]. Electrophysiological recordings of visual cortical neurons during presentations of natural visual stimuli reveal that stimulation of cholinergic drive alters the firing rate, rate variability, and correlated variability of population neural responses in a manner that closely mimics attention [63,64]. Strikingly, optogenetic inactivation of cholinergic input reverses these attention-like patterns in the population response [64] (Figure 4). These effects are unlikely to be due to changes in the general state of arousal. Concurrent measures of running speed taken during the brief optogenetic activation and inactivation periods remain

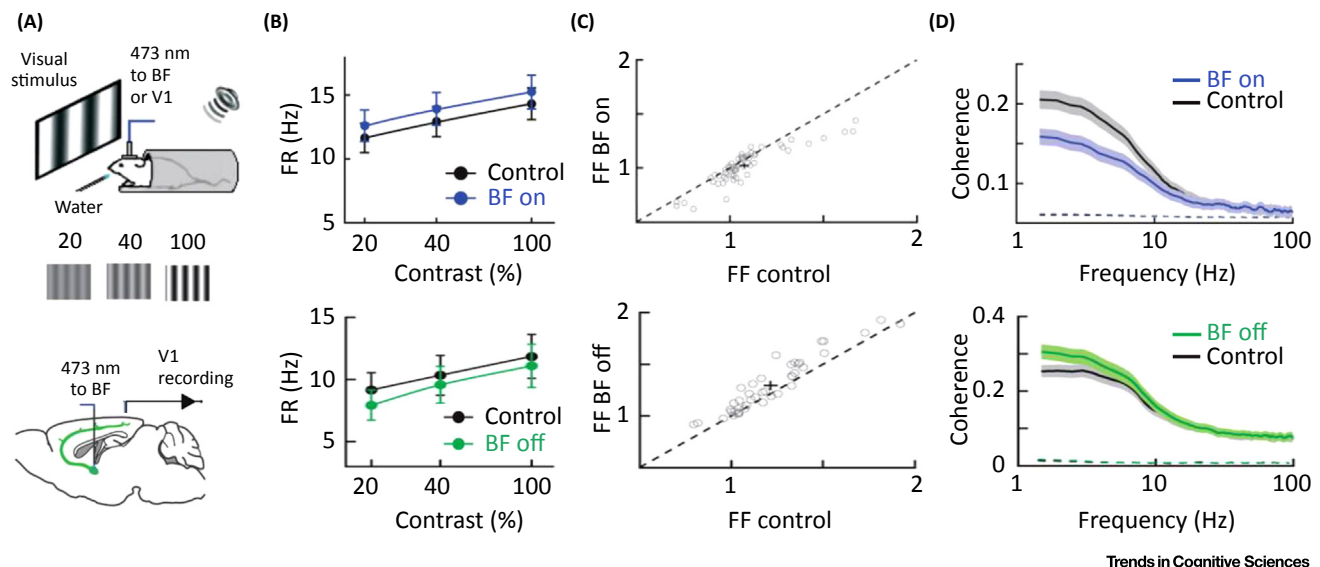


Figure 4. Rapid and Reversible Cholinergic Modulation of Signal-to-Noise Ratio (SNR) and Synchrony. (A) Mice viewed drifting gratings of varying contrast (20%, 40%, and 100%) or movie clips of natural scenes. Conditions were defined by one of three types of input on cortically projecting cholinergic neurons located in the mouse basal forebrain: optogenetic activation of cholinergic basal forebrain (BF) neurons, optogenetic inactivation of cholinergic BF neurons, and control stimulation. Multiunit electrophysiology recordings were acquired in visual cortical area V1. (B–D) Effects of cholinergic drive on mean firing rate (FR) over trials (B), trial-to-trial variability (C), and correlated variability indexed from cortical synchrony of local field potentials (D). For parallel comparison with attention-modulated effects, consult Figure 1B–D. (B) For all grating contrasts, optogenetic activation increased the mean firing rate of single units in V1 relative to control stimulation. Optogenetic inactivation decreased the mean firing rate relative to control. The same effects were observed for natural scenes. (C) Trial-to-trial variability (FF, Fano factor) in the responses of individual V1 neuron to natural scenes (open gray circles). Optogenetic activation (y axis) decreased variability in V1 over trials relative to control stimulation (x axis). Optogenetic inactivation increased the variability of V1 neurons over trials relative to control. The same effects were observed for gratings. (D) Intratrial coherence (y axis) was computed between the responses of single V1 neurons against the population average during natural movies. Low-frequency (x axis) coherence between neurons decreased under optogenetic activation compared to control stimulation. Optogenetic inactivation increased low-frequency coherence between neurons compared to control. The same effects were observed for gratings. For both plots, dotted lines represent baseline coherence levels estimated from random trial shuffles. Adapted, with permission, from [64].

unaffected. As in monkey electrophysiology work on directed visual attention [66], the influence of cholinergic drive on noise correlation also appears to depend on the similarity of the feature preferences of a neuron (their **signal correlation**) [67]. The largest changes in noise correlation are observed among neurons that share similar feature preferences and, probably, suppressive fields [34] (Figure 5).

How do these cholinergically driven changes in population neuronal responses affect population neuronal coding? Like directed attention, stimulation of ACh release in the visual cortex improves the amount of stimulus information carried in the population code, as indexed by decoding performance for different types of visual stimuli from population neural responses [63,67]. Cholinergic drive of cortical population responses may thus constitute a key biochemical basis of the rapid and spatially localized population coding dynamics of attention. Another major implication of the rodent research is that, like attention, cortical release of ACh modulates normalization. Below we discuss several avenues of research that motivate further exploration of this hypothesis.

First, if ACh modulates normalization, then its influence on this canonical computation should not be restricted to the primary sensory cortex. Demonstrations of cholinergically driven changes in SNR and noise correlation elsewhere in the cortical hierarchy, and under a variety

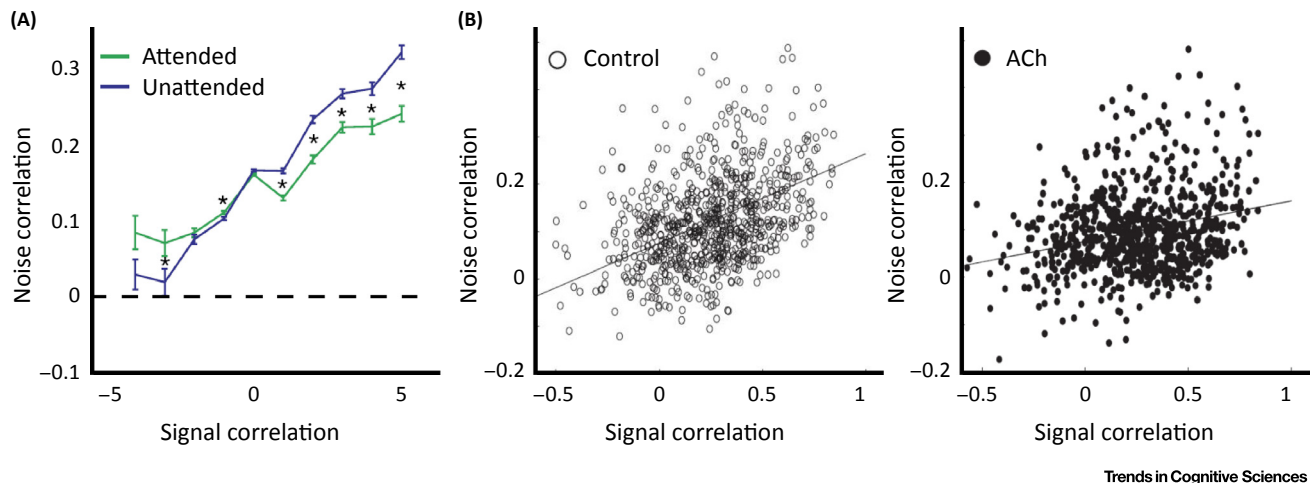


Figure 5. Relationship between Signal Correlation (x Axes) and Noise Correlation (y Axes) in the Visual Cortex. Neurons preferring the same stimulus features (signal correlation $r > 0$) exhibit the most shared noise. The effects of directed visuospatial attention (A) and optogenetically controlled release of cortical acetylcholine (ACh) (B) both shift the magnitude of this correlation downward. Hence, attention and cholinergic drive both exert the strongest decrease in noise correlation among neurons which share similar feature preferences. (A) and (B) adapted, with permission, from [66] and [67], respectively.

of task conditions, are still needed. Nevertheless, recent work using state-of-the-art electrochemical biosensors implanted in the prefrontal cortex, which are sensitive to millisecond-scale changes in synaptic ACh, have shown that phasic changes in ACh and firing rate are linked to specific behaviorally relevant task events such as cue detection [68,69]. As this exciting line of research evolves to include multi-electrode arrays, more directed experiments testing the links among cholinergic drive, population neuronal response patterns, and normalization across different areas of cortex will come into reach.

Second, if directed attention coordinates normalization through hierarchical integration, then the cholinergic system should mirror this functional architecture. However, this proposal rests on major assumptions about the form and function of the cholinergic system itself, which we discuss in the remainder of this section.

Hierarchical integration implies that populations of neurons at different levels of integration are functionally connected. If the cortical cholinergic projections modulate normalization in this scheme, they should be organized in a way that allows simultaneous modulation of interconnected cortical areas. Emerging evidence from studies using simultaneous retrograde cell labeling of cholinergic neurons from different areas of cortex strongly supports such an organization. The cortical cholinergic projections of the basal forebrain exhibit a complex topography characterized by pools of neurons with either segregated or overlapping cortical projections, and where the degree of overlap appears to depend on the interconnectivity among the cortical regions to which they project [70,71]. Similarly, the axonal branches (collaterals) emerging from individual cholinergic neurons also appear to target functionally interconnected cortical areas [72–74]. At multiple spatial scales, the cholinergic projections are well organized to simultaneously modulate hierarchically integrated cortical areas.

Despite an architectural plan consistent with hierarchical integration, the cholinergic projections must also be endowed with a mode of neurotransmission to rapidly and reversibly modulate the balance between E and S inputs that determine normalization at each level. Here too emerging

evidence supports the changing view that, in addition to its slower and more diffuse effects [75–77], cortical release of ACh can resemble ‘wired’ synaptic transmission [58,59]. Optogenetic stimulation of cholinergic basal forebrain projections rapidly alters the local spatiotemporal pattern of inhibition in three types of GABAergic interneurons: parvalbumin (PV)-, somatostatin (SOM)-, and vasoactive intestinal peptide (VIP)-expressing cells [62,78–85]. These three interneuron subtypes are thought to form a canonical cortical circuit [82–85]. A core property of this circuit is that it allows both inhibitory and disinhibitory signaling in populations of excitatory glutamatergic neurons (Figure 6A). Moreover, recent optogenetics work [81] exploring the stimulus and task-dependent response profiles in this circuit has revealed that, like excitatory neurons, individual interneurons of each subtype also exhibit a diversity of feature preferences (Figure 6B). At the population level, simultaneous cholinergic modulation of multiple interneurons can therefore evoke diverse stimulus- and task-dependent responses which, depending on the subtype, are either inhibitory or disinhibitory. This diversity of cortical inputs, in turn, leads to a diversity of excitatory cortical outputs (Figure 6C).

Why might cholinergic input recruit the multiple parallel signaling pathways of this cortical circuit? The emerging evidence indicates that parallel processing within this circuit enables it to generate different responses to the same stimuli depending on context [81]. Returning to the example in Figure 2F, the relevance of these findings to attention becomes readily apparent. Consider a slightly different scenario where covert attention must be shifted from the 30° to the 45° orientation stimulus. The E inputs for the two stimuli must flip in their weighting, along with resulting changes in S interactions, effectively reversing modulation of SNR and correlated variability within each population, even though the external stimuli remain constant. To accomplish this context-dependent modulation, some component of the circuit must maintain the current task context (attend 30°), and another component must convey rapid and reversible changes in context (shift attention to 45°). At the circuit level, this might be achieved by means

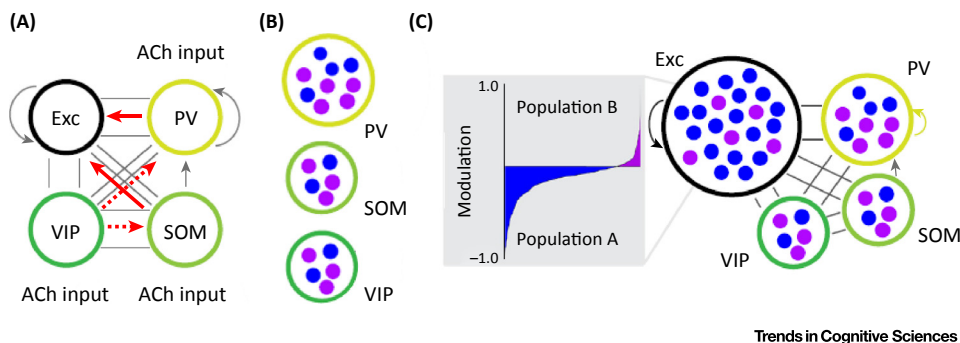


Figure 6. A Cell Type-Specific Model of a Cholinergically Modulated Cortical Microcircuit Underlying Normalization. (A) Cortical release of acetylcholine (ACh) due to stimulation of basal forebrain cholinergic neurons primarily affects three types of inhibitory interneurons: parvalbumin (PV)-, somatostatin (SOM)-, and vasoactive intestinal peptide (VIP)-expressing cells. Cholinergically driven changes in these inhibitory interneurons explain both response facilitation and response suppression of excitatory pyramidal neurons (Exc). For instance, PV and SOM interneurons appear to directly inhibit Exc neurons (unbroken red arrows), whereas VIP interneurons disinhibit Exc neurons (broken red arrows). (B) For individual cells of each subtype of inhibitory interneuron (blue and purple units), the strength of their inhibitory or disinhibitory drive on Exc populations also depends on their distinct feature preferences; for example, interneurons might prefer 30° or 45° orientations similarly to populations A (blue) or B (purple) from Figure 2. (C) A multiunit model of the microcircuit reveals that these diverse inhibitory inputs can, in turn, lead to diverse excitatory outputs (gray box) for Exc populations with distinct feature preferences A and B. Hence, parallel processing of cholinergic modulation by diverse inhibitory inputs enables the microcircuit to generate different responses to the same stimuli depending on the context, for example, shifting attention between two competing orientations (Figure 2F). Cholinergic modulation of this cortical microcircuit might therefore constitute a key interface between attention and normalization. Adapted, with permission, from [81].

of parallel adjustment to the inhibitory and disinhibitory inputs targeting both populations A and B. Parallel cholinergic modulation of this cell type-specific cortical circuit (Figure 6C) might therefore constitute a key biophysical basis underlying the computational implementation of attention and normalization (Figure 2F).

Our understanding of this cortical circuit is far from complete. Many additional cholinergic signaling pathways may exist. For instance, optogenetic stimulation of basal forebrain cholinergic neurons has been found to modulate the responses of individual orientation-tuned V1 neurons to specific orientation stimuli via a non-neuronal signaling pathway involving astrocytes [86]. In another optogenetics study, the cholinergic basal forebrain neurons themselves were shown to be capable of manufacturing, transporting, and phasically coreleasing both ACh and GABA [87]. There are also diverse cholinergic receptors in the cortex, belonging to either muscarinic or nicotinic subtypes, whose distinct roles in this cortical circuit are not well characterized. Finally, more work will be necessary to determine whether the observed effects of cholinergic drive on population neuronal responses are uniform across cortical layers or exhibit layer-specific differences [88–90]. Although beyond the scope of the current review, finer-grained elucidation of this cortical circuit will likely provide further insights at the direct interface between neurobiology and cognition.

In sum, the cholinergic basal forebrain projection system is exquisitely wired to coordinate simultaneous, rapid, and reversible modulation of normalization across hierarchically integrated cortical areas. Very few of the many overlapping electrophysiological findings observed under either directed attention (primarily in monkeys) or cholinergic modulation (primarily in rodents) have been evaluated systematically, although there are already several compelling examples of interdependence between the two [91,92]. In the next section we highlight several avenues of future research where a better understanding of this interdependence may prove extremely fruitful.

Concluding Remarks and Future Directions

Cholinergic neurons are particularly vulnerable to age-related neurodegeneration [61,93], in part because of their enormous axonal projections [94–98]. Longitudinal decreases in basal forebrain gray matter volume are observed among older adults with cerebrospinal fluid biomarkers of Alzheimer's disease (AD) but no apparent deficits in short-term memory or entorhinal cortical degeneration [99], suggesting that the cholinergic system is among the earliest affected in the disease progression. However, is the cholinergic lesion in preclinical AD indeed 'clinically silent'? In preclinical stages of AD, the loss of cortically projecting cholinergic inputs may reduce the availability of ACh to modulate normalization in the cortex. With reduced capacity to bias competing neural populations and increased noise, one potential outcome would be a greater susceptibility to encoding errors and distraction during sensory processing. Many studies of visual attention in older adult samples report exactly this pattern [100–103]. Such a deficit might predispose gradual degradation in the fidelity of memory encoding and, ultimately, short-term memory retrieval [100,101,104]. Precisely how nascent cholinergic deafferentation in preclinical stages of AD affects the cortical cholinergic microcircuit, normalization, and population coding is unknown.

Cholinergic dysfunction is not limited to neurodegenerative diseases of aging. Psychiatric conditions including anxiety, depression, and post-traumatic stress disorder have been linked to abnormal cholinergic signaling [105–110]. Indeed, a growing body of cross-species translational psychiatry research indicates that, in contrast to AD, a common mechanism of dysfunction across these disorders is persistent hyperactivity of cortical ACh. Of particular

Outstanding Questions

Can we model normalization in circumstances where the distinct feature preferences of neuronal populations are difficult to infer, for example, at high levels of integration such as the frontal cortex? If so, how?

Are the dynamic changes in noise correlation induced by attention detectable with non-invasive brain imaging techniques such as fMRI and scalp EEG?

What are the temporal and spatial scales at which cortical release of ACh modulates population coding during tasks requiring 'top-down' directed attention?

How does the spatiotemporal scale of cholinergic neuromodulation relate to the capacity limits of attention?

For diseases in which pathophysiology targets the cholinergic basal forebrain neurons, can chronic hyper- or hypo-cholinergic modulation predict reliable changes in population coding?

interest are the direct cholinergic projections from the septal nucleus of the basal forebrain to the hippocampus – the septohippocampal pathway – which generate a range of stable oscillatory network states [111–113] that are important for maintaining cognitive functions such as attention [114]. Both pharmacological blockade and genetic knockdown of the acetylcholinesterase enzyme – which breaks down ACh in the synapse upon neurotransmission [59] – yield hyperactive cholinergic signaling in the mouse hippocampus as well as anxiety- and depression-like behaviors [115].

One possibility is that the hippocampal cholinergic hypersensitivity observed in anxiety, depression, and post-traumatic stress disorder arises from, or interacts with, deficient inhibitory function of GABAergic interneurons. Muted GABAergic tone local to the hippocampus is also well characterized in each of these disorders, as well as in schizophrenia [116–128]. Moreover, in cognitively normal populations, lower concentrations of hippocampal GABA are associated with increased susceptibility to intrusive thought [129]. An imbalance between the inhibitory and disinhibitory GABAergic drives of hippocampal interneurons, for example, in the three classes of interneurons described in Figure 6, might therefore disrupt normalization of their responses, increasing susceptibility to runaway excitation. Consequently, at the population level, this imbalance might lower the threshold of activation for unwanted or arbitrary memories and thoughts, potentiating a constellation of disruptions to attention such as flash-backs, rumination, persistent worry, and hallucination.

In the preceding sections we proposed that both directed attention and ACh can modulate normalization to optimize population coding throughout the cortical mantle. These lines of evidence raise many questions about the ‘sameness’ of attentional and cholinergic function. If directed attention manifests from biased competition across levels of integration in the cortical hierarchy, does ACh influence the ‘directedness’ of this hierarchical integration? Attention is a fast and flexible resource; if cholinergic drive is crucial, mechanisms must exist for routing its effects to highly specific neural populations, perhaps through interaction with corticocortical ‘control’ inputs. At the same time, attention is also a capacity-limited resource [130]. We can process only so much information at any moment; one possibility is that the cholinergic system also constrains attention, to some degree, by its spatiotemporal upper bound. While the evidence discussed in this review hints at the plausibility of ACh as a key neurobiological basis of attention, more work will be necessary to clarify precisely how and when attention and ACh interact to improve population coding (see Outstanding Questions). Finally, if normalization is a canonical cortical computation, then dysfunction in its cortical microcircuitry will have severe widespread consequences for population coding. Multiple diseases of the central nervous system characterized by pathophysiology in the cholinergic system may thus share abnormal population coding – and specific disturbances of attention – as a common feature.

Acknowledgments

Financial support was provided by the Medical Research Council (programme SUAG/002/RG91365).

References

1. Roelfsema, P.R. *et al.* (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376–381
2. Spitzer, H. *et al.* (1988) Increased attention enhances both behavioral and neuronal performance. *Science* 240, 338–340
3. Martinez-Trujillo, J.C. and Treue, S. (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751
4. Burkhalter, A. and Van Essen, D.C. (1986) Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J. Neurosci.* 6, 2327–2351
5. Reynolds, J.H. *et al.* (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26, 703–714
6. Moran, J. and Desimone, R. (1985) Selective attention gates visual motion processing in the extrastriate cortex. *Science* 229, 782–784
7. Treue, S. and Maunsell, J.H. (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382, 539–541

8. Roberts, M. *et al.* (2007) Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. *Nat. Neurosci.* 10, 1483–1491
9. Rauschecker, J.P. *et al.* (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114
10. Gomez-Ramirez, M. *et al.* (2014) Temporal correlation mechanisms and their role in feature selection: a single-unit study in primate somatosensory cortex. *PLoS Biol.* 12, e1002004
11. Mante, V. *et al.* (2005) Independence of luminance and contrast in natural scenes and in the early visual system. *Nat. Neurosci.* 8, 1690–1697
12. Averbeck, B.B. *et al.* (2006) Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366
13. Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222
14. Treue, S. and Martinez Trujillo, J.C. (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579
15. Maunsell, J.H. and Newsome, W.T. (1987) Visual processing in monkey extrastriate cortex. *Annu. Rev. Neurosci.* 10, 363–401
16. Mitchell, J.F. *et al.* (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55, 131–141
17. Cohen, M.R. and Maunsell, J.H. (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600
18. Churchland, M.M. *et al.* (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378
19. Cohen, M.R. and Kohn, A. (2011) Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819
20. Ruff, D.A. *et al.* (2016) Relating normalization to neuronal populations across cortical areas. *J. Neurophysiol.* 116, 1375–1386
21. Cohen, M.R. and Maunsell, J.H. (2010) A neuronal population measure of attention predicts behavioral performance on individual trials. *J. Neurosci.* 30, 15241–15253
22. Cohen, M.R. and Maunsell, J.H. (2011) Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* 70, 1192–1204
23. Abbott, L.F. and Dayan, P. (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101
24. Kanitscheider, I. *et al.* (2015) Origin of information-limiting noise correlations. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6973–E6982
25. Kohn, A. *et al.* (2016) Correlations and neuronal population information. *Annu. Rev. Neurosci.* 39, 237–256
26. Moreno-Bote, R. *et al.* (2014) Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417
27. Carandini, M. and Heeger, D.J. (2011) Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62
28. Reynolds, J.H. *et al.* (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753
29. Reynolds, J.H. and Heeger, D.J. (2009) The normalization model of attention. *Neuron* 61, 168–185
30. Ni, A.M. and Maunsell, J.H.R. (2017) Spatially tuned normalization explains attention modulation variance within neurons. *J. Neurophysiol.* 118, 1903–1913
31. Heeger, D.J. (1992) Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181–197
32. Carandini, M. and Heeger, D.J. (1994) Summation and division by neurons in primate visual cortex. *Science* 264, 1333–1336
33. Carandini, M. *et al.* (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17, 8621–8644
34. Verhoef, B.E. and Maunsell, J.H.R. (2017) Attention-related changes in correlated neuronal activity arise from normalization mechanisms. *Nat. Neurosci.* 20, 969–977
35. Kanashiro, T. *et al.* (2017) Attentional modulation of neuronal variability in circuit models of cortex. *Elife* 6, e23978
36. Rabinowitz, N.C. *et al.* (2015) Attention stabilizes the shared gain of V4 populations. *Elife* 4, e08998
37. Goris, R.L. *et al.* (2014) Partitioning neuronal variability. *Nat. Neurosci.* 17, 858–865
38. Harris, K.D. and Thiele, A. (2011) Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523
39. Ohshiro, T. *et al.* (2017) A neural signature of divisive normalization at the level of multisensory integration in primate cortex. *Neuron* 95, 399–411
40. DeNardo, L.A. *et al.* (2015) Connectivity of mouse somatosensory and prefrontal cortex examined with trans-synaptic tracing. *Nat. Neurosci.* 18, 1687–1697
41. Hussar, C. and Pasternak, T. (2010) Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21842–21847
42. Qi, X.L. and Constantinidis, C. (2012) Variability of prefrontal neuronal discharges before and after training in a working memory task. *PLoS One* 7, e41053
43. Falkner, A.L. *et al.* (2013) Spatial representation and cognitive modulation of response variability in the lateral intraparietal area priority map. *J. Neurosci.* 33, 16117–16130
44. Thiele, A. *et al.* (2016) Attention induced gain stabilization in broad and narrow-spiking cells in the frontal eye-field of macaque monkeys. *J. Neurosci.* 36, 7601–7612
45. Leavitt, M.L. *et al.* (2017) Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. *Proc. Natl. Acad. Sci. U. S. A.* 114, E2294–E2503
46. Markowitz, D.A. *et al.* (2015) Multiple component networks support working memory in prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11084–11089
47. Tremblay, S. *et al.* (2015) Attentional filtering of visual information by neuronal ensembles in the primate lateral prefrontal cortex. *Neuron* 85, 202–215
48. Ohshiro, T. *et al.* (2011) A normalization model of multisensory integration. *Nat. Neurosci.* 14, 775–782
49. Ruff, D.A. and Cohen, M.R. (2017) A normalization model suggests that attention changes the weighting of inputs between visual areas. *Proc. Natl. Acad. Sci. U. S. A.* 114, E4085–E4094
50. Ruff, D.A. and Cohen, M.R. (2016) Attention increases spike count correlations between visual cortical areas. *J. Neurosci.* 36, 7523–7534
51. Ruff, D.A. and Cohen, M.R. (2016) Stimulus dependence of correlated variability across cortical areas. *J. Neurosci.* 36, 7546–7556
52. Rigotti, M. *et al.* (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590
53. Rigotti, M. *et al.* (2010) Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front. Comput. Neurosci.* 4, 24
54. Fusi, S. *et al.* (2016) Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74
55. Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
56. Dehaene, S. *et al.* (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14529–14534
57. Norman, D.A. and Shallice, T. *et al.* (1986) Attention to action: willed and automatic control of behavior. In *Consciousness and Self-Regulation* (Davidson, R.J., ed.), pp. 1–18, Plenum Press
58. Munoz, W. and Rudy, B. (2014) Spatiotemporal specificity in cholinergic control of neocortical function. *Curr. Opin. Neurobiol.* 26, 149–160
59. Sarter, M. *et al.* (2009) Phasic acetylcholine release and the volume transmission hypothesis: time to move on. *Nat. Rev. Neurosci.* 10, 383–390

60. Ballinger, E.C. *et al.* (2016) Basal forebrain cholinergic circuits and signaling in cognition and cognitive decline. *Neuron* 91, 1199–1218
61. Mesulam, M.M. (2013) Cholinergic circuitry of the human nucleus basalis and its fate in Alzheimer's disease. *J. Comp. Neurol.* 521, 4124–4144
62. Chen, N. *et al.* (2015) An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nat. Neurosci.* 18, 892–902
63. Goard, M. and Dan, Y. (2009) Basal forebrain activation enhances cortical coding of natural scenes. *Nat. Neurosci.* 12, 1444–1449
64. Pinto, L. *et al.* (2013) Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nat. Neurosci.* 16, 1857–1863
65. Gu, Z. and Yakel, J.L. (2011) Timing-dependent septal cholinergic induction of dynamic hippocampal synaptic plasticity. *Neuron* 71, 155–165
66. Ruff, D.A. and Cohen, M.R. (2014) Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* 17, 1591–1597
67. Mincses, V. *et al.* (2017) Cholinergic shaping of neural correlations. *Proc. Natl. Acad. Sci. U. S. A.* 114, 5725–5730
68. Howe, W.M. *et al.* (2017) Acetylcholine release in prefrontal cortex promotes gamma oscillations and theta-gamma coupling during cue detection. *J. Neurosci.* 37, 3215–3230
69. Parikh, V. *et al.* (2007) Prefrontal acetylcholine release controls cue detection on multiple timescales. *Neuron* 56, 141–154
70. Zaborszky, L. *et al.* (2015) Neurons in the basal forebrain project to the cortex in a complex topographic organization that reflects corticocortical connectivity patterns: an experimental study based on retrograde tracing and 3D reconstruction. *Cereb. Cortex* 25, 118–137
71. Zaborszky, L. *et al.* (2015) Organization of the basal forebrain cholinergic projection system: specific or diffuse? In *The Rat Nervous System* (4th edn) (Paxinos, G., ed.), pp. 491–507, Elsevier
72. Chandler, D. and Waterhouse, B.D. (2012) Evidence for broad versus segregated projections from cholinergic and noradrenergic nuclei to functionally and anatomically discrete subregions of prefrontal cortex. *Front. Behav. Neurosci.* 6, 20
73. Chandler, D.J. *et al.* (2013) Identification and distribution of projections from monoaminergic and cholinergic nuclei to functionally differentiated subregions of prefrontal cortex. *Brain Res.* 1522, 38–58
74. Chavez, C. and Zaborszky, L. (2016) Basal forebrain cholinergic-auditory cortical network: primary versus nonprimary auditory cortical areas. *Cereb. Cortex* 27, 2335–2347
75. Descarries, L. *et al.* (1997) Diffuse transmission by acetylcholine in the CNS. *Prog. Neurobiol.* 53, 603–625
76. Descarries, L. and Mechawar, N. (2000) Ultrastructural evidence for diffuse transmission by monoamine and acetylcholine neurons of the central nervous system. *Prog. Brain Res.* 125, 27–47
77. Mechawar, N. *et al.* (2000) Cholinergic innervation in adult rat cerebral cortex: a quantitative immunocytochemical description. *J. Comp. Neurol.* 428, 305–318
78. Poorthuis, R.B. *et al.* (2014) Cholinergic circuit modulation through differential recruitment of neocortical interneuron types during behaviour. *J. Physiol.* 592, 4155–4164
79. Arroyo, S. *et al.* (2012) Prolonged disynaptic inhibition in the cortex mediated by slow, non-alpha7 nicotinic excitation of a specific subset of cortical interneurons. *J. Neurosci.* 32, 3859–3864
80. Bennett, C. *et al.* (2012) Mechanisms generating dual-component nicotinic EPSCs in cortical interneurons. *J. Neurosci.* 32, 17287–17296
81. Kuchibhotla, K.V. *et al.* (2017) Parallel processing by cortical inhibition enables context-dependent behavior. *Nat. Neurosci.* 20, 62–71
82. Fu, Y. *et al.* (2014) A cortical circuit for gain control by behavioral state. *Cell* 156, 1139–1152
83. Kepecs, A. and Fishell, G. (2014) Interneuron cell types are fit to function. *Nature* 505, 318–326
84. Pi, H.J. *et al.* (2013) Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524
85. Xue, M. *et al.* (2014) Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* 511, 596–600
86. Chen, N. *et al.* (2012) Nucleus basalis-enabled stimulus-specific plasticity in the visual cortex is mediated by astrocytes. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2832–E2841
87. Saunders, A. *et al.* (2015) Corelease of acetylcholine and GABA from cholinergic forebrain neurons. *Elife* 4, e06412
88. Obermayer, J. *et al.* (2017) Cholinergic modulation of cortical microcircuits is layer-specific: evidence from rodent, monkey and human brain. *Front. Neural Circuits* 11, 100
89. Verhoog, M.B. *et al.* (2016) Layer-specific cholinergic control of human and mouse cortical synaptic plasticity. *Nat. Commun.* 7, 12826
90. Wall, N.R. *et al.* (2016) Brain-wide maps of synaptic input to cortical interneurons. *J. Neurosci.* 36, 4000–4009
91. Herrero, J.L. *et al.* (2008) Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* 454, 1110–1114
92. Thiele, A. *et al.* (2012) Contribution of cholinergic and GABAergic mechanisms to direction tuning, discriminability, response reliability, and neuronal rate correlations in macaque middle temporal area. *J. Neurosci.* 32, 16602–16615
93. Hampel, H. *et al.* (2017) Revisiting the cholinergic hypothesis in Alzheimer's disease: emerging evidence from translational and clinical research. *Alzheimers Dement.* Published online October 10, 2017. <http://dx.doi.org/10.1016/j.jalz.2017.08.016>
94. Lewis, N.E. *et al.* (2010) Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.* 28, 1279–1285
95. Mattson, M.P. (2004) Pathways towards and away from Alzheimer's disease. *Nature* 430, 631–639
96. Mattson, M.P. and Magnus, T. (2006) Ageing and neuronal vulnerability. *Nat. Rev. Neurosci.* 7, 278–294
97. Wang, X. *et al.* (2010) Functional genomics of brain aging and Alzheimer's disease: focus on selective neuronal vulnerability. *Curr. Genom.* 11, 618–633
98. Wu, H. *et al.* (2014) Complete morphologies of basal forebrain cholinergic neurons in the mouse. *Elife* 3, e02444
99. Schmitz, T.W. and Spreng, R.N. (2016) Basal forebrain degeneration precedes and predicts the cortical spread of Alzheimer's pathology. *Nat. Commun.* 7, 13249
100. Schmitz, T.W. *et al.* (2010) Failing to ignore: paradoxical neural effects of perceptual load on early attentional selection in normal aging. *J. Neurosci.* 30, 14750–14758
101. Schmitz, T.W. *et al.* (2014) Distinguishing attentional gain and tuning in young and older adults. *Neurobiol. Aging* 35, 2514–2525
102. Quigley, C. *et al.* (2010) Feature-selective attention: evidence for a decline in old age. *Neurosci. Lett.* 474, 5–8
103. Campbell, K.L. *et al.* (2010) Hyper-binding: a unique age effect. *Psychol. Sci.* 21, 399–405
104. Romberg, C. *et al.* (2013) Paying more attention to attention: towards more comprehensive cognitive translation using mouse models of Alzheimer's disease. *Brain Res. Bull.* 92, 49–55
105. Soreq, H. and Seidman, S. (2001) Acetylcholinesterase – new roles for an old actor. *Nat. Rev. Neurosci.* 2, 294–302
106. Kaufner, D. *et al.* (1998) Acute stress facilitates long-lasting changes in cholinergic gene expression. *Nature* 393, 373–377
107. Gilad, G.M. (1987) The stress-induced response of the septo-hippocampal cholinergic system. A vectorial outcome of psychoneuroendocrinological interactions. *Psychoneuroendocrinology* 12, 167–184

108. Mineur, Y.S. and Picciotto, M.R. (2010) Nicotine receptors and depression: revisiting and revising the cholinergic hypothesis. *Trends Pharmacol. Sci.* 31, 580–586
109. Furey, M.L. and Drevets, W.C. (2006) Antidepressant efficacy of the antimuscarinic drug scopolamine: a randomized, placebo-controlled clinical trial. *Arch. Gen. Psychiatry* 63, 1121–1129
110. Higley, M.J. and Picciotto, M.R. (2014) Neuromodulation by acetylcholine: examples from schizophrenia and depression. *Curr. Opin. Neurobiol.* 29, 88–95
111. Cobb, S.R. and Davies, C.H. (2005) Cholinergic modulation of hippocampal cells and circuits. *J. Physiol.* 562, 81–88
112. Freund, T.F. and Buzsáki, G. (1996) Interneurons of the hippocampus. *Hippocampus* 6, 347–470
113. Cobb, S.R. *et al.* (1995) Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* 378, 75–78
114. Ekstrom, A.D. *et al.* (2005) Human hippocampal theta activity during virtual navigation. *Hippocampus* 15, 881–889
115. Mineur, Y.S. *et al.* (2013) Cholinergic signaling in the hippocampus regulates social stress resilience and anxiety- and depression-like behavior. *Proc. Natl. Acad. Sci. U. S. A.* 110, 3573–3578
116. Benes, F.M. (2015) Building models for postmortem abnormalities in hippocampus of schizophrenics. *Schizophr. Res.* 167, 73–83
117. Crestani, F. *et al.* (2002) Trace fear conditioning involves hippocampal alpha5 GABA(A) receptors. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8980–8985
118. Crestani, F. *et al.* (1999) Decreased GABAA-receptor clustering results in enhanced anxiety and a bias for threat cues. *Nat. Neurosci.* 2, 833–839
119. Czeh, B. *et al.* (2015) Chronic stress reduces the number of GABAergic interneurons in the adult rat hippocampus, dorsal-ventral and region-specific differences. *Hippocampus* 25, 393–405
120. Gilani, A.I. *et al.* (2014) Interneuron precursor transplants in adult hippocampus reverse psychosis-relevant features in a mouse model of hippocampal disinhibition. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7450–7455
121. Heckers, S. and Konradi, C. (2015) GABAergic mechanisms of hippocampal hyperactivity in schizophrenia. *Schizophr. Res.* 167, 4–11
122. Konradi, C. *et al.* (2011) Hippocampal interneurons are abnormal in schizophrenia. *Schizophr. Res.* 131, 165–173
123. Müller, I. *et al.* (2015) The GAD65 knock out mouse – a model for GABAergic processes in fear- and stress-induced psychopathology. *Genes Brain Behav.* 14, 37–45
124. Osuch, E.A. *et al.* (2001) Regional cerebral blood flow correlated with flashback intensity in patients with posttraumatic stress disorder. *Biol. Psychiatry* 50, 246–253
125. Patel, R. *et al.* (2012) Neurocircuitry models of posttraumatic stress disorder and beyond: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 36, 2130–2142
126. Schobel, S.A. *et al.* (2013) Imaging patients with psychosis and a mouse model establishes a spreading pattern of hippocampal dysfunction and implicates glutamate as a driver. *Neuron* 78, 81–93
127. Tregellas, J.R. *et al.* (2014) Intrinsic hippocampal activity as a biomarker for cognition and symptoms in schizophrenia. *Am. J. Psychiatry* 171, 549–556
128. Yee, B.K. *et al.* (2004) GABA receptors containing the alpha5 subunit mediate the trace effect in aversive and appetitive conditioning and extinction of conditioned fear. *Eur. J. Neurosci.* 20, 1928–1936
129. Schmitz, T.W. *et al.* (2017) Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nat. Commun.* 8, 1311
130. Kahneman, D. (1973) *Attention and Effort*, Prentice-Hall