Forebrain cholinergic signaling: Wired and phasic, not tonic, and causing behavior.

Martin Sarter & Cindy Lustig
Dept. Psychology & Neuroscience, University of Michigan, Ann Arbor, MI (USA)

Running title: Phasic ACh

Words counts: abstract 388; main text 4,396

Acknowledgements: The authors’ research was supported by PHS grants R01DA045063 and P50NS091856. A preprint of this paper was previously made available (https://www.preprints.org/manuscript/201904.0010/v1). We thank Drs. Peter Dayan (Max Planck Institute for Biological Cybernetics), Howard Gritton (Boston University), William M. Howe (Icahn School of Medicine at Mt. Sinai), and Vinay Parikh (Temple University) for comments on a draft of this paper.

Competing Financial Interests: The authors declare no competing financial interest.

Addresses: Martin Sarter, Department of Psychology, University of Michigan, 530 Church Street, 4030 East Hall, Ann Arbor, MI 48109; email: msarter@umich.edu; Cindy Lustig, Department of Psychology, University of Michigan, 530 Church Street, 4036 East Hall, Ann Arbor, MI 48109; email: clustig@umich.edu
Abstract

Previous evidence in support of a slowly acting (scale of 100s of seconds) and volume-transmitted component of cholinergic signaling was based largely on studies using measures of extracellular brain acetylcholine (ACh) levels which required several minutes to generate a single data point and typically employed AChEsterase inhibitors (AChEIs) to foster the measurement of ACh. Moreover, collecting such data points in correlation with relatively stable behavioral states has supported the view that extracellular ACh levels vary at a relatively slow rate. Here we argue that forebrain cholinergic signaling is exclusively phasic (milliseconds to perhaps seconds), unlikely to be volume-transmitted, and that previous neurochemical evidence and associated behavioral correlates may be re-interpreted in terms of integrated phasic cholinergic activity and specific behavioral and cognitive operations. The highly potent catalytic enzyme for ACh, AChE, limits the presence of an ambient extracellular ACh level and thus renders it unlikely that ACh influences target regions via relatively slow changes in extracellular ACh concentrations. Real-time amperometric recordings of cholinergic signaling have suggested a specific function of rapid, phasic or transient cholinergic signaling in attentional contexts. Optogenetic studies support a causal relationship between these transients on behavior. Combined electrochemical and neurophysiological recordings revealed that the powerful behavioral control by cholinergic transients involves the generation of high-frequency oscillations. Such oscillations are thought to recruit efferent circuitry to (re)activate dormant task sets. Evidence showing the impact of genetic variations of the capacity of cholinergic synapses likewise can be interpreted in terms of their impact on the ability to sustain generation of repeated phasic cholinergic signals, as opposed to effects on ambient ACh levels. Further, while notions of slowly-changing, sleep stage-associated variations in extracellular ACh levels and their functions are widely accepted, the evidence is in fact limited. An alternative hypothesis offers a role for high-frequency cholinergic transient signaling during REM sleep. By employing a theoretical framework that focuses on the phasic and causative characteristics and functions of cholinergic signaling, results from human cognitive neuroscience studies of cholinergic function may be substantially clarified and simplified. Compared to the current treatment of cholinergic deficits using AChEIs, the conceptualization of forebrain cholinergic signaling as wired, phasic, and causative predicts that drugs that either rescue transient presynaptic signaling or amplify or rescue the postsynaptic impact of phasic signals will be more efficacious in treating age- and dementia-related cognitive and cognitive-motor disorders.
Introduction: ACh as a phasic modulator

Traditional descriptions of the anatomical organization of the basal forebrain cholinergic projections to telencephalic regions emphasize the hallmarks of a neuromodulatory system. These hallmarks include the presence of a relatively small number of soma in the basal forebrain giving rise to a relatively large innervation space, a limited topographical organization of cholinergic projections, a substantial degree of axonal collateralization, and the presence of extrasynaptic, or non-classical, receptors and, by implication, volume-transmission. Consequently, theories of cholinergic function have primarily described it in terms of slowly (over minutes) changing extracellular ACh levels (Yu and Dayan, 2002) and volume transmission (e.g., Lean et al., 2019). By these views, ACh acts in a spatially and temporally diffuse way to influence the excitability of widespread cortical target regions and thus primarily modulates relatively global functions such “arousal”. The main goal of this article is to critically probe these traditional descriptions, including our own prior interpretations of the evidence (Sarter and Bruno, 1997).

As an alternative, we discuss the evidence in support of the view that ACh mediates neuromodulatory effects based on highly phasic and probably largely synaptic signaling. This re-conceptualization of ACh signaling as phasic, synaptic, and behaviorally causal fosters the integration of diverse levels of analysis of cholinergic functions in rodents, non-human primates and humans, the development of computational models, and more effective approaches to the psychopharmacological development of pro-cholinergic treatments.

Anatomical foundations of locally-specific cholinergic signaling

Contemporary neuroanatomical research has revealed a heretofore unexpected degree of anatomical and functional parcellation of basal forebrain cholinergic neurons and a highly topographical organization of the basal forebrain cholinergic projection system, including complex relationships between basal forebrain afferent and efferent projection patterns (Zaborszky et al., 2008; Zaborszky et al., 2015b; Zaborszky et al., 2015a; Gielow and Zaborszky, 2017; Huppé-Gourgues et al., 2018; Lean et al., 2019). Combined with a limited degree of axonal collateralization (Price and Stern, 1983), this evidence suggests a neuronal projection system that can support regionally discrete cholinergic stimulation (see also Chavez and Zaborszky, 2017). The presence of neuronal subpopulations and topographic projections also supports proposals about cholinergic modules which can selectively impact information processing in individual cortical areas and layers (see also Tingley et al., 2015).
The relatively high density of cholinergic contacts, relative to axonal lengths and neuron number (Mechawar et al., 2000), would seem to significantly limit the spatial selectivity of cholinergic function. However, further differentiation of cholinergic actions may be derived from the presence of target area-specific organization of microcircuits, involving diverse and regionally-specific populations of interneurons (e.g., Xiang et al., 1998; Chen et al., 2015; Eggermann and Feldmeyer, 2009). Moreover, evidence indicating neuronal activity-dependent cholinergic modulation of dendritic computation (e.g., Williams and Fletcher, 2018), and region-specific wiring of cholinergic terminals, in part via heteroreceptors expressed at cholinergic terminals (Parikh et al., 2010; Parikh et al., 2008; Lambe et al., 2003; Poorthuis et al., 2013), offer additional mechanisms for differentiated, locally-specific cholinergic signaling. Thus, evidence at both the system and microcircuit level combine to render the view that ACh acts uniformly across large regions to, for example, “enhance cortical arousal”, increasingly obsolete.

The catalytic power of acetylcholinesterase (AChE) supports spatially and temporally constrained cholinergic signaling

The catalytic power of AChE has been called “amazing” and “a hallmark of an evolutionarily perfect enzyme” (Quinn, 1987). Indeed, the rate of ACh hydrolysis is limited by the rate of ACh diffusion to the active site, rather than by how quickly AChE can break it down (Botti et al., 1999; Hasinoff, 1982; Antosiewicz et al., 1995). AChE is present in the dendrites, perikarya, axons, and synaptic clefts (Blotnick-Rubin and Anglister, 2018). Thus, proposals suggesting extra-synaptic presence of “ambient” extracellular ACh levels, capable of reaching targets across tens of micrometers of extracellular space (Descarries, 1998), require mechanisms that limit the synaptic hydrolysis of ACh. Such an escape from hydrolysis has been proposed for ACh released from synapses with relatively large pre- to post-synaptic distances, based on the view that AChE is largely bound to presynaptic membranes (Dobbertin et al., 2009). However, the role of neuronally released, soluble forms of AChE (Andres et al., 1990; Appleyard, 1992) in terminating ACh action in vivo would also need to be considered. The finding that knockout of AChE in mice increased brain basal ACh levels from nanomolar to micromolar concentrations, but yielded only relatively minor functional impairments (Farar et al., 2012), may also be considered evidence against an essential role of AChE in terminating cholinergic signaling. However, little remains known about the compensatory role of other esterases capable of hydrolyzing ACh (but see Hartmann et al., 2007). Importantly, prior discussions in support of an ambient extracellular ACh level have relied largely on morphological evidence; what is needed are in vivo demonstrations that newly released ACh can escape hydrolysis.
We conducted one such test by measuring extracellular choline generation - a main product of ACh hydrolysis - with choline-sensitive electrodes. In addition, we (co-)immobilized AChE on these electrodes to hydrolyze ACh that potentially escaped, that is, was not hydrolyzed by, endogenous AChE (Giuliano et al., 2008). *In vitro*, these electrodes were able to detect “spared” extra-synaptic ACh at low femtomolar concentrations. In the cortex *in vivo*, we injected KCl into the vicinity of the electrodes to produce depolarization-evoked, relatively large increases of ACh release to optimize the possibility that a portion of such ACh might escape the endogenous AChE. However, even in such conditions, choline currents did not indicate that a portion of ACh “escaped” the endogenous AChE. In other words, these experiments did not reveal the presence of ACh spared by endogenous AChE.

Related to the presence or absence of ambient extracellular ACh levels, the presence or absence of classical cholinergic synapses (in cortex) has remained in dispute (Umbriaco et al., 1994; Descarries and Mechawar, 2000; Smiley et al., 1997; Turrini et al., 2001). However, if ACh indeed is nearly completely hydrolyzed by endogenous AChE, a significant degree of volume transmission would appear unlikely. Although burst firing patterns of basal forebrain cholinergic neurons may support increases in ACh release that continue for several seconds (e.g., Unal et al., 2012; Manns et al., 2000; Lee et al., 2005), for ACh to exert relatively distant effects, akin to effects of monoamines across several millimeters (Schneider et al., 1994; Puopolo et al., 2005), it would be necessary to demonstrate additional regulatory constraints of the efficacy of AChE. Conclusive experiments that could reject the presence of volume transmission do not appear straight forward, and raising such a binary question may not be very useful (see also Sarter et al., 2009). However, as discussed above, the cholinergic synapse seems exquisitely equipped to limit the spatial range of cholinergic signaling (see also Dunant and Gisiger, 2017). The recent demonstration that electrical stimulation yielded a very limited spread of activated (fluorescent) G-protein-coupled ACh receptors (Jing et al., 2018) is consistent with this view.

**Slow ACh - methodological artifact? New insights from amperometric recordings**

The view that levels of cholinergic neurotransmission vary across minutes has been supported by attributing relatively long-lasting (several minutes) arousal states to different extracellular ACh levels (e.g., Marrosu et al., 1995; Kametani and Kawamura, 1990). However, to a substantial degree, this view has been driven by the limited temporal resolution of previously predominant methods for monitoring changes in extracellular ACh levels. Using microdialysis to collect ACh from the extracellular space typically yields samples containing pM to low nM concentrations which are close to the detection limit of traditional analytical methods. Thus, it has been necessary...
to collect samples over several minutes. Moreover, such collections typically occurred while an AChEI was reverse-dialyzed to artificially increase levels of recoverable ACh\(^1\). In other words, ACh levels were long considered to vary at the scale of minutes because that was the scale at which they could be measured.

This view is challenged by experiments using newer methods that allow real-time monitoring of ACh release. Using amperometric measures of evoked choline currents, which reflect newly released and hydrolyzed ACh (Parikh and Sarter, 2006), we observed phasic, or “transient” cholinergic activity in the prefrontal cortex of rats performing a signal detection task. Such transients reliably predicted “switch hits” – correct signal detections following either a long temporal delay or a perceived nonsignal trial (i.e., after a correct rejection or miss (Parikh et al., 2007; Howe et al., 2013; Howe et al., 2017). These transients did not occur for other trial types, including correct rejections, misses, or hits following other hits (there were too few false alarms to analyze).

Critically, optogenetic studies (Gritton et al., 2016) demonstrated that cholinergic transients cause behavior: Optogenetic inhibition of such transients during signal trials reduced hits, but did not affect correct rejections, similar to the effects of cholinergic lesions (McGaughy et al., 1996). Moreover, optogenetic generation of cholinergic transients during cued trials, which therefore coincided with, or substituted the occasional absence of, endogenously-generated transients, increased detection rates (or hits). However, the most conclusive evidence for the causal power of cholinergic transients comes from the effects of optogenetically-generated cholinergic transients during non-cue (or blank) trials - in which normally no such transients are observed. Evoked transients in such trials drastically increased the rate of false alarms (incorrect reports of a signal) from around 20% to nearly 50% (Gritton et al., 2016). We further demonstrated that the behavioral power of cholinergic transients is due to the generation of high frequency oscillations in cortex, requiring muscarinic M1 acetylcholine receptor (mAChR) stimulation (Howe et al., 2017).

Further experimentation will be needed to disambiguate the precise computation driven by cholinergic transients. The task circumstances in which they have been demonstrated thus far –

\(^1\)If it is correct that the AChE effectively limits, or even prevents, the presence of extracellular ACh concentrations, it would need to be postulated that the successful recovery of ACh by microdialysis, in the absence of an AChE-inhibitor in the perfusion medium (Herzog et al., 2003; Himmelheber et al., 1998; Chang et al., 2006), results from the protection of ACh from the AChE by the glia barrier formed in response to the probe penetration injury (Jaquins-Gerstl and Michael, 2009; for more discussion of such technical issues see Sarter and Kim, 2015).
i.e., “shift-hits”, or signal detection after a long temporal delay or non-detection - suggest two possibilities: The first builds on decision theory and describes a noisy and imperfect balance between competitive “signal-absent” and “signal-present” representations of the current task context (Yu and Dayan, 2005). By this view, the cholinergic transient shifts the excitatory/suppressive balance away from the dominant ‘signal-absent’ context representation to the ‘signal-present’ one (Schmitz and Duncan, 2018). As an extreme experimental demonstration of this possibility, optogenetic generation of invalid cholinergic transients during signal-absent trials led to false alarms, i.e., incorrect reports of signal presence (Gritton et al., 2016).

The second view also emphasizes the (re)activation of the ‘signal-present’ taskset, but via a slightly different route. This interpretation starts from the observation that in humans performing the same signal-detection task as used in the rodent studies, shift-hits primarily activate a prefrontal region associated with switching from externally-oriented (monitoring) processes to internal processing (specifically, memory or task-set retrieval; (Burgess et al., 2005; Chun and Johnson, 2011; see Howe et al., 2013 for additional evidence that the fMRI findings related to shift-hits are cholinergically mediated).

The differences between these views are relatively subtle and careful experimentation will be required to differentiate between them (or other possibilities). However, they both replace the traditional view describing ACh in terms of functionality of relatively undefined variations in “states” related to presumed extracellular “levels” of ACh with more specific operations determined by the presence or absence of discrete cholinergic transients. Evidence indicating that phasic and precisely timed ACh release events are sufficient to produce cortical synaptic strength changes (Urban-Ciecko et al., 2018) that may be essential for the detection of attention-demanding cues are consistent with this proposal.

**Cholinergic “tone”: an intuitive, method-derived but unneeded concept?**

As already mentioned, the traditionally dominant view that ACh acts relatively slowly to influence widespread target regions has been based in part on evidence obtained by using microdialysis to monitor extracellular ACh levels. Data obtained from this method have necessarily suggested the functionality of slowly-changing levels of cholinergic tone (e.g., Coppola et al., 2016; Lecrux et al., 2017; Savage, 2012). Correlations between slowly-changing ACh levels with slowly-changing brain (arousal) states (e.g., Anaclet et al., 2015; Xu et al., 2015; Zant et al., 2016; Yang et al., 2017; Teles-Grilo Ruivo et al., 2017) have further supported the view that variations in “tonic” ACh levels are functional.
Above we argued that the cholinergic synapse is equipped to support highly phasic cholinergic signaling. This view raises the question of whether dialysate-derived tonic ACh levels reflect the integration of transients. Because the dimensions of the neurochemical measures obtained from microdialysis versus enzyme-coated microelectrodes and amperometry cannot be readily unified, and because the measurement compartments and terminal fields monitored by these two methods differ rather profoundly (microdialysis probe insertion-induced millimeter-sized cavity versus reactions of enzyme immobilized on a micrometer-sized, relatively slim electrode; (e.g., Fig. 2 in Howe et al., 2017), a direct test of this possibility has remained elusive. To complicate the issue further, the amperometric method is optimized for the measurement of transients and probably not capable of tracking slow changes in ACh (should those exist), largely because hydrolyzed choline spikes are rapidly cleared by cholinergic synapses and also diffuse into the interstitial space.

For a test of the possibility that dialysis-derived ACh levels represent integrated cholinergic transients, we measured choline currents using amperometry and ACh levels using microdialysis in (necessarily separate groups of) rats performing a cued appetitive response task, with long temporal delays between cues (60-120 s). In this task, amperometrically measured choline spikes occur in trials in which rats indicated behaviorally that they detected a cue which predicted subsequent reward delivery. Measures were obtained from prefrontal cortex and from motor cortex. To compare amperometric data with ACh levels measured in 8-min dialysate collections, we expressed both types of data dimension-free and collapsed transient amplitudes over 8-min periods (methods and results are detailed in Supplemental Data in Parikh et al., 2007). Statistical comparisons between these two data sets indicate the absence of a significant difference, suggesting that microdialysis levels were reproduced by folding transient data into time bins which matched the dialysis collection intervals.

Several caveats are important here. First, it should be acknowledged that we originally interpreted some aspects of this data, particularly the spatially-specific nature of cue-evoked transients (exclusively in mPFC and not motor cortex) versus the equivalent results for mPFC and motor cortex using either the microdialysis or re-analyzed amperometric data, as supporting different timescale mechanisms. However, although cue-evoked transients were confined to mPFC, amperometric activity did occur in motor cortex as well at various points in the trial – interestingly, the patterns suggested they may occur during shifts in motor behavior (e.g., from grooming to rearing). This leads to the second question of how transients could be integrated to lead to microdialysis results in light of the fast, highly-efficient action of AChE. As noted above, this may
be related to the glial barrier created in response to the microdialysis probe penetration injury (Footnote 1).

In short, the evidence that transient signaling is sufficient to describe forebrain cholinergic signaling is currently tentative but appears to be at least quantitatively possible. More critically, the evidence for longer-timescale action is methodologically problematic, and on first principles appears contradictory to the known efficiency of AChE. Definitive evidence on this point likely awaits further methodological development. However, to test the potential strength of a ‘phasic only’ conceptualization, below we assess the usefulness of this hypothesis in the context of evidence from two areas of research, or cases, on arousal states and on the impact of genetic variations of the synaptic capacity for cholinergic signaling.

**Case 1: Arousal states**

It has been widely accepted that forebrain cholinergic tone is elevated during REM sleep, and that ACh levels in that stage are nearly comparable with levels seen in the awake state. Indeed, evidence connecting arousal-states to ACh levels has remained a major source of support for the idea of a cholinergic tone. However, this may once again be at least partially a methodological artifact.

The primary evidence comes from classical studies which preceded even the availability of microdialysis. Sealed chambers were placed onto the pial surface of the cortex of anesthetized and immobilized animals and perfused with AChEIs to prevent ACh hydrolysis. Individual samples were collected over 10-15 min periods. ACh levels in these samples, in response to electrical stimulation of the reticular formation, formed the basis of the notion that arousing events increase cortical ACh levels (Celesia and Jasper, 1966; Szerb, 1971; Phillis, 1968). Subsequent microdialysis studies measured extracellular ACh levels in 5-60-min dialysate samples. Results from these studies seemed to confirm that ACh levels were higher during wakefulness and paradoxical sleep when compared with slow-wave-sleep (Kametani and Kawamura, 1990, 1991; Marrosu et al., 1995; Jimenez-Capdeville and Dykes, 1996). However, as noted above, both of these measurements occurred in the presence of AChE inhibitors in the perfusion fluid to prevent ACh hydrolysis.

Overall, the available evidence showing elevated cholinergic tone during arousal states seems unexpectedly limited and is largely based on older methods which relied on inhibiting ACh hydrolysis and which, by default, generated measurement time points incapable of revealing underlying potential phasic release patterns – patterns that are predicted based on the
neurophysiological activity of cholinergic soma during paradoxical sleep and wakefulness. In particular, recordings from cholinergic neurons in the basal forebrain indicate phasic, high frequency bursts during wake and REM sleep stages, that is, activity on a time scale that mirrors the time scale of transient ACh release events (Lee et al., 2005).

Consistent with these prior findings, we observed cholinergic transients, recorded at a sampling rate of 20 Hz in cortex and hippocampus, across all stages of the sleep/wake cycle but at a relatively higher frequency during REM sleep (Gritton et al., 2009). These transients had amplitudes of 5-40 pA and decay rates of 3-5 pA/sec. During REM sleep, the frequency of transients was about 4-fold higher than during slow-wave sleep (0.4 versus 0.1 transients/min), and also significantly higher than during wakefulness (0.25 transients/min). While such transient frequencies appear unexpectedly low, we observed non-correlated, or desynchronized, transients at recordings sites that were separated by only about 100 µm. This finding suggests that within a neuronal space of 500 µm³, approximating the space contributing to analyte recovery in microdialysis studies (Dykstra et al., 1992), transients during REM sleep may occur at a rate of over 10-50/min. Such a rate would be robustly higher than the rate observed during behavior (above) and thus could readily account for the elevated ACh levels seen in studies which used microdialysis or other low-temporal resolution methods to monitor ACh.

Case 2: Cholinergic top-down control – evidence for a relatively “tonic” action of ACh?

Thus far our description of the cognitive operations supported by cholinergic transients has focused on dynamic operations – shifts from one task or context representation (non-signal) to another (signal detection). However, successful cognition also requires the ability to maintain stability and stay “on task”, especially in the face of distractors or other challenges. The cholinergic system also plays a critical role in this aspect of cognition, one which we and others have previously ascribed to longer-term (seconds-to-minutes) cholinergic activity.

For example, right frontal and parietal ACh levels measured using microdialysis in rats performing the same signal-detection task used to demonstrate cholinergic transients (above) are elevated relatively to pre-task baseline and increase further in the face of a perceptual-attentional challenge (changing background illumination) that disrupts performance (St Peters et al., 2011; see Gill et al., 2000; Kozak et al., 2006 for additional evidence of the cholinergic system’s essential role in responding to challenge). Humans performing a parallel task show parallel increases in activation along the right middle/inferior frontal gyrus (Berry et al., 2017; Demeter et al., 2011). These increases in ACh levels and activation appear to be more strongly related to attempts to maintain or regain the task set, and thus performance, than with successful performance *per se* (see also
Gritton et al., 2013; Paolone et al., 2012). They have thus been described as related to “attentional effort”, or the motivated activation of attentional systems in order to stabilize or recover performance, especially in the face of challenge (Sarter et al., 2006; for evidence from other investigators and tasks reaching similar conclusions see Passetti et al., 2000; McGaughey et al., 2002).

Support for this interpretation also comes from humans with a genetic variant that reduces the capacity of the neuronal choline transporter (CHT) in vitro (Okuda et al., 2002) and, expressed in mice, choline clearance in vivo (Donovan et al., 2019). CHT capacity is essential for, and the rate-limiting step of, ACh synthesis and release (for reviews see Okuda and Haga, 2003; Ferguson and Blakely, 2004; Sarter and Parikh, 2005). We showed that the attentional performance of humans expressing this sub-capacity CHT variant is drastically impaired in the presence of a distractor (Berry et al., 2014; for review of evidence from humans and from a mouse model of impaired CHT function see Sarter et al., 2016). Additional support for a cholinergic role in “attentional effort” has been derived from investigations in patients with Parkinson’s disease with PET-based determination of cholinergic losses, in addition to the disease defining striatal dopaminergic degeneration. In these patients, reduced signal detection is associated with denervation of thalamic, rather than cortical, cholinergic pathways (Kim et al., 2017b). Cortical cholinergic denervation is associated instead with an increased vulnerability to irrelevant external stimuli (Kim et al., 2017a).

Together these data would seem to present a strong case for a dissociation between a “shifting” function associated with cholinergic transients, and a “stabilization” function associated with more sustained cholinergic firing, and indeed that was our initial interpretation (e.g., Sarter et al., 2001). However, the dissociation may be anatomical, rather than temporal. The fMRI activation patterns associated with shift-hits are observed in an anterior PFC region associated with retrieval and turning attention towards internal representations (see above). In contrast, those associated with responding to distraction and other attentional challenges occur along the right middle/inferior frontal gyrus, in a region frequently discussed as a “hub” for the network-level neural representation of relevant task sets, so that cognition and behavior are driven by these goal-relevant task sets, rather than being stimulus-driven (e.g., Braver et al., 2009; Lustig and Sarter, 2016; Berry et al., 2017).

Critically, maintaining representations in working memory - including task-set representations - does not require persistent neuronal firing (Lundqvist et al., 2018). Instead, they can be maintained by shifts in synaptic weights or coordinated variability and oscillatory behavior (e.g.,
Schmitz and Duncan, 2018; Lustig et al., 2007; Sadaghiani et al., 2015; Dehaene et al., 1998). Explicit activity may only be required during the initial acquisition, to recover the task set after an error, or to ‘protect’ the representation in the face of competing inputs (see especially the discussion in Dehaene et al., 1998), or more occasionally to ‘refresh’ the representation to counteract degradation in network coherence that would otherwise occur as a result of stochastic variability among its components (Lustig et al., 2009). Recent computational work demonstrates how cholinergic activity supporting the same fundamental operation – normalization, or shared variability among neurons – can support both stimulus and goal-driven attention by operating at different levels of the cortical hierarchy (Schmitz and Duncan, 2018).

In other words, the current evidence suggests that working memory representations – including those of the current task context – do not require constant, sustained neuronal spiking activity. In the absence of perturbation by external distractors or competing task sets, they can instead be maintained by correlated variability and shifts in synaptic weights, with occasional ‘refreshing’ needed to counteract stochastic variability that over time degrades their synchronization. The introduction of competing stimuli/task sets increases the need to “reinforce” the correct representation, but again this may be accomplished by short-burst firing – albeit at a more closely-spaced intervals. This view predicts that populations with low CHT function should have largely preserved, though somewhat less stable, performance in the absence of competition, with increasingly degraded performance with increasing salience and frequency of competing inputs – exactly the pattern shown by humans with genetically reduced CHT capacity and Parkinson’s patients with cortical cholinergic degeneration but relatively preserved thalamic cholinergic innervation (see above).

We recognize that the distinction between “closely spaced cholinergic transients” and “persistent neuronal firing” may be difficult to empirically discern (but see Cui and Strowbridge, 2019 for a neuronal mechanism via which cholinergic transients can induced persistent firing of cortical cells). However, there are critical conceptual distinctions: by this view the frequency of cholinergic activity is driven quantitatively by situational needs to refresh the task-set representation in the face of interference, rather than being a qualitatively different physiological “mode” operating at a different timescale (see also Fiebelkorn and Kastner, 2019).

**Conclusions**

Traditional assumptions about relatively lasting brain states controlled by the forebrain cholinergic system have coalesced with traditional neurochemical methods which generate minute-based measures of cholinergic activity and sample from relatively large neuronal spaces. The
widespread uses of AChEIs to optimize ACh measures and as a pharmacological tool have further cemented the view that tonic (scale of 100s of seconds) changes in extracellular ACh levels mediate relatively large-scale cognitive functions (such as arousal or top-down attentional control). Based on the demonstration of the presence and functions of fast, phasic or “transient” cholinergic signaling, here we argue that cholinergic signaling and functions can be sufficiently described by the presence of cholinergic transients which mediate a single computation that, behaviorally, favors the detection of behaviorally significant cues in attentional settings, specifically when such detection involves shifts between modes of attention (e.g., intrinsic to extrinsic, or monitoring to cue-oriented responding). The interpretation of evidence from behavioral, neurophysiological as well as human imaging studies on the role of cholinergic signaling will be more constrained and eventually heuristically more powerful by focusing on the role of fast cholinergic signaling for defined computational processes. Moreover, the search for effective pro-cholinergic, pro-cognitive treatments may benefit significantly from moving away from drugs the effects of which conform with views about tonic cholinergic activity and function, such as AChEIs, to drugs that enhance and rescue transient cholinergic signaling or their post-synaptic processing (e.g., Kucinski et al., 2019; Uslaner et al., 2018; Moran et al., 2018; Howe et al., 2010).
References


