A Standard Model of Mind:
AAAI Technical Report FS-17-05

An Integrated Computational Framework for Attention, Reinforcement Learning, and Working Memory

Andrea Stocco
Department of Psychology, Campus Box 351525
University of Washington
Seattle, WA 98195

Abstract

This paper proposes a reinterpretation of selective attention as a form of control of working memory based on self-generated reward signals and model-free reinforcement learning. In addition to being simple and parsimonious, this approach systematizes a number of classic psychological constructs without calling for additional, specific mechanisms. Finally, the papers presents the results of an empirical test of this framework, and elaborates on the implications of our findings for general models of control and intelligent behavior, as well as neurobiological models of the basal ganglia.

Selective attention is the capacity to focus and concentrate processing resources in the face of multiple competing stimuli. As such, it is an essential part of any adaptive, intelligent system.

In psychology, the ability to efficiently allocate processing resources is often studied and discussed under the terms of "executive function" or "cognitive control", and typically studied with paradigms that pitch a contextually inappropriate but prepotent response against a contextually appropriate but weaker one. For example, to successfully perform the Stroop task (in which participants are asked to name the color a word is printed in while ignoring the word itself), one must divert attention from the prepotent act of reading a word and instead focus on the internal goal of retrieving and naming the color name.

Unfortunately, it is difficult to distinguish executive function, selective attention, and working memory, as the three concepts are intimately connected. For instance, working memory correlates highly with measures of cognitive control, and even with tests of intelligence (Harrison, Shipstead, and Engle 2015).

Finally, according to an influential account, working memory is, in fact, just a form of "executive" attention—that is, the ability to concentrate and focus on some relevant information while ignoring the rest (Engle 2002). In other words, there is a high degree of overlap between these concepts, and, as a consequence, it is hard to describe computationally their specific mechanisms.

In this paper, I will introduce the idea that attention, can be understood in terms of reinforcement-based procedural control of working memory. Furthermore, I will introduce some experimental data that illustrates the point. Finally, I will show how this evidence has important consequences for the architecture of the standard model, and for how the standard model can be related to the neural architecture of the brain.

A Framework for Selective Attention

The proposed framework is outlined in Fig. 1; the red line marks the specific component that corresponds to selective attention. According to this framework, cognitive control is achieved through four interrelated mechanisms:

1. *Spreading Activation from Working Memory*. This assumption states that the availability of different units of declarative knowledge changes over time based on the contents of working memory. A common implementation of this idea consists of having activation spread from working memory contents to the associated contents of long-term memory.
2. *Procedural Control of Working Memory*. This assumption states that contents of working memory are updated or deleted by the activation and selection by specific units of procedural knowledge.
3. *Reinforcement Learning Control of Procedural Memory*. This assumption states that the units of procedural memory are selected on the bases of their expected reward, using a form of model-free reinforcement learning (RL) to progressively refine these expectations. A corollary of this assumption is that an agent can learn which is the most promising information to put in working memory through feedback.
4. *Continuous Performance Monitoring*. This assumption states that the agent regularly generates internal feedback signals about its own performance, with or without explicit feedback from the environment.

All of these assumptions can be defended on the bases of current neuroscientific research (Miller and Cohen 2001; McNab and Klingberg 2008; Schultz, Dayan, and Montague 1997). In this framework, working memory is thought of as a limited-capacity store. However, its strategic allocation is what ultimately determines the behavioral outcome, and is controlled by procedural memory.

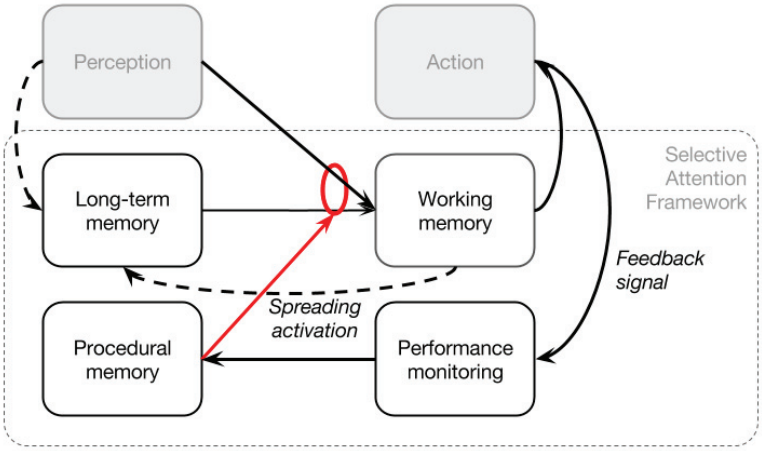

Figure 1: The proposed framework for selective attention

Implementation

The framework described herein was initially implemented in the ACT-R cognitive architecture (Anderson 2007), although, as it will be clear, it is easily generalizable to other architectures. For reference, a brief overview of ACT-R is given in Figure 2. ACT-R stores knowledge in two formats, declarative and procedural. *Declarative knowledge* is stored into table-like structures called *chunks*, which are used to represent static information, such as semantic memories (Paris is the capital of France), perceptual inputs (A black triangle is on the screen), or motor commands (Press the spacebar). In Fig. 2, chunks are represented by the flowchart symbol of a document.

Procedural knowledge, on the other hand, consists of *productions*, that is, state-action rules that implement the basic procedures needed to perform different tasks. In Fig. 2, productions are represented with the flowchart symbol of a process, with incoming arrows representing the state in which the rule can be applied and outgoing arrows representing their consequent actions.

The relationship between chunks and productions is mediated by a set of functional modules (solid rounded rectangles in Fig. 2). For instance, perceptual modules create new chunks to represent the contents of the outside world, and a memory module maintains chunks in long-term memory. Chunks are made available to production rules via a set of module-specific buffers, (dashed rounded rectangles in Fig. 2). Only when chunks are placed into buffers, can they be inspected, modified, and copied by production rules.

Simplifying Assumptions

ACT-R is a complex architecture: it has several dozen parameters, and the same task can be potentially modeled in very different ways. Thus, to implement a general framework on top of it, two simplifying assumptions were made:

1. *Single-Buffer Working Memory* Although ACT-R is made of several buffers and components, only one of them (the imaginal buffer) was used as a working memory store. Note that, because ACT-R's buffers can only contain a single chunk, the contents of working memory are by nature unstructured.
2. *Unconstrained Retrieval*. That retrieval of information from Long-Term Memory is completely characterized by chunk activation, which includes both base-level activation and spreading activation. Any time the retrieval buffer is empty, it simply retrieves the most active chunk.

Notice how these two assumptions significantly reduce ACT-R's number of degrees of freedom. The second assumption, in particular, is in direct contrast with the natural mechanisms of ACT-R, in which productions can place constraints on which chunk will be retrieved by specifying the desired pattern. For example, a production can directly retrieve the arithmetic fact "3 + 4 = 7" from long-term memory by specifically requesting to find a chunk whose first addend is "3" and whose second added is "4". In other words, an ACT-R model can always be programmed to retrieve the desired information efficiently.

In contrast, the retrieval of correct information in the proposed framework must rely on the strategic allocation of spreading activation, which, in turn, depends on which information is placed in working memory. For example, when trying to retrieve the arithmetic fact "3 + 4 = 7", a model developed in this framework would need to keep in working memory both "3" and "4", and let activation spread so that the baseline activation of the relevant arithmetic fact surpasses that of other, competing rivals.

There are three consequences of this assumption. The first is that, in this framework, attention comes at a price. In particular, the cost associated with deploying attention is the allocation of increasingly larger resources of working memory. Thus attention and working memory are by nature connected. The second consequence is that working memory is not only needed to store intermediate processing results but also to overcome interference. If a model can specify a unique retrieval pattern, as in canonical ACT-R, then there is no way it can make a mistake. In contrast, in the proposed

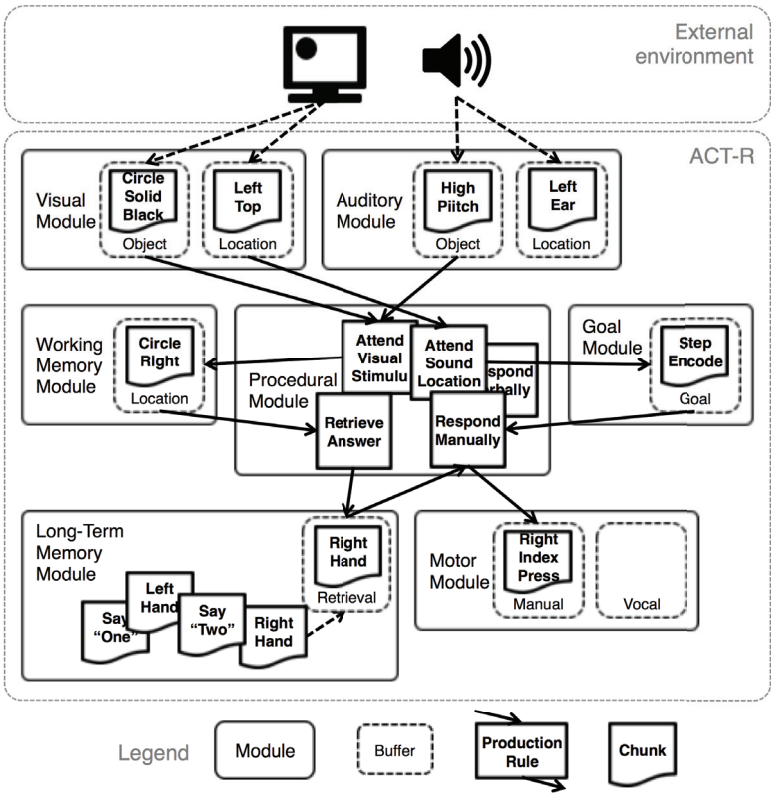

Figure 2: Overview of the ACT-R cognitive architecture

framework, errors occur naturally whenever spreading activation from working memory is not sufficient to overcome the activation of irrelevant chunks. Thus, the presence of specific retrieval cues in working memory is necessary to properly guide retrieval. The third and last consequence is that a model can learn to properly allocate attention through reinforcement learning. This is the key for any model conforming to this framework to improve its own performance over time.

Of these three consequences, the first two are the least contentious. In fact, many authors agree upon either the connection between attention and working memory (Engle 2002) or the connection between working memory and interference management (Miller and Cohen 2001). The reinforcement-learning nature of attention allocation, on the contrary, is a novel prediction of this framework and requires further examination. The next sections will introduce the specific form of reinforcement learning implemented in this framework, and summarize the results of a study that successfully tested this prediction.

Reinforcement Learning and Procedural Knowledge

In ACT-R, production rules are selected on the basis of their *utility*, an associated scalar quantity that can be thought of as the equivalent of the Q-value in reinforcement learning. Like a Q-value, the utility U_p of a production p is updated at every time based on the difference between actual and expected rewards. Specifically, the update follows the delta rule:

$$U_t^p = U_{t-1}^p + \alpha(R_t - U_{t-1}^p) \quad (1)$$

where α is the learning rate and R_t is the current reward.

The use of RL theory to implement procedural learning is common to both ACT-R and Soar, the two principal cognitive architectures in existence. It is also consistent with the current neuroscientific consensus on the nature of procedural knowledge. According to this consenses, procedural knowledge is related to the basal ganglia, a set of highly interconnected subcortical nuclei which receive extensive dopamine projections. Most importantly, the activity of dopamine neurons has been shown to closely mimic the reward prediction error (the difference $R_t - U_{t-1}^p$ in Eq. 1) in single-cell recordings in primates (Schultz, Dayan, and Montague 1997; Schultz 2000).

While reinforcement-learning models have often been used as a computational approximation of the basal ganglia (Joel, Niv, and Ruppin 2002), some noteworthy differences remain. Most importantly, the biological basal ganglia contain at least two pathways, whose different number of inhibitory synapses results in distinct mechanisms to excite and inhibit the release of thalamic projections to the prefrontal cortex, thus facilitating or preventing information from entering working memory. These two pathways can be

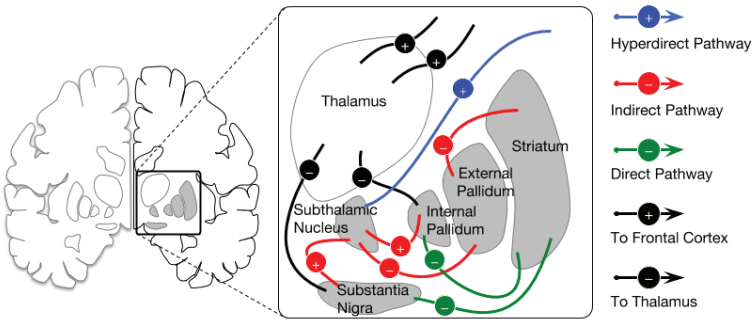

Figure 3: Functional anatomy of the basal ganglia

observed by looking at the physiology of the basal ganglia, and are highlighted in green (direct pathway) and red (indirect pathway) in Fig. 4.

The existence of these two pathways has important clinical effects. In a recent model, I was able to show that this distinction can be captured within ACT-R by modifying its procedural module to contain two set of productions (corresponding to the two pathways), controlled by partially independent parameters (Stocco 2017). Thus modified, ACT-R models were capable of reproducing important results from the literature and yield qualitative patterns that could not be previously obtained from the standard architecture.

The existence of two separate pathways fits nicely with an intuitive understanding of what attention is, and the fact that it might require the competing processes of focusing on a stimulus while actively ignoring and inhibiting the others. In fact, much evidence links attention, inhibition, and basal ganglia (Casey et al. 1997).

Testing the Framework

The previous sections have described a general framework for selective attention that integrates several psychological constructs. The framework's central tenet is that attention allocation depends on procedural learning mechanisms—which are, in turn, controlled by the specific RL dynamics of the basal ganglia. One straightforward way to test this assumption is to demonstrate the existence of a correlation between some measurable aspect of reinforcement learning and some performance aspect of selective attention. In turn, this requires measuring the performance of a group of participants ($N = 50$) across two behavioral tasks, one that measures RL parameters and one that measures selective attention.

The PSS Task

The relative strength of the two pathways of the basal ganglia can be behaviorally measured through the Probabilistic Stimulus Selection (PSS) task (Frank, Seeberger, and O'Reilly 2004). The PSS task is an iterative, two-alternative forced-choice paradigm in which participants repeatedly choose from pairs of non-verbalizable stimuli, each of which has a different probability of yielding a binary reward. Participants are first trained to select the most rewarding stimulus out of three different pairs. Note that the correct response can be made by either learning to *choose* the most rewarding stimulus, or by learning to *avoid* the least rewarding one. To distinguish between these two strategies, participants are then tested over the remaining combinations of stimuli, so that their sensitivity for learning to choose high-reward probability stimuli ("Choose" accuracy) and their sensitivity for learning to avoid low-reward probability stimuli ("Avoid" accuracy) can be measured independently. Multiple patient, pharmacological, and genetic studies (Frank, Seeberger, and O'Reilly 2004; Frank et al. 2007a) have shown that Choose accuracy reflects the contribution of the direct pathway, while Avoid accuracy reflects that of the indirect pathway.

The Simon Task

In the Simon task (Simon 1990), participants are asked to respond with their left and right hand to the specific visual feature (e.g., shape) of a stimulus that appears on a screen. For example, they might be asked to respond with their left hand when the stimulus is a square, and with their right hand if the stimulus is a circle. Interference occurs when a stimulus is presented on the side of the screen that is contralateral to the desired response. As a result, these "incongruent" trials are less accurate and take longer than "congruent" trials in which the stimulus appears on the side of the desired response. This extra time reflects the additional cost necessary to resolve conflict generated by the activation of two competing responses (one for the shape, one for the position). A review of the literature (Lu and Proctor 1995) concluded that interference in the Simon task occurs early on, at the moment in which the relevant and irrelevant features of the stimuli are being processed. In addition, because the stimuli of the Simon task do not convey any information about the correct response, the response rule is likely retrieved from long-term memory (Diamond 2013). Thus, it is safe to assume that performance in the Simon task reflects the allocation of attention to different stimulus features, and that the behavioral response reflects the retrieval of the instructed response rule from long-term memory. Taken together, these characteristics make the Simon task an ideal paradigm to test our framework.

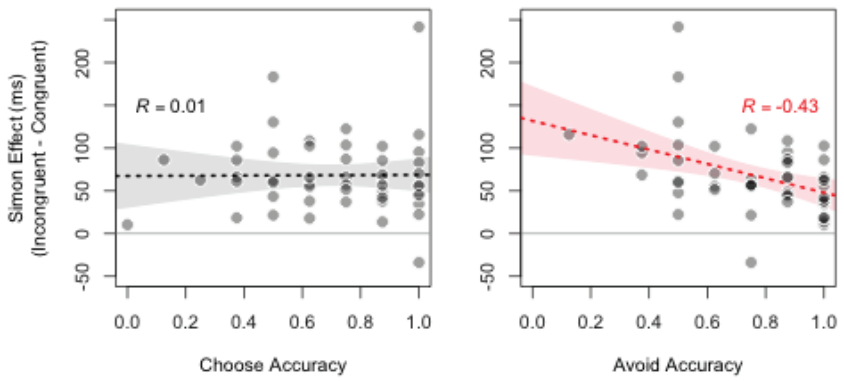

Figure 4: A successful test of the framework: Performance on a selective attention task (the Simon task) is correlated with the activity of the inhibitory pathway of the basal ganglia (as indexed by the Avoid accuracy in the PSS task)

Results

Such an experiment was recently carried out in my lab (Stocco et al. 2017), and its results confirmed our prediction: a negative correlation was found between the capacity to inhibit the location of the stimulus (as indexed by the Simon effect, or the response time difference between incongruent and congruent trials) and the strength of the basal ganglia's indirect pathway as indexed by the Avoid accuracy $[r(50) = -0.45, p = 0.001]$. Even more remarkably, this correlation was found to be specific to the Avoid accuracy; no correlation was found between the size of the Simon effect and Choose accuracy $[r(50) < 0.10.\ p > 0.40]$.

Thus, these experimental results confirm two facts simultaneously. The first is that a reliable relationship exists between the capacity to strategically allocate attention and procedural reinforcement learning. This confirms the broad framework of Fig. 1. The second is that the particular implementation of procedural memory that was suggested by Stocco (2017) captures the essential characteristics of the basal ganglia (the difference between the two pathways) that are necessary for implementing this framework in a biologically–plausible manner.

Discussion

This paper has outlined a minimal and general framework to interpret selective attention as the interaction between a limited-capacity working memory store and a procedural learning system driven by RL, and has successfully tested one of this framework's crucial predictions.

It is worth noting that the proposed framework has implications for the functional interpretation of basal ganglia physiology, and, by extension, for how the role of such circuit should be modeled within a unified cognitive architecture. Broadly speaking, models of the basal ganglia that include detailed physiology divide into two families. One family, which can be traced back to Albin, Young, and Penney's original work (Albin, Young, and Penney 1989), assumes that the most important characteristic of the circuit is the competition between the direct and indirect pathways. Because of the competing activity of the two pathways, models belonging to this family are also known as "brake and accelerator" models. Models belonging to this family include those by O'Reilly and Frank (2006) and Stocco, Lebiere, and Anderson (2010).

According to the other family, however, the main distinction is between the basal ganglia pathway that passes through the subthalamic nucleus (blue line in Fig. 4) and the two pathways that pass through the striatum. This interpretation was initially proposed by Kevin Gurney (Gurney, Prescott, and Redgrave 2001), and has been implemented in the spiking neuron models by Chris Eliasmith's group (Eliasmith et al. 2012; Stewart, Bekolay, and Eliasmith 2012).

The distinction between these two families might appear exquisitely academic, and, indeed, models exist that integrate both views (Frank et al. 2007b; Nambu 2004). However, the difference between the two types of models reflects a deeper disagreement on the role of the basal ganglia. Gurney's (2001) landmark model was explicitly proposed to frame the basal ganglia as an action selection mechanism. In contrast, while brake–accelerator models *might* be used to perform action selection, they tend to have a more general function. For instance, in the Prefrontal–Basal Ganglia Working Memory model (O'Reilly and Frank 2006) all the basal ganglia actions (which correspond to the opening and closing of cortical "gates") can fire simultaneously, while the Conditional Routing model (Stocco, Lebiere, and Anderson 2010) is only limited by a bottleneck of how much information can be transferred per cycle of operation.

Since our results favor the first family of models, it worth pondering whether procedural should also include an action selection bottleneck. Because the proposed framework was implemented in ACT-R, it did inherit the architecture's serial nature and single-production bottleneck. However, as noted in previous work (Stocco 2017), this is not a strict requirement of the proposed procedural module.

A second interesting consequence is related to the relationship between basal ganglia and cortex. For instance, the Conditional Routing model (Stocco, Lebiere, and Anderson 2010) assumes that all cortical areas are constantly exchang-

ing information, so that their contents are constantly subject to the risk of being accidentally overwritten. According to this model, the basal ganglia are mostly needed to block and inhibit most signals and prioritize weaker ones that would be quickly overwritten by more established cortico-cortical connections. This is not the case, however, in PBWM, where the connections between prefrontal and posterior regions are organized in a much tighter way, and the basal ganglia are mostly needed to let information in at the appropriate times. While both models could be, in principle, correct, our results identified a specific correlation between attention and the (inhibitory) indirect pathway only, thus suggesting that inhibition of incoming signals is more important that their excitation. In turn, this seems to favor the Conditional Routing model over PBWM. It also suggests that, in a more realistic cognitive architecture, buffers should be directly connected with each other, with large opportunities for working memory contents to be accidentally overwritten—and correspondingly greater needs for inhibitory processes in attention control.

Acknowledgments

This research was supported by a grant from the Office of Naval Research (ONRBAA13-003) entitled Training the Mind and Brain: Investigating Individual Differences in the Ability to Learn and Benefit Cognitively from Language Training" and by a start-up grant from the University of Washington.

References

Albin, R. L.; Young, A. B.; and Penney, J. B. 1989. The functional anatomy of basal ganglia disorders. *Trends in neurosciences* 12(10):366–375.

Anderson, J. R. 2007. *How can the human mind occur in the physical universe?* Oxford University Press.

Casey, B.; Castellanos, F. X.; Giedd, J. N.; Marsh, W. L.; Hamburger, S. D.; Schubert, A. B.; Vauss, Y. C.; Vaituzis, A. C.; Dickstein, D. P.; Sarfatti, S. E.; et al. 1997. Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* 36(3):374–383.

Diamond, A. 2013. Executive functions. *Annual review of psychology* 64:135–168.

Eliasmith, C.; Stewart, T. C.; Choo, X.; Bekolay, T.; DeWolf, T.; Tang, Y.; and Rasmussen, D. 2012. A large-scale model of the functioning brain. *science* 338(6111):1202–1205.

Engle, R. W. 2002. Working memory capacity as executive attention. *Current directions in psychological science* 11(1):19–23.

Frank, M. J.; Moustafa, A. A.; Haughey, H. M.; Curran, T.; and Hutchison, K. E. 2007a. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences* 104(41):16311–16316.

Frank, M. J.; Samanta, J.; Moustafa, A. A.; and Sherman, S. J. 2007b. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science* 318(5854):1309–1312.

Frank, M. J.; Seeberger, L. C.; and O'Reilly, R. C. 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306(5703):1940–1943.

Gurney, K.; Prescott, T. J.; and Redgrave, P. 2001. A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics* 84(6):401–410.

Harrison, T. L.; Shipstead, Z.; and Engle, R. W. 2015. Why is working memory capacity related to matrix reasoning tasks? *Memory & cognition* 43(3):389–396.

Joel, D.; Niv, Y.; and Ruppin, E. 2002. Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks* 15(4):535–547.

Lu, C.-H., and Proctor, R. W. 1995. The influence of irrelevant location information on performance: A review of the simon and spatial stroop effects. *Psychonomic bulletin & review* 2(2):174–207.

McNab, F., and Klingberg, T. 2008. Prefrontal cortex and basal ganglia control access to working memory. *Nature neuroscience* 11(1):103.

Miller, E. K., and Cohen, J. D. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 24(1):167–202.

Nambu, A. 2004. A new dynamic model of the cortico-basal ganglia loop. *Progress in brain research* 143:461–466.

O'Reilly, R. C., and Frank, M. J. 2006. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation* 18:283–328.

Schultz, W.; Dayan, P.; and Montague, P. R. 1997. A neural substrate of prediction and reward. *Science* 275:1593–1599.

Schultz, W. 2000. Multiple reward signals in the brain. *Nature Reviews Neuroscience* 1:199–207.

Simon, J. R. 1990. The effects of an irrelevant directional cue on human information processing. *Advances in psychology* 65:31–86.

Stewart, T. C.; Bekolay, T.; and Eliasmith, C. 2012. Learning to select actions with spiking neurons in the basal ganglia. *Frontiers in neuroscience* 6.

Stocco, A.; Murray, N. L.; Yamasaki, B. L.; Renno, T. J.; Nguyen, J.; and Prat, C. S. 2017. Individual differences in the simon effect are underpinned by differences in competitive dynamics of the basal ganglia: An experimental verification and a computational model. *Cognition* 164:31–45.

Stocco, A.; Lebiere, C.; and Anderson, J. R. 2010. Conditional routing of information to the cortex: A model of the basal ganglias role in cognitive coordination. *Psychological review* 117(2):541.

Stocco, A. 2017. A biologically plausible action selection system for cognitive architectures: Implications of basal ganglia anatomy for learning and decision-making models. *Cognitive Science*.