Serial order and working memory in a spiking global neuronal workspace model

CN720 Final Project
Sean Lorenz
December 11, 2007

1. Introduction

Lashley's groundbreaking article, "The Problem of Serial Order in Behavior," showed that associative chaining models for serial order can be problematic, therefore other alternatives must be explored. In lieu of chaining models, competitive queuing (CQ) models (Grossberg, 1978; Henson et al., 1996; Rhodes et al., 2004; Cooper & Shallice, 2006) have provided a feasible alternative. CQ models propose that the typical "sawtooth" errors seen in sequencing tasks such as Immediate Serial Recall (ISR) or the 2xN task can be explained by way of a primacy gradient of activation across list item representations rather than by associations between these items. The two basic premises of CQ models are: 1) several plan representations can be active in a planning layer at the same time, and 2) the most active plan representation is chose in a competitive choice layer which also decides serial order of list item recall. Although this relatively new approach to modeling serial recall tasks could accurately replicate error trends from experimental data, it was not until recently that physiological proof for CQ was demonstrated (Averbeck et al., 2002).

In this paper I will attempt to explain Averbeck's results from the global workspace theory (GWT) perspective. More specifically, I plan to correlate CQ models of working memory and sequencing tasks, like the ones reference above, with recent advances in biologically-plausible models based on the GWT framework. I will begin with a general explanation of GWT and also elaborate on some of the more interesting, pertinent work being done in this field, paying special attention to the new spiking neural models of Dehaene and Shanahan. Averbeck's results can then be interpreted within a global neuronal workspace (GNW) perspective of how to execute a sequencing task. Lastly, a proposal will be discussed for applying Shanahan's general spiking cortical model to working memory areas of the human brain as well as methods for simulating such a model within a 3D robot simulation environment.

2. Global Workspace Theory

GWT has undergone several stages of growth since the foundational cognitive groundwork of Bernard Baars two decades ago (Baars, 1988 & 1997). Some of these advances include Stan Franklin's more abstract, module-based software application of Baars' theory (Baars & Franklin, 2003), Murray Shanahan's robotic implementation with biological analogues (Shanahan, 2006 & 2007), as well as Stanislas Dehaene's physiologically-based cortical models of cognitive function (Dehaene et al., 1998 & 2003). Much of the recent work in GWT is based in actual brain area research using more realistic spiking network models, thus I will be referring to this more specific form of GWT as the global neuronal workspace (GNW).

Before going into further depth concerning recent research, I will first give a brief overview of GWT's primary tenets. Albeit fashioned as an explanation for "consciousness" – and I use the scare quotes intentionally – GWT's strength comes from its core hypothesis of linking *conscious* and *unconscious* processes in the brain. For Baars, conscious experience is best represented using a theater stage metaphor. Working memory serves as a stage for actors to perform (the contents of conscious experience), where the spotlight is seen as a focal point of attention on the stage. The audience, then, behaves as a system of unconscious networks which offer their expertise as to what actor should remain in the spotlight (Baars, 1997). These conscious experiences act serially whereas unconscious actions act in parallel, illustrating the competitive nature of neuronal networks (and biological systems in general for that matter).

At the heart of Baars' theory is the idea that conscious problem-priming depends upon unconscious problem-solving which loops back to conscious solution display. In other words, unconscious systems are working hard to return conscious answers based on "working residues" of earlier conscious thoughts. The theater image, then, is meant to be a general metaphor for understanding direction of information flow in the brain. GWT also acts as a method for relating autonomic, subcortical, or pre-percept cortical function with volitional, neocortex function.

Lastly, the concept of a global neuronal workspace steers clear of the homunculus fallacy since there may be hundreds of separate workspaces nodes distributed throughout the brain which interact in a variety of manners. Executive function areas such as those found in prefrontal cortex (PFC) may have stronger connection activities projecting to other cortical or subcortical areas, however, PFC is still just one of many interacting workspace areas.

An overarching theory such as GWT gives a general plan for approaching the study of interacting brain area function, which has been developed into a computationally-based cognitive schematic by Franklin. The most recent implementation from Franklin is LIDA (Learning Intelligent Distribution Agent), a working model of machine consciousness based on GWT, which has associative perceptual memory, workspace, episodic memory, functional consciousness (as opposed to phenomenal consciousness), procedural memory, and action selection mechanisms. Although these modules are not directly related to biological function in his model, they nicely summate the procedures needed to produce action or accurately model cognitive processes such as working memory as seen in Figure 1. In order to get a more biological GWT perspective, however, we will need to look at the work of Dehaene as well as that of Shanahan.

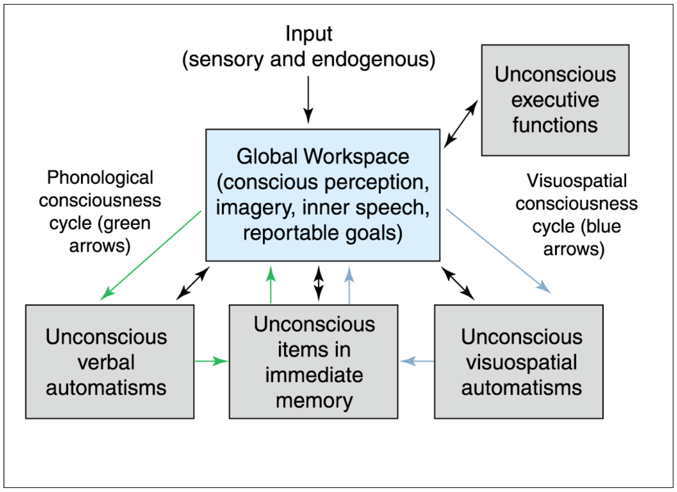

Figure 1: Relation of conscious and unconscious elements in working memory processing. Phonological (green arrows) and visuospatial (blue arrows) consciousness cycles (Baars & Franklin, 2003).

3. Dehaene and Shanahan Models

Although Baars continues to publish on conscious experience and working memory, GWT has also progressed under Dehaene (and collaborators such as Jean-Pierre Changeux and Claire Sergent) at the INSERM-CEA Cognitive Neuroimaging Unit in Paris. Dehaene claims "*recurrent* interactions between distal brain areas as a necessary condition for conscious perception (Dehaene et al, 1998; italics mine)." These areas process a stimulus dynamically in the workspace which has recurrent connections to distal areas, allowing "auto-amplification" of the activation. If a primary stimulus activation exceeds a certain threshold, reverberation takes place and the current stimulus can gain access to the workspace, which in turn will permit broadcasting to a wide range of brain areas. This is meant to help explain cognitive control tasks such as verbal report, voluntary manipulation, voluntary action, and long-term memorization. The key concept here is that a GNW predicts an *all-or-nothing* transition between conscious and unconscious action.

In the same 1998 paper, Dehaene lays out five major processor categories – along with their neural correlates – which interact with the workspace neurons:

1. perceptual circuits (present stimulus): object-oriented ventral and lateral temporal lobe areas, temporal and inferior parietal areas, and Wernicke's area;
2. motor programming circuits (future intentional behavior): premotor cortex, posterior parietal cortex, supplementary motor area, basal ganglia (caudate nucleus), cerebellum, and inferior frontal lobe (Broca's area);
3. long-term memory circuits (past percepts and events): hippocampal and para-hippocampal areas distributed throughout cortex according to context/modality;
4. evaluation circuits (positive /negative value assessment): orbitofrontal cortex, anterior cingulate , hypothalamus, amygdala, ventral striatum, mesocortical catecholoaminergic and cholinergic projections to prefrontal cortex; and,
5. attentional circuits (focusing): mobilized workspace circuits independent of the external world, i.e. overt behavior versus covert attention.

In addition to these five parallel processors dispersed throughout the brain, the global workspace exists as a set of cortical neurons in numerous regions that send excitatory, long-range, horizontal cortico-cortical projections (typically from layers 2 and 3 pyramidal cells) to other distal workspace cortical neurons. Vertical connections within a cortical workspace column are also reciprocally connected to layer 5 neurons. Dehaene also notes that the amount of pyramidal neurons in layers 2 and 3 of are especially dense in dorsolateral prefrontal cortex (DLPFC), a region long-known for its connection to working memory, as well as inferior parietal cortex (IP).

From this information, a neuronal architecture was built which included stimulus inputs, the five processors, workspace neurons and two generalized workspace input systems – vigilance and reward. This network, shown below in Figure 2, was simulated in order to see if it could accurately predict accurate results using the word-color Stroop task as a basis for experimentation. In the first task, the network was rewarded for turning a color unit on. For the second task, the network was rewarded for turning on a unit appropriate to the word, but not the ink color. For the final task, conflicting word and color inputs were introduced to the network, however, for this task, connections to and from the workspace unit were required unlike in the first two tasks which relied on the specialized processors.

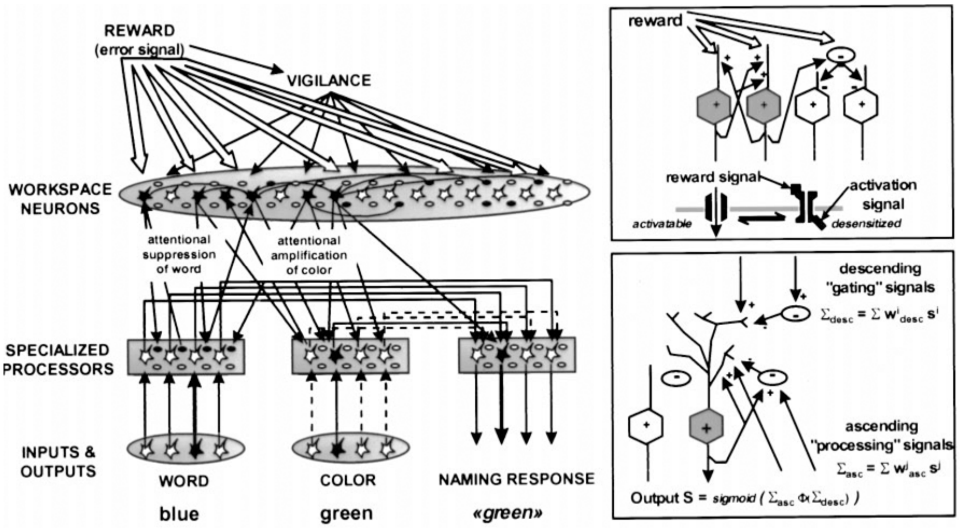

Figure 2: GWT neuronal architecture used to simulate an effortful Stroop task (Dehaene et al. 1998)

The same research group has since built upon their neuronal workspace by switching to an integrate-and-fire modeling method, applying this new schema to behavioral tasks such as the attentional eye blink (Dehaene et al., 2003). Their conclusions state that attentional blink in humans involves nonlinear dynamics of bottom-up and top-down reverberation, thus a GNW formed by DLPFC (among other areas) acts as a “bottleneck” which cannot process a second target if the stimulus time interval is small enough.

Shanahan has also built a cognitive architecture built on GWT which contains an internal sensorimotor loop used by a Webot robot controller to perform tasks in a simple virtual environment (Shanahan, 2006). Unlike Franklin, however, Shanahan built into the architecture several brain area analogues such as the motor and sensory cortices, association cortex, basal ganglia, and amygdala. The model includes a first-order, reaction-based system which is a loop closed by the outside world. In addition to this loop, a second-order system consisting of the basal ganglia with its "go" or "no-go" functionality (among other areas including motor and assocation cortices), is required for *selective* action. The Webots robot used this minimal "brain" to choose either to roll forward, rotate left, or rotate right when its tunnel-vision camera encountered cylinders of different colors. With the higher-order system, the robot does not make automatic motor movements to look at the different cylinders, rather, it can be vetoed by the basal ganglia or amygdala pre-wired associations with the different stimuli. What is most interesting about this experiment is its departure from most current robotics research, replacing symbolic reasoning with a "recurrent cascade of attractor networks."

Although this system is, I believe, a giant leap ahead for a biologically-based version of cognitive control than that presented by Franklin (if mimicking brain behavior is one's end goal), the Shanahan model described above still requires a switch to a more neurobiological focus. For this reason, following the lead of Dehanene and his colleagues, Shanahan has also switched to a spiking neuron GNW model in order to better understand cortical broadcast and competition (Shanahan, 2007; in press). Instead of using the standard integrate-and-fire neuron model for simulation, Shanahan adapted Izhikevich's simplified Hodgkin-Huxley model of spiking behavior.

In this updated proposal, focus is placed on interaction among different cortical columns which possess "workspace nodes" – a small set of neurons within a column that project to other workspace nodes in other cortical areas. In this way, the network of long-range corticocortical connections throughout the brain is proposed to be the global neuronal workspace. Thus, each workspace node can broadcast its information to other cortical regions as proposed by Baars. Figure 3 shows how Shanahan's goal here is to build upon the research of Dehaene et al.; Figure 4 shows how the workspace nodes interact.

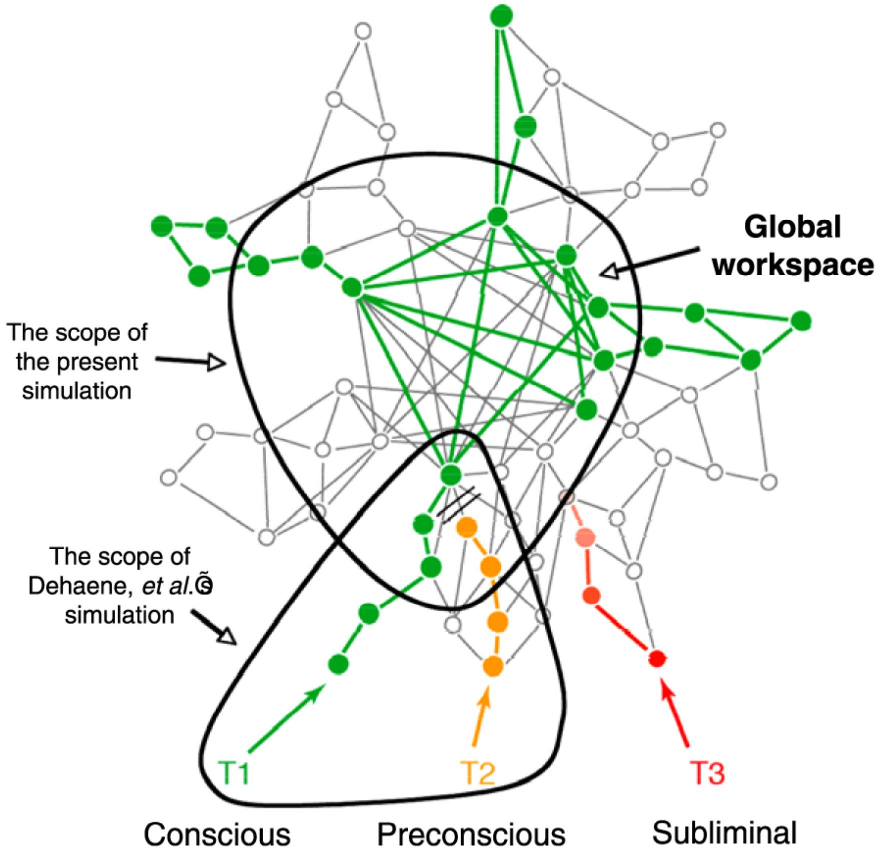

Figure 3: Shanahan's adaptation of the GNW model of Dehaene et al., 2006

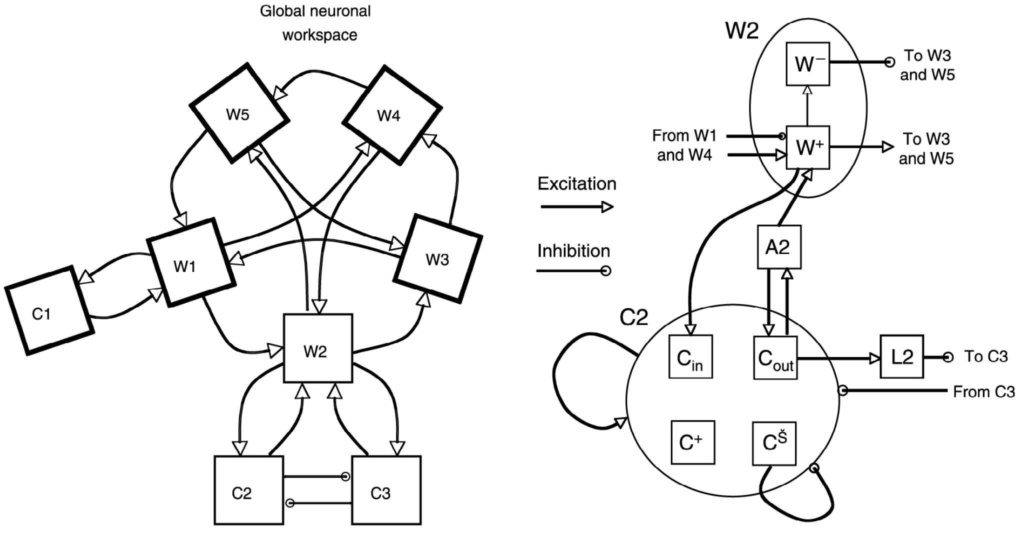

Figure 4: On the left – the model with five workspace nodes (W1 – W5) along with interaction between non-workspace cortical column areas (C1, C2-C3); on the right – close-up W2 and C2 interaction.

To be brief, Figure 3 shows how, rather than solely model unconscious competitive access to the workspace as in Dehaene et al., 2006, Shanahan elaborates on this idea by modeling workspace node reverberation in addition to unconscious competition for the stage spotlight. Figure 4 shows how the workspace nodes interact in different cortical areas. It also details the interaction between workspace neurons and non-workspace neurons within a single cortical area (as shown in the W2 – C2 detailed interaction). In this model of sequence learning and retrieval, Shanahan does not make reference to certain cortical areas yet it seems likely that a GNW framework such as this can be applied as a general principle for long-distance corticocortical interaction.

GNW and Averbeck serial order data

Now that a general view of GNW has been presented, I will relate this type of theory to the experimental results of Averbeck et al.'s 2002 paper which gave physiological plausibility to Lashley's notion of parallel processing in serial order tasks. The researchers trained two monkeys to produce geometric shapes (a triangle, square, trapezoid, and inverted triangle) using a joystick to control an on-screen cursor, then recorded from 511 neurons in Brodmann's area 46 of the prefrontal cortex. The task and results shown in the neural activity can be seen below in Figure 5.

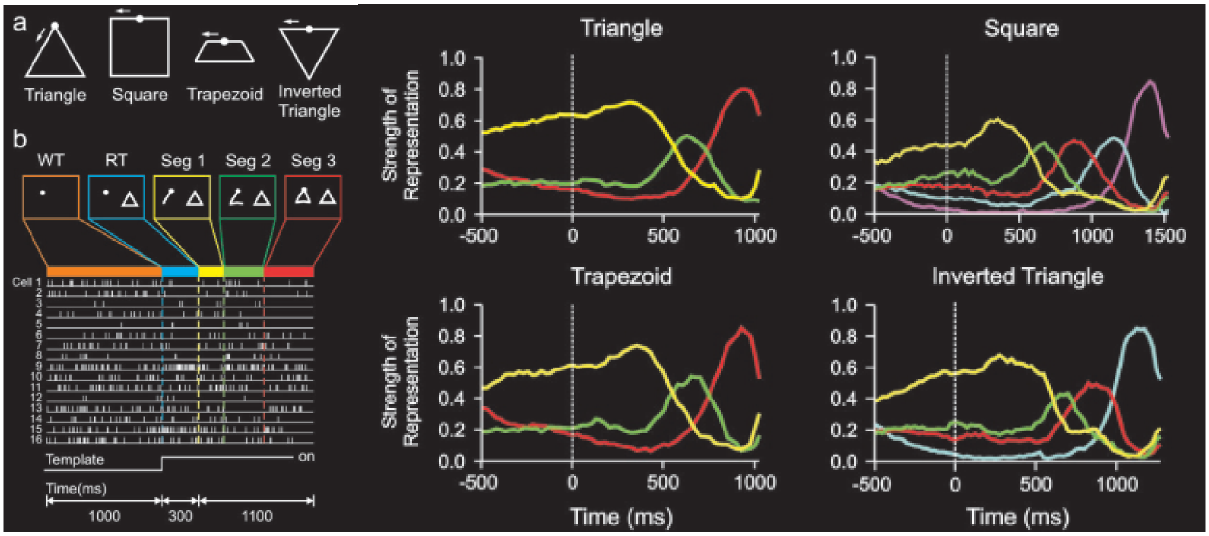

Figure 5: Averbeck et al. 2002 results from 16 independently-driven microelectrodes (right) in area 46

for the triangle shape; strength of representation in this area over time for four different shapes (left).

Figure 5 demonstrates that *rank order* of strength representation for each segment of the triangle corresponds to *serial order* of each segment or what the authors conclude to be “the neural code for serial order.” This study shows a cotemporal activation of the different segment representations before drawing of any segment shape was made, a result which supports both Lashley’s claims as well as the results presented in competitive queuing (CQ) models (Bullock, 2004; Cooper & Shallice, 2006; Page & Norris, 1998).

I would also contend that the Figure 5 data above is complimentary to a GNW perspective on determining serial order. This preconscious parallel loading of segment representation could be viewed as competition for access to the workspace nodes since Dehaene says the global workspace is associated with a “fleeting memory capacity” which integrates competing and cooperating input networks. In this way, workspace neurons seem to work similarly to the competitive choice layer found in CQ models. Also, higher PFC activation is shown when the monkey learns a sequence yet even though areas like cerebellum become more active with learned sequences, PFC activation remains even though the amount of activation diminishes.

This might be explained in Shanahan’s most recent model by saying that cortical columns in the PFC may be competing for access to its workspace nodes more frequently at the beginning of training, yet this diminishes as the cerebellum teaching signal is learned and fed back to PFC columns which now require less competition for representation across its workspace node neurons. One possibility is that the workspace may be encoding a serial order from the cerebellum for motor output once the sequence is overlearned, making PFC workspace neurons increasingly quiescent over numerous trials.

Referring back to Figure 4, perhaps the new C1 pattern of activity that reaches W1 is the method by which one item (or shape segment in this case) in working memory is presented. Then, once activated, a new pattern is introduced and the old pattern fades thanks to a wave of inhibition spreading around the workspace after the first pattern is executed. This may also help to explain where the cotemporal activations in area 46 prior to drawing shape segments are receiving their preloaded information from. Area 46 shows parallel loading of all segment representations, but where are these neural signals entering from? As mentioned above, the cerebellum is probably one piece of the puzzle but there are plenty of other areas involved in even a simple task such as the one performed by the monkeys in the Averbeck experiments. The strength of a GNW interpretation of serial order tasks is its reliance upon global activity which "is more likely to be achieved if there is '*resonance'* between bottom-up sensory information and top-down signals (Dehaene et al., 2003)."

Future possibilities

From the tasks simulated by Dehaene and Shanahan, it seems that GNW models could explain the cotemporal activation of area 46 in a similar way to CQ-based models such as N-STREAMS (Rhodes, et al., 2004). N-STREAMS is unique in its representation of both cortical and subcortical brain regions, a theory which nicely compliments a GNW model by relying on both parallel and serial representations. One issue with the Shanahan 2007 paper which requires further exploration is the fact that all "unconscious" elements of the model are still cortical. Many unconscious computations occur in subcortical regions such as the cerebellum, basal ganglia, amygdala, and hippocampus which rely on their subsequent thalamocortical circuits.

One obvious idea for expanding such a model would be to simulate a serial order task, such as that shown in Figure 5, including both cortical *and* subcortical regions that do not possess laminar structure. To do so would require a stronger reliance on physiological data for how subcortical areas such as basal ganglia or the amygdala interact with cortical regions such as the parietal cortex or PFC. Also, rather than using generalized workspace nodes, analogues to actual cortical regions could be made based on models such as N-STREAMS or via fMRI neural activity data from similar serial order tasks (Koechlin et al., 2004). Models such as N-STREAMS, PBWM (O'Reilly & Frank, 2006), or VITE-

FLETE (Contreras-Vidal et al., 1997) have already explored the interaction between cortical and subcortical brain areas, hence these modeling approaches could be adapted into a GNW model. Thus, investigating this interaction for cognitive tasks in a GNW model simulation would be an interesting next step.

Another area of exploration within a GNW model would be to develop a treatment for how numerous conscious sensory percepts, such as smelling *and* seeing a Christmas tree simultaneously, act as contexts for cognitive capabilities. This ties into how environmental stimuli and sensory percepts mix with emotions and volitional action based on internal decisions, i.e. Dehaene's "attentional circuits." In other words, how does a GNW model combine sensory percepts and higher-order cognitive states into one framework? These are extremely large questions which surely will not be answered for quite some time, and are exacerbated by the fact that a better GNW model should also account for regulatory mechanisms such as cholinergic and noradrenergic projections as well as take into consideration the effect of different neurotransmitters on certain cell types across different regions of the brain. Nonetheless, despite the countless avenues to be explored, a global workspace theory foundation seems to be a promising modeling approach for these topics.

As was mentioned earlier, Dehaene and colleagues as well as Shanahan have recently adopted the spiking model approach to global workspace models since this method gives more realistic representation of neural activity. Different neurons are known to elicit a number of spiking and bursting patterns dependent not only on the type of spike but the temporal nature of spiking in conjunction with conduction delays between neurons – a concept which Izhikevich calls "polychronization" (Izhikevich, 2006). The Izhikevich simplified Hodgkin-Huxley model can account for dynamics within neural networks such as plasticity and firing rhythms, a fact that has been shown in a recent paper exploring global workspace access and resulting γ-band oscillations during an attentional blink simulation (Dehaene et al., 2003).

The downside of such an approach is the amount of computational power necessary to simulate a full global neuronal workspace. Shanahan furthered the work of Dehaene's model by simulating not just access to a workspace node, but also broadcast from one node to other workspace nodes creating a sort of "mini" global workspace. Even in this network, however, only non-conscious cortical column activity for one workspace node could be efficiently computed. Thus, to create a biologically-plausible, fully-functioning simulation of the Averbeck serial order task would require a great deal of processing power in order to capture the dynamics of visual stimuli input, previous associations with that input, "where" and "what" visual stream processing, cerebellar and basal ganglia loops, prefrontal cortex competitive queuing, motor output to the spine, etc.

Lastly, there is the issue of embodiment. One problem with modeling of motor sequence tasks is that learning occurs in a simulated programming structure rather than in a physical brain with an external environment. To address this issue, several attempts have been made (Shanahan, 2006; Valpola & Joensuu master's thesis work at Helsinki University of Technology) to create virtual environments in a program called Webots (http://www.cyberbotics.com) which demonstrate motor adaptation and selective motor actions by programming brain analogue models into a 3-dimensional robot. Shanahan's Webots robot simulation did not take advantage of the Izhikevich-based Matlab code, so it would be interesting to see if such a robot could perform a simple serial order task via a spiking model implementation.

Conclusion

This last section presented a number of options for furthering a GNW as part of a larger research endeavor, however, for the purposes of simulating serial order tasks, such a schema can be reduced to focus on conscious and unconscious interaction among areas known to be vital for such tasks. Eventually it would be nice to include cortical areas throughout the brain in order to better explain a number of tasks such as ISR or 2x10 yet a GNW model which explains a host of working memory tasks is still a ways off. To conclude, I propose that a spiking GNW model implemented within a Webots robot simulation in order to accurately simulate serial order task behavior could be a promising avenue for future serial order research.

References

1. Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 99, 13172–13177.
2. Averbeck, B. B., Sohn, J.-W., and Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nature Neuroscience*, 9(2):276-282.
3. Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.
4. Baars, B. J. (1997). *In the theated of consciousness: The workspace of the mind*. Oxford: Oxford University Press.
5. Baars, B. J. and Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences*, 7(4):166-172.
6. Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences*, 8(9):426-433.
7. Cooper, R. and Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*, 113(4):887-916.
8. Dehaene, S., Kerszberg, M., and Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A, 95*(24):14529-14534.
9. Dehaene, S. and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1-2):1-37.
10. Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the United States of America,* 100(14), 8520–8525.
11. Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12(7-8):961-974.
12. Grossberg, S. (1980). How Does a Brain Build a Cognitive Code? *Psychological Review*, 87, 1-51.
13. Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, 49A, 80–115.
14. Kim, J.-J., Kwon, J. S., Park, H. J., Youn, T., Kang, D. H., Kim, M. S., Lee, D. S., and Lee, M. C. (2003). Functional disconnection between the prefrontal and parietal cortices during working memory processing in schizophrenia: A [15o]h2o pet study. *Am J Psychiatry*, 160(5):919-923.
15. Lashley, K. S. (1951). *The problem of serial order in behavior*. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley (Reprinted in The neuropsychology of Lashley, pp. 506–528, by Beach, F.A., Hebb, D.O., Morgan, C.T., & Nissen, H.W. (Eds.), 1960, New York: McGraw-Hill).
16. Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781.
17. Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., and Page, M. P. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, 23(5):699-746.
18. Rorie, A. E. and Newsome, W. T. (2005). A general mechanism for decision-making in the human

brain? *Trends Cogn Sci*, 9(2):41-43.

19. Sergent, C. & Dehaene, S. (2004). Is Consciousness a Gradual Phenomenon? Evidence for an All-or-None Bifurcation During the Attentional Blink. *American Psychological Society*, Nov;15(11):720-8.
20. Shanahan, M. P., & Baars, B. J. (2005). Applying global workspace theory to the frame problem. *Cognition*, 98(2), 157–176.
21. Shanahan, M. (2006). A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition*, 15(2):433-449.
22. Shanahan, M. A spiking neuron model of cortical broadcast and competition. *Consciousness and Cognition*, In Press, Corrected Proof.