Effects of Guanfacine and Phenylephrine on a Spiking Neuron Model of Working Memory

Peter Duggins (psipeter@gmail.com)
Terrence C. Stewart (testewar@uwaterloo.ca)
Xuan Choo (fchoo@uwaterloo.ca)
Chris Eliasmith (celiasmith@uwaterloo.ca)
Centre for Theoretical Neuroscience, University of Waterloo
200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

Abstract
We utilize a spiking neural network model of working memory (WM) capable of performing the spatial delayed response task (DRT) to investigate the functional effects of two drugs that affect WM: guanfacine (GFC); and phenylephrine (PHE). In this model, the loss of information over time results from changes in the spiking neural activity due to recurrent connections. We reproduce the standard forgetting curve, then show that this curve changes in the presence of GFC and PHE, whose application is simulated by manipulating various neuron properties. In particular, applying GFC causes increased firing in neurons that are sensitive to the information currently being remembered, while applying PHE leads to decreased firing in these same neurons. Interestingly, these memory-specific effects emerge from network-level interactions, because GFC and PHE affect all neurons equally. We compare our model to both electrophysiological data from neurons in monkey dorsolateral prefrontal cortex and to behavioral evidence from monkeys performing the DRT.

Keywords: working memory; delayed response task; guanfacine; phenylephrine; Neural Engineering Framework

Introduction
Working memory (WM) is a central component of cognitive systems which use it to temporarily store information during the execution of complex tasks. Models of WM differ greatly between contemporary cognitive architectures, leading to diverse predictions about how information is represented and altered over time. Because WM is biologically realized in networks of neurons, one goal for researchers studying WM is to understand how networks of spiking neurons implement information storage and retrieval in the brain. In so doing, such models can be used to characterize deficits of WM associated with mental disorders, such as attention deficit hyperactivity disorder (ADHD) and Tourette's syndrome (Scanhill et al., 2014), and be used to understand the biochemical mechanisms behind drugs used to treat such deficits (Avery, Frannowicz, Studholme, van Dyck, & Arnsten, 2000). Due to the complexity of the interactions involved, few studies have characterized the relationships between drug chemistry, neurobiology, and cognitive abilities, including working memory.

In this paper we present a spiking neural network model of WM and action selection applied to a mnemonic cognitive test, the delayed response task (DRT). Computational models are well-suited to investigate multilevel interactions, including those between drugs that alter the brain’s biochemistry and the resulting disruptions in cognitive abilities. We construct such a model using the Neural Engineering Framework (NEF) (Eliasmith & Anderson, 2003), a general method for building cognitive models from spiking neurons. The NEF has previously been used to create biologically-constrained models of list memory (Choo & Eliasmith, 2010) and action selection (Stewart, Choo, & Eliasmith, 2010) that are consistent with neural and behavioral data. This paper extends these models by simulating the effects of two drugs, guanfacine (GFC) and phenylephrine (PHE), which enhance and inhibit WM respectively.

In the next sections we describe the biological and computational basis of WM in the brain, examine the biophysical mechanisms of the applied drugs on neural activity, and advance a hypothesis for the relationship between them. We then present our model, describing how information is stored, forgotten, and retrieved in the delayed response task. When GFC (PHE) is applied to the model, we observe a shifted firing rate in those neurons whose spatial mnemonic tuning (preferred space/time direction) is aligned with the cue’s location. This in turn affects the value stored in WM, leading to an increase (decrease) in performance on the DRT. The magnitude and timing of this effect is comparable to empirical data from monkeys. We conclude by proposing biophysical and anatomical extensions of the model.

Biological Background
WM is realized in the prefrontal cortex (PFC), a brain region whose prominent size in highly-evolved primates suggests its importance in complex cognitive tasks that require a flexible mental workspace. The PFC represents information that is temporarily held in mind and used to guide behavior and decision-making, and is thought to be maintained through recurrent excitatory connections between neurons with similar tuning properties (Goldman-Rakic, 1995). Computationally, this recurrence realizes an extended temporal integration that preserves the represented item without external stimulation (Singh & Eliasmith, 2006).

The stable representation of items stored in WM is particularly sensitive to the synaptic connections of intra-PFC loops and the biochemical environment of PFC neurons. Drugs that are used to treat WM disorders such as ADHD and Tourette’s Syndrome target these biophysical mechanisms and have been shown to affect WM in healthy animals (Avery et al., 2000; Scanhill et al., 2014). For example, guanfacine (GFC), an agonist for the α2A-adrenoreceptor, influences
WM in PFC neurons expressing Hyperpolarization-activated Cyclic Nucleotide-gated (HCN) ion channels (Franowicz et al., 2002). At rest, HCN channels permit the influx of non-specific cations, but deactivate in response to depolarization. HCN channels are unevenly distributed along the dendritic tree, with an almost sevenfold increase in density from the soma to distal end of the dendrites. These properties allow HCN channels to reduce the temporal variability of dendritic excitatory postsynaptic potentials (EPSPs) that exists due to spatial distribution along the dendritic tree (Magee, 1999). It is believed that these channels control the excitability of pyramidal neurons in PFC by modulating the temporal dendritic summation and resting potential (Poolos, Migliore, & Johnston, 2002).

GFC, acting through a cAMP-mediated intracellular signalling cascade, closes HCN channels, resulting in less damping of excitatory dendritic spikes and increasing the overall excitability of the neuron. A study by Wang et al. (2007) showed that GFC increased the firing rate of PFC neurons with weak mnemonic tuning in the direction of spatial cues on the delayed response task, while having no effect on cells tuned in the opposite direction. Similarly, the α2A-adrenoceptor antagonist PHE opened HCN channels and decreased firing rates of preferred-direction cells. These results are consistent with increased (decreased) behavioral performance on the delay-response task (Mao, Arnsten, & Li, 1999; Ramos, Stark, Verdutzo, van Dyck, & Arnsten, 2006). We hypothesize that GFC raises the firing rate of spatially tuned neurons, causes a slower decay of items stored in PFC neural integrators, induces lower rates of forgetting, and consequently increases performance on the delay-response task. We test this hypothesis in a spiking neuron model.

A Spiking Neuron Model of Working Memory

The core requirement in a neural model of WM is a population of neurons that can maintain its state over time. That is, given a brief input, the internal connectivity should cause the neural activity pattern that results from that input to persist after the input has stopped. This persistence will not be perfect – over time the neural activity will drift away from its initial value.

However, this population of neurons cannot maintain any possible pattern of firing: we expect there to be correlations in the structure of this neural activity. Indeed, it has become common to analyze neural activity in WM areas (and elsewhere in the brain) by performing dimensionality reduction through techniques such as jPCA (Shenoy, Sahani, & Churchland, 2013). These approaches characterize the underlying patterns of correlation between the spiking neurons, identifying a lower-dimensional subspace that the neural activity represents. That is, rather than treating each neuron independently, we assume there is some vector x that is being represented by the population of neurons. The dimensionality of this vector is much smaller than the number of neurons, which means the information is redundantly encoded across these neurons. In particular, each neuron i will have some particular vector e_i for which that neuron fires most strongly (these are often known as “preferred direction vectors” or “encoders” and have been widely used as a useful way of characterizing cortical activity (e.g., Georgopoulos, Kalaska, Caminiti, and Massey (1982)). We can consider the total overall current going into a neuron to be proportional to e_i · x (the similarity between x and the preferred vector e_i). To produce a variety of tuning curves and firing rates that matches those in PFC, we randomly chose a gain α_i and bias current β_i for each neuron, resulting in a total input current of α_i e_i · x + β_i. This current can be fed into any neuron model, but here we simply use the standard leaky integrate-and-fire (LIF) model.

Given that the neural spiking activity encodes some vector x, it should be possible to recover that information by observing the spikes. The simplest method is to “decode” this spiking information via a weighted sum of the spikes, such that ˆx(t) = ∑ a_i(t) d_i · h(t), where a_i(t) is the spiking activity of the ith neuron, h(t) is the shape of the post-synaptic current caused by the spikes, and d_i is the weighting factor for each neuron. The decoder (i.e., d_i) values can be found by performing a least-squares optimization that minimizes the difference between x (the original vector) and ˆx (the vector recovered by observing the spiking activity). This method of characterizing neural representation is the first principle of the Neural Engineering Framework (NEF) (Eliasmith & Anderson, 2003).

Now that we have defined how a population of neurons can represent a value x, we can construct recurrent connections within this population such that the neural activity continues to represent x over time. To realize such a WM, we must find recurrent connection weights that stabilize dynamical neural activity, regardless of the value x being represented. Using the third principle of the NEF, this can be characterized as another least-squares minimization problem: previous work has shown that the optimal weights from neuron i to neuron j are w_ij = α_j e_j · d_i (Eliasmith & Anderson, 2003). The result is a population of spiking neurons that maintains its activity over time, and has been the basis of multiple WM models (Singh & Eliasmith, 2006; Choo & Eliasmith, 2010).

To simulate the WM component of the delayed response task, we let x be two-dimensional, where the first dimension is the value to be remembered, and the second dimension is the amount of time it has been remembered for. Empirical and modeling evidence are consistent with the claim that PFC neurons explicitly encode the passage of time (Lewis & Miyall, 2006; Bekolay, Laubach, & Eliasmith, 2014; Singh & Eliasmith, 2006). For example, some PFC neurons start firing only after a given amount of time has passed, while others gradually decrease their firing rate over time (Romo, Brody, Hernández, & Lemus, 1999). These “positive monotonic” and “negative monotonic” neurons can be thought of as neurons that are sensitive to both the value being represented and...
the amount of time the memory has been held; in other words, these are spatial mnemonic neurons whose \( e_i \) values are large for both the first and second dimension. Other neurons may only be sensitive to one or the other dimension (i.e. would have small \( e_i \) values for one of those two dimensions). This variability in \( e_i \) matches well to the observed variability in WM tuning curves (Singh & Eliasmith, 2006).

**Variability and Drug Effects**

The WM model used here is based on that in Singh and Eliasmith (2006), with the addition of randomly varying background current to each neuron, to reflect the stochastic variability found in the brain. Without this random “noise”, the information stored in WM is stable for a very long time (minutes to hours). However, with a small amount of background current added, the memory decays over tens of seconds as shown in Figure 1, consistent with decay rates of human WM (Choo & Eliasmith, 2010).

We use this model to investigate how WM is affected by the drugs GFC, which increases the excitability of neurons, and PHE, which decreases excitability. We simulate their effects using two alternative methods which simplify the aforementioned biophysics while maintaining the core functional properties in the NEF. In the first method, we model excitability as a global increase (or decrease) in somatic current to all WM neurons. Importantly, even though Wang et al. (2007) showed that, in vivo, an increase in firing activity was only observed for neurons whose preferred direction was aligned with the stimulus being remembered, we do not apply this extra current only to those neurons. This is because there is no direct mechanism by which GFC or PHE could affect only those neurons that are actively encoding information. Rather, we apply the simulated drug effect to all the neurons in the WM model. While this seems counter-intuitive, we show below that when we simulate this system, the network effects of the recurrent connections are sufficient to cause the observed differential response (Mao et al., 1999).

In the second method, we attempt to more faithfully reproduce the biophysical effects of HCN channel closure by manipulating the neurons’ internal properties. HCN channels allow positive ions to flow into the cell, so closing HCN effectively induces a negative current, lowering the resting membrane potential. We model this effect by lowering the bias current \( b_i \) of each neuron in the WM. Additionally, closing HCN channels modulates neurons’ dendritic summation, such that small, desynchronized dendritic spikes more strongly influence the somatic membrane potential. This effectively increases neurons’ response to a given synaptic input, which we model by increasing the gain \( g_{ii} \) of each neuron. We calibrate the competing effects of these manipulations using data from Nolan et al. (2004), which compares the subthreshold voltages of HCN-knockout mice and normal mice as a function of input current, Figure 2.

**Modeling the Delayed Response Task**

In the spatial delayed-response task (DRT), monkeys are presented with a brief (1s) visual cue positioned relative to their fixed gaze. The cue is removed. During the delay period (2, 4, 6, or 8s), the monkey stores the cue location in WM, then recalls that location in the response period by pressing the corresponding button or making a saccade. In terms of our model, the cue is considered to be a numerical value between -1 and 1 (the first dimension of the vector \( x \)). This value is fed into the model by directly injecting current into the WM neurons, causing them to spike with frequency determined by the similarity between their preferred vector \( e_i \) and the represented value \( x \), computed as \( e_i \cdot x \). This external current is injected for the duration of the cue period (1s) then removed; after this, the memory must be maintained by activity fed back through the WM recurrent connections.

To produce a response, the model must access that stored value and produce one of two outputs (-1 or +1). While a mechanism to perform this is straightforward to design with the NEF (Sharma, Kromer, Stewart, & Eliasmith, 2016), this part of the model does not alter the drug effects, so for simplicity we do not consider it here. Instead, we take the neural activity of the WM neurons and compute their weighted sum,
giving an estimate of the original value \( \hat{x}(t) = \sum_i a_i(t)d_i * h(t) \). Since a neural mechanism to convert this value into a decision will include some degree of variability, we approximate this by adding normally distributed noise to this value. If the result is above zero we interpret this as the model giving the first response, and if it is below zero we interpret it as giving the second response.

### Results

To simulate the cellular effects of GFC and PHE, we tested two methods for perturbing the neurons, as described above. In the first, we injected a noisy signal\(^2\) into neurons in the WM population, essentially using additive bias to increase (decrease) neural excitability. In the second, we manipulated the gains and biases of the LIF neurons used in the simulation, effectively decreasing each neuron’s resting potential while increasing its gain to synaptic inputs\(^3\). Both perturbations produced the desired effects; we hereafter report results from the first method.

We began by comparing delay-related neural firing rates in the WM population\(^4\) with activity from neurons in monkey dorsolateral PFC (Wang et al., 2007). We selected model neurons that were tuned to the preferred direction during control conditions, as per their hypothesized importance in representing the cue’s location during the delay period. Wang et al failed to provide a precise definition of “weak spatial mnemonic tuning” or their procedure for choosing such neurons, so we selected model neurons based on the magnitude of their encoders (\(e\)), the change in their firing rate when presented with preferred-direction stimuli (\(da/dt\)), and their differential response to drug application (\(\Delta a\)). Figures 3 and 4 show the normalized firing rate of neurons before and after the simulated application of GFC and PHE. Both empirically and in simulation, GFC increased the firing rate of preferred-direction neurons while having little effect on neurons in the nonpreferred direction. Similarly, PHE decreased the firing rate of preferred, but not nonpreferred, neurons.

Next, we investigated whether the firing rate of preferred-direction neurons encoded the location of the cue stored in WM. Using the NEF, we decoded, from the neural activities, the value stored in the WM during the delay period. As the model forgot the original stimulus, this value decayed exponentially. In response to GFC (PHE), and concurrent with the increased (decreased) firing rate of preferred neurons, the WM value decayed less (more) rapidly, Figure 5.

Lastly, we tested whether the value stored in the WM coincided with the accuracy of the model on the DRT. Figure 6 shows the likelihood of correct response as a function of delay period length for a one-dimensional DRT (left-right cues). Both for monkeys (solid line) and the model (dashed line),

\[^2\] Normally distributed and proportional to the maximum firing rate, \(N(0.002, 0.09)\) for GFC, \(N(-0.002, 0.09)\) for PHE

\[^3\] GFC: \(\beta_{pre} = 0, \beta_{post} = -0.04, \alpha_{pre} = 1.00, \alpha_{post} = 1.036\); PHE: \(\beta_{pre} = 0, \beta_{post} = 0.046, \alpha_{pre} = 1.00, \alpha_{post} = 0.960\)

\[^4\] \(N = 3000\) neurons, neuron noise \(\sigma = 0.009\), synaptic time constant \(\tau = 0.1\), dimension \(D = 2\) (stimulus, time).
Figure 5: Value stored in WM during the delay period. Applying GFC (PHE) results in higher (lower) neural firing rates that shifts the curve up (down), altering the model’s ability to distinguish the represented value from zero following the delay period. Reported values are averaged over $N = 50$ realizations with confidence intervals plotted in gray.

accuracy decreased steadily from 2-6s then dropped sharply at 8s\(^5\). Our model shows increased (decreased) performance on the DRT following application of these drugs that fits the baseline empirical data with root mean-squared error between 0.0001 – 0.005. The accuracy dropoff occurs when the value stored in the WM become indistinguishable from zero to the model’s noisy decision procedure following the delay period.

**Discussion**

In this paper, we presented a minimal model of WM applied to the DRT that reproduces neural spiking and behavioral results under various drug manipulations with surprising accuracy. The model extends classical works on WM dynamics and the effects of neuromodulation (Brunel & Wang, 2001) by (a) incorporating the NEF, an approach that allows for the principled decoding of information represented in large-scale spiking neural networks, and (b) demonstrating that neuromodulation of WM (and its behavioral effects) can be studied through simple manipulation of neuron properties, bypassing the need to build complex circuits using Hodgkin-Huxley-type neurons.

Future work will address several simplifying assumptions made in the study. First, a detailed sensitivity analysis would reveal the robustness of the model to parameter variation. Exploratory experiments showed the decision noise and synaptic time constant altered the shape of the recall curves and increased the RMSE, but a more systematic investigation is needed to discover interactions between the remaining free parameters.

Second, the use of LIF point neurons to represent delay-related activity in WM necessitated an approximation of HCN opening and closing. Surprisingly, we found that both simple manipulations (biasing neurons or increasing their gain) produced changes in firing rate and behavioral response that match the empirical data. This suggests that biophysical simulations of drug-neuron interactions may be unnecessary, so long as the qualitative effect of the drugs on firing rate can be discerned from electrophysiological data. That said, replacing LIF neurons with more detailed neurons that include explicit HCN channels (which can be closed or opened by GFC or PHE) would expand the range of biochemical processes we could simulate. To progress in this direction, we have integrated the NEURON simulation package with the NEF-style modeling performed here.

Finally, while our model focused on the representational aspects of WM, the processes by which information is placed in, and retrieved from, WM are equally important for its implementation in unified cognitive systems. Adding input and output neural subsystems to the model, such as a visual hierarchy and a basal ganglia, would greatly expand the range of cognitive tasks that our model could potentially perform, avoid the use of arbitrary parameters such as the “misperception probability” and “decision noise”, and present new targets for drugs that affect different aspects of cognition.\(^6\) Many of these systems have already been built using the NEF (Eliasmith et al., 2012). In future work, we plan to implement both these extensions in pursuit of a deeper understanding of the neural basis, psychological dysfunction, and pharmaceutical modulation of working memory.

---

\(^{5}\)To capture the inaccuracy of monkeys after a 2s delay, we introduced a 7% chance of misperceiving/ignoring the stimulus.

\(^{6}\)For example, dopamine (D1) receptors are present both in PFC and hippocampus, and abnormal neurotransmitter/receptor levels have been implicated in WM deficits related to Parkinson’s and schizophrenia (including performance on the spatial DRT).
Acknowledgments
This work was supported by CFI and OIT infrastructure funding as well as the Canada Research Chairs program, NSERC Discovery grant 261453, ONR grant N000141310419, AFOSR grant FA8655-13-1-3084 and OGS graduate funding.

References