A Neuro-Mimetic Realization of the Common Model of Cognition via Hebbian Learning and Free Energy Minimization

Alexander G. Ororbia[1*], **Mary Alexandria Kelly**[2*]

[1] Rochester Institute of Technology, Rochester, NY, USA
[2] Carleton University, Ottawa, ON, Canada
ago@cs.rit.edu, mary.kelly4@carleton.ca

Abstract

Over the last few years, large neural generative models, capable of synthesizing semantically rich passages of text or producing complex images, have recently emerged as a popular representation of what has come to be known as "generative artificial intelligence" (generative AI). Beyond opening the door to new opportunities as well as challenges for the domain of statistical machine learning, the rising popularity of generative AI brings with it interesting questions for Cognitive Science, which seeks to discover the nature of the processes that underpin minds and brains as well as to understand how such functionality might be acquired and instantiated in biological (or artificial) substrate. With this goal in mind, we argue that a promising research program lies in the crafting of cognitive architectures, a long-standing tradition of the field, cast fundamentally in terms of neuro-mimetic generative building blocks. Concretely, we discuss the COGnitive Neural GENerative system, such an architecture that casts the Common Model of Cognition in terms of Hebbian adaptation operating in service of optimizing a variational free energy.

Introduction

Thanks to deep learning, the state-of-the-art in artificial intelligence (AI) has been advancing rapidly. Many problems deemed impossible a decade ago can now be performed by AI at an expert-level, such as complex board games and video games including Go (Silver et al. 2016) and StarCraft (Vinyals et al. 2019). For creative activities like art, writing, and conversation, it has become difficult to distinguish AI from human performance, leading to generative AI winning art contests (Roose 2022) and, conversely, human artists being mistaken for AI (Stokel-Walker 2023). Between tech CEOs claiming significant strides towards artificial general intelligence (AGI) and Geoffrey Hinton, "godfather of AI," warning of a potential "risk to humanity" (Metz 2023), it can be easy to believe that human-level AI is imminent.

It is the role of the cognitive scientist to ask critical questions of the recent developments in artificial intelligence:

1. Do humans and current AI agents behave similarly?
2. In what ways do humans and current AI agents differ?
3. Where and why do current AI agents and humans differ?

4. How can the differences be bridged?

Our focus is the fourth question. Where there is a gap between current AI and human behaviour, can we bridge such a gap by exploring novel architectures? We are working towards one possible answer to this question in form of a new cognitive architecture, CogNGen (the COGnitive Neural GENerative system; Ororbia and Kelly 2021, 2022, 2023). CogNGen is built on two neurobiologically and cognitively plausible models, namely a variant of predictive processing (Clark 2015; Salvatori et al. 2023) known as neural generative coding (NGC; Ororbia et al. 2020) and vector-symbolic (a.k.a. hyperdimensional computing; Gayler 2003; Kanerva 2009) models of human memory (Hintzman 1986; Kelly, Mewhort, and West 2017). Desirably, the use of these particular building blocks yields scalable, local update rules, based on variants of Hebbian learning (Hebb 1949). Such a form of plasticity allows the system to adapt its parameters while facilitating robustness in acquiring, storing, and composing distributed representations of encountered tasks.

In what follows, we motivate CogNGen by situating it within the Common Model of Cognition and recent generative neural network research. We provide an overview of the CogNGen architecture, discuss preliminary results on maze-learning tasks, and detail future directions for CogNGen's development. By explaining our approach at a high level, we aim to both motivate our own research and provide inspiration for other research programs that seek to progressively work towards bridging the gap between humans and AI.

The Common Model of Cognition

Cognitive architectures serve as unified theories of cognition and as computational frameworks for implementing models of specific tasks. Countless architectures have emerged over the past forty years, often with many similarities to one another (Kotseruba and Tsotsos 2018). On the basis of the strong commonalities between existing cognitive architectures, Laird, Lebiere, and Rosenbloom (2017) propose the *Common Model of Cognition* (CMC), a high-level theory of the modules of the mind and their interactions (see Fig. 1).

The Common Model of Cognition describes several key modules: perceptual and motor modules for interacting with the agent's environment, short-term/working memory buffers for holding the active data in the agent's mind, a
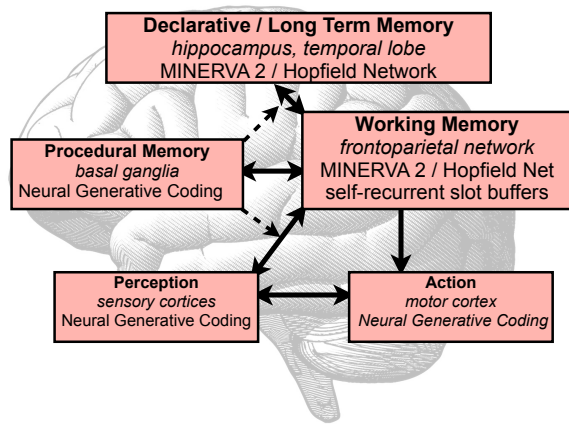
Figure 1: Common Model of Cognition (Laird, Lebiere, and Rosenbloom 2017), associated brain areas (Stocco et al. 2021; Steine-Hanson, Koh, and Stocco 2018; Stocco et al. 2018) and our approach to modelling each module. Solid arrows pass data while dashed arrows modulate data passing.

declarative (long-term) memory that stores world knowledge, and a procedural memory that controls information flow and evaluates possible actions (Laird, Lebiere, and Rosenbloom 2017). An evaluation of fMRI data taken from 200 participants across a diverse of tasks found correlations in patterns of activity across brain areas consistent with the Common Model of Cognition (Stocco et al. 2021).

Although research in cognitive psychology and cognitive neuroscience supports the Common Model of Cognition (CMC, Common Model) as a framework for understanding cognition, is the Common Model the correct framework for developing a novel cognitive architecture that is based on modern machine learning techniques? Our answer is *yes* and, as such, we have modelled CogNGen after the CMC, as illustrated in Figure 1. However, to further motivate our choice of adopting the CMC as a framework, we will next discuss the current state of generative networks and situate the CMC within current machine learning approaches.

Attention is All You Need

In Vaswani et al. (2017), *Attention is All You Need*, a new type of neural model was proposed, the transformer, which demonstrated superior performance on natural language processing (NLP) tasks when compared to recurrent networks, which had been the dominant paradigm for connectionist approaches to NLP since the Elman (1990) network. The crucial difference between the transformer and earlier models was the inclusion of a large number of "attention layers" to facilitate the selective weighting of different words or subword language units based on contextual relevance.

The claim that attention is *all* that is needed for NLP is a bold one. Yet, chatbots powered by transformers (e.g., OpenAI's ChatGPT, Microsoft's BingChat, Google's Bard) are fluent and versatile conversational partners. OpenAI's GPT-4 is able to write essays and solve math problems at the level of a C+ first-year student (Rudolph, Tan, and Tan 2023).

Notably, attention has not just been useful for NLP, as the transformer has also achieved state-of-the-art performance in visual (Bi, Zhu, and Meng 2021) and audio processing tasks (Lin et al. 2022; Verma and Berger 2021). The ability of transformers to communicate as if they "understand" has even raised questions as to whether transformers could have consciousness (Chalmers 2023). Juliani, Kanai, and Sasai (2022) evaluated Perceiver, a variant transformer model, on a variety of working memory and selective attention tasks, and found that the Perceiver implements Global Workspace Theory (Baars 1993; Mashour et al. 2020), a leading theory of access consciousness.

It should be noted, however, that the Perceiver architecture adds more than just "attention." Perceiver consists of input and ouptut modules, a shared central workspace for data, and attentional mechanisms that select data from the modules, modify the contents of the workspace, and pass data from the workspace back to the modules. Thus, the Perceiver has almost all of components of the Common Model of Cognition (see Figure 1): perception (input) and motor (output) modules, a working memory (a shared workspace), and the ability to modulate the flow of data in and out of the workspace (procedural memory). Perceiver lacks only the CMC's declarative memory. Juliani, Kanai, and Sasai (2022) found that the removal of any single component causes the Perceiver to fail on some of the working memory and selective attention tasks and thus fails to satisfy the criteria for implementing Global Workspace theory. As a result, all CMC components (except for declarative memory) are *necessary* to implement the functionalities of human consciousness, as postulated by Global Workspace Theory.

While Juliani et al. (2022) found that state-of-the-art networks satisfy many of the criteria for consciousness across different theories, many other capacities associated with consciousness are absent, chiefly episodic memory, which is necessary for the sense of self and history required for self-awareness and long-term planning. Juliani et al. noted that generative networks could be augmented with a memory network to address the lack of a long-term memory. Yet, Juliani et al. also point out other capacities required to achieve human-like consciousness, including theory of mind, causal reasoning, and meta-cognition. Thus, while adding attention layers to modern neural networks has led to tremendous progress, attention is *not* all that is needed for generative networks to achieve human-level performance. Rather, all components of the CMC are necessary for generative networks to exhibit human-level performance, and further components not specified by the CMC may also be necessary.

What is Attention?

It is undeniable that the addition of attention layers has resulted in a significant step forward in the capabilities of deep learning. Even so, Vaswani et al. (2017) did not set out to evaluate the transformer as a model of human attention; rather, the term *attention layer* was a mere act of naming a useful computational mechanism.

Is Vaswani et al.'s "attention" what a cognitive scientist would call attention? Attention is nebulously defined in cognitive psychology. However, a mathematical analysis of the

attention layer yields a correspondence to a real mechanism in the brain. Hebb (1949) critically observed of neurons that:

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, ... A's efficiency, as one of the cells firing B, is increased

Or, as Hebbian learning is pithily summarized, "neurons that fire together wire together." Each occurrence of an input pattern strengthens the ability of the network to reproduce it later. In effect, Hebbian learning allows neural networks to encode memories in their synaptic connectivity.

An early use of Hebbian learning is the Hopfield (1982) network. Given a Hopfield network that has memorized m patterns (i.e., traces) and a pattern $\mathbf{m_p}$ that is the input to the network (i.e., probe), an *echo* $\mathbf{m_e}$ is retrieved as a sum of all memory traces, each trace weighted by its similarity to the probe as measured by the · dot product:

$$\mathbf{m_e} = \sum_{i=1}^{m} (\mathbf{m_p} \cdot \mathbf{m}_i)\mathbf{m}_i \tag{1}$$

where $\mathbf{m}_i$ is the i-th trace memorized by the network.

Unfortunately, the memory storage capacity of the Hopfield (1982) network is very limited. As more traces are stored, the more likely the network is to return an echo that is a confused muddle of different traces. The storage capacity can be improved by adding non-linearity – we can raise the similarities between the probe and traces to an exponent b, such that highly similar memory traces contribute disproportionately more to the echo than dissimilar memory traces:

$$\mathbf{m_e} = \sum_{i=1}^{m} (\mathbf{m_p} \cdot \mathbf{m}_i)^b \mathbf{m}_i. \tag{2}$$

With the addition of the exponent b we have the memory retrieval equation for the Hintzman (1986) MINERVA 2 model of human memory. Hintzman's model is a Hopfield network with a non-linear weighting of the memories to improve retrieval (Kelly, Mewhort, and West 2017). Over the 35 years of research since it was first proposed, MINERVA 2 has accounted for human memory and learning in experimental paradigms too numerous to exhaustively cite, including judgement of frequency and recognition (Hintzman 1984), category learning (Hintzman 1986), implicit learning (Jamieson and Mewhort 2011), associative learning in humans and animals (Jamieson, Crump, and Hannah 2012), heuristics and biases in decision-making (Dougherty, Gettys, and Ogden 1999), hypothesis-generation (Thomas et al. 2008), word learning (Jamieson et al. 2018), and sentence production (Johns et al. 2016; Kelly et al. 2020b).

Hopfield networks can achieve exponential storage capacity by using an exponential function of the similarity to weight memory traces (Demircigil et al. 2017). We can adopt the widely-used *softmax* function, or normalized exponential function, for the purpose of weighting memory traces at retrieval (Ramsauer et al. 2021):

$$\mathbf{m_e} = \sum_{i=1}^{m} \frac{(e^{\mathbf{m_p} \cdot \mathbf{m}_i})}{\sum_{j=1}^{m} (e^{\mathbf{m_p} \cdot \mathbf{m}_j})} \mathbf{m}_i. \tag{3}$$

Modern Hopfield networks use the *softmax* to weight memory traces for efficient retrieval, which is also *exactly* what an attention layer in a transformer uses to apply attentional weighting across a set of items (Ramsauer et al. 2021). As such, transformers can be best understood as networks with a mix of conventional, back-propagation (backprop) trained layers and Hopfield memory layers whose contents are rapidly updated using Hebbian learning. Desirably, upon presentation of input pattern $\mathbf{m_p}$, the energy functional $\mathcal{F}$ associated with a Hopfield network can be formulated as:

$$\mathcal{F}(\Theta) = \frac{1}{2} \sum_{i=1}^{N_f} (\mathbf{m_p})_i^2 - \log \left(\sum_{k=1}^{m} \exp \left(\sum_{i=1}^{N_f} \mathbf{M}_{ki}(\mathbf{m_p})_i \right) \right) \tag{4}$$

where N_f is the number of input features, $\Theta = \{\mathbf{M}\}$, and $\mathbf{M}$ is the memory matrix of currently memorized patterns.

While backprop is a powerful learning algorithm, as yet, there is no consensus on identifying a neurological equivalent to backprop in the brain (Crick 1989). The primary challenge to realizing back-propagation of the error signal in the brain is that it requires that non-local information be used to modify synaptic connectivity globally in a highly coordinated manner. A number of potential biological mechanisms have been proposed, such as error signalling via neuropeptides (Liu et al. 2022) or high-frequency bursts of spikes from pyramidal neurons (Payeur et al. 2021). Alternatively, one can commit to strictly local synaptic update rules that can approximate the optimization of backprop (e.g., Betti and Gori 2020), which is the approach we adopt (Ororbia et al. 2020). Furthermore, when a network is trained sequentially using backprop to learn a series of distinct tasks, *catastrophic interference* occurs, with each new task causing the network to forget all previous tasks (French 1999; Mannering and Jones 2021; McCloskey and Cohen 1989).

To avoid the neural and behavioural pitfalls of backprop, we only use Hebbian learning in CogNGen. For working memory and declarative memory, we use MINERVA 2, implemented as Equation 2 with $b = 100$ for the simulations discussed in this paper, though future versions of the model will use the more efficient *softmax* retrieval as in Equation 3. As a complement to MINERVA 2's Hebbian learning, we craft neural circuits that also minimize their own energy functionals derived from the free energy principle (Friston 2005, 2010), instantiated as a form of predictive coding (Rao and Ballard 1999; Ororbia et al. 2020; Salvatori et al. 2023).

Free Energy and Predictive Processing

One of the two neural building blocks of CogNGen (alongside Hopfield networks), and the one that also endows it with a fundamental generative quality, is predictive coding (PC; Clark 2015; Friston 2005), based on a generalization known as neural generative coding (NGC; Ororbia and Kifer 2022). NGC circuits can be expressed in terms of the quantity that they optimize at each step in time, an objective function known as the variational free energy (VFE; Friston 2010).
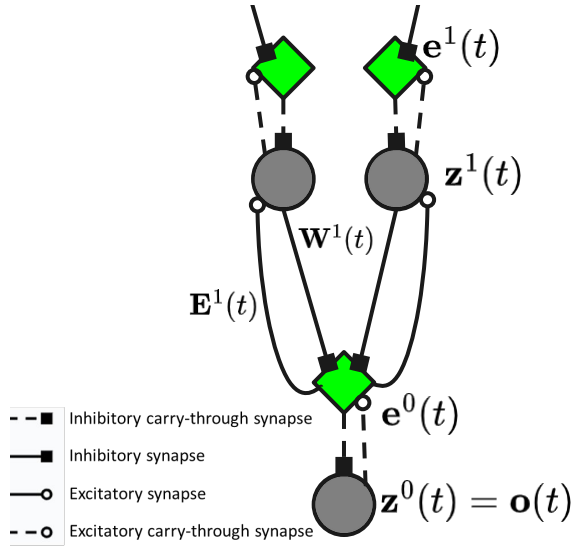
Figure 2: NGC circuit with state units $\{\mathbf{z}^0(t), \mathbf{z}^1(t)\}$ (dark grey circles), error neurons $\{\mathbf{e}^0(t), \mathbf{e}^1(t)\}$ (green diamonds), generative synapses $\mathbf{W}^1(t)$, and feedback synapses $\mathbf{E}^1(t)$.

Formally, the VFE functional can be written down as:

$$E(p, q, \mathbf{o}) = \underbrace{\sum_{\mathbf{z}} q(\mathbf{z})\log\left(\frac{q(\mathbf{z})}{p(\mathbf{z})}\right)}_{\text{Complexity Term}} + \underbrace{\sum_{\mathbf{z}} q(\mathbf{z})\log\left(\frac{1}{p(\mathbf{o} \mid \mathbf{z})}\right)}_{\text{Accuracy Term}}$$

(5)

where $p(\mathbf{o}|\mathbf{z})$ is the underlying directed generative model (or likelihood) that an NGC circuit embodies and $p(\mathbf{z})$ is the prior over its latent variables (or neural activities). $q(\mathbf{s})$ is an approximate posterior distribution over the latent variables $\mathbf{z}$ (given an observation), the choice of which often depends on a set of parameters that are to be optimized. Observe that VFE is a functional that balances two terms—one that encourages improving the likelihood mapping/generative component (or accuracy) while the other term penalizes model complexity (aligning with the idea of Occam's razor) and encourages the underlying model to ensure that the (Kullback-Leibler) divergence between its proxy posterior (or recognition model) and its prior is as small as possible.

Given that we have committed ourselves to viewing our cognitive architecture's foundational neural circuitry as a generative process in of itself, we next briefly specify the form of the generative model that each circuit embodies. A neural circuit's marginal probability has the dependencies:

$$p(\mathbf{z}^0, \ldots, \mathbf{z}^L) = p(\mathbf{z}^L) \prod_{\ell=0}^{L-1} p(\mathbf{z}^\ell \mid \mathbf{z}^{\ell+1})$$

(6)

and we further consider the distribution $p(\mathbf{z}^\ell \mid \mathbf{z}^{\ell+1})$ to be a multivariate Gaussian distribution with a mean given by a transformation f^ℓ of the latent variables of the level above. In our formulation, f^ℓ is a linear map consisting of activation function ϕ^ℓ and synaptic matrix $\mathbf{W}^\ell(t)$, i.e., $\bar{\mathbf{z}}^\ell = f^\ell(\mathbf{z}^{\ell+1}) = \mathbf{W}^\ell(t) \cdot \phi^\ell(\mathbf{z}^{\ell+1})$, where $\cdot$ indicates

matrix-vector multiplication. As a result, we arrive at:

$$p(\mathbf{z}^L) = \mathcal{N}(\bar{\mathbf{z}}^L, \boldsymbol{\Sigma}^L), \; p(\mathbf{z}^\ell \mid \mathbf{z}^{\ell+1}) = \mathcal{N}(f^\ell(\mathbf{z}^{\ell+1}), \Sigma^\ell).$$

(7)

To optimize the generative structure above, all that remains is to specify its energy functional. Due to the particular assumptions made above, coupled with a mean-field approximation that enforces independence among the neurons that compose the circuit (see Friston 2010; Ororbia et al. 2022; Ororbia and Kifer 2022; Ororbia et al. 2023; Salvatori et al. 2023, for details on derivations/assumptions), we obtain a form of the VFE of Equation 5 as follows:

$$\mathcal{F}(\Theta) = \sum_{\ell=0}^{L} \frac{1}{2\Sigma^\ell} \sum_{i=1}^{\mathcal{J}_\ell} \left(\mathbf{z}_i^\ell(t) - \bar{\mathbf{z}}_i^\ell\right)^2$$

(8)

where $\mathcal{J}_\ell$ denotes the number of neuronal cells in layer ℓ and $\Theta = \{\mathbf{W}^\ell(t), \mathbf{E}^\ell(t)\}_{\ell=1}^L$ (i.e., Θ is the construct that houses all of the time-evolving generative $\mathbf{W}^\ell(t)$ and message passing $\mathbf{E}^\ell(t)$ synaptic matrices). The above VFE functional highlights two neurobiological commitments made by NGC: 1) there exists a neuronal unit type that specializes in calculating mismatch signals, i.e., the precision-weighted *error neuron* $\mathbf{e}^\ell(t) = \frac{1}{\Sigma^\ell}\left(\mathbf{z}^\ell(t) - \bar{\mathbf{z}}^\ell\right)^1$, and 2) a neuron i in one layer (inside of vector $\mathbf{z}^{\ell+1}$) guesses the activities of another neuron j in layer ℓ. The dynamics of any neuronal layer, i.e., the *state neurons*, within our circuit follow:

$$\tau_m \frac{\partial \mathbf{z}^\ell(t)}{\partial t} = -\gamma \mathbf{z}^\ell(t) + \mathbf{d}^\ell \odot f_D(\mathbf{z}^\ell(t)) - \mathbf{e}^\ell(t)$$

(9)

where $\odot$ indicates the Hadamard product, $\mathbf{d}^\ell = \mathbf{E}^\ell(t) \cdot \mathbf{e}^{\ell-1}(t)$, i.e., the perturbations produced by error messages that are passed back along feedback synapses $\mathbf{E}^\ell(t)$, $f_D()$ is a neural activity dampening function, and τ_m is the cellular membrane time constant. The generative synapses within the circuit are adjusted according to the following:

$$\tau_w \frac{\partial \mathbf{W}^\ell(t)}{\partial t} = -\gamma_w \mathbf{W}^\ell(t) + \mathbf{e}^{\ell-1}(t) \cdot \left(\mathbf{z}^\ell(t)\right)^T$$

(10)

$$\tau_e \frac{\partial \mathbf{E}^\ell(t)}{\partial t} = -\gamma_e \mathbf{E}^\ell(t) + \mathbf{z}^\ell(t) \cdot \left(\mathbf{e}^{\ell-1}(t)\right)^T$$

(11)

where τ_w and τ_e are generative and error feedback synaptic plasticity time constants, respectively, and γ_w and γ_e are generative and error feedback synaptic plasticity decay constants, respectively. These differential equations indicate that synaptic plasticity is a function of a local update (i.e., a simple two-factor Hebbian rule based on error neuron activity and state unit activity values). See Figure 2 for an illustration of an NGC neural circuit.

While there are many ways to utilize NGC circuitry, upon presentation of input stimulus $\mathbf{o}(t)$ and possibly top-down context or input vector $\mathbf{c}(t)$ (i.e., meaning we clamp $\mathbf{z}^0(t) = \mathbf{o}(t)$ and $\mathbf{z}^L(t) = \mathbf{c}(t)$) our cognitive system iteratively applies Equation 9 for T steps and then applies Equations 10 and 11, implying that synapses evolve at a slower time-scale than neural activities.

---

[1]Note that in this case, we have assumed a fixed scalar precision $\frac{1}{\Sigma^\ell}$ but, in general, this can be constructed to be a learnable synaptic matrix (Ororbia and Kifer 2022; Salvatori et al. 2023).
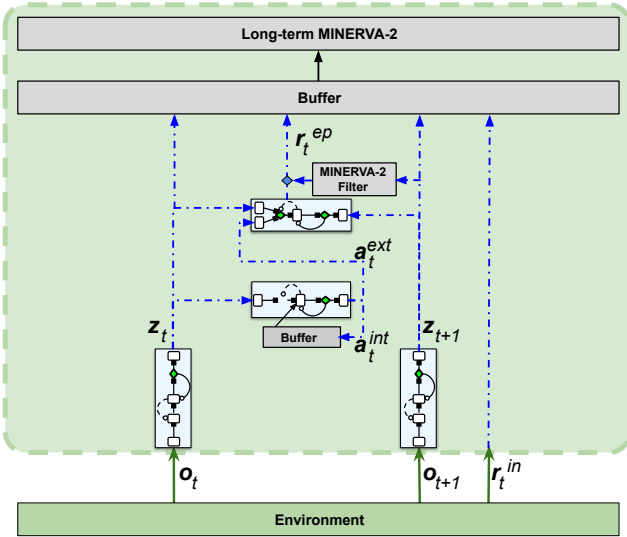
Figure 3: The CogNGen architecture, composed of vector-symbolic memory and free energy minimizing circuitry.

The CogNGen Architecture

CogNGen's design entails implementing, using either (assemblies of) vector-symbolic memory or free-energy minimizing NGC circuits, a perception system (for one or more modalities), a motor-action module, a procedural memory system, working memory buffers, and a declarative memory. In Figure 3, we depict a minimal implementation of the CogNGen architecture (i.e., referred to as a "kernel", a term we borrow from operating systems design). Note that the diagram depicts CogNGen in "processing mode" (i.e., no synaptic update is performed in this mode, only circuit latent states are updated, actions are taken, and internal reward signals are computed/filtered). However, relevant information is stored in a working memory buffer, which interfaces with the long-term MINERVA 2 memory. Transitions are sampled from MINERVA 2 when the "learning process" is triggered (i.e., after state $\mathbf{z}_{t+1}$ has been encountered by the perceptual module and stored, samples are replayed from memory to update the motor and procedural circuits).

Perception: In general, an NGC circuit can be constructed to serve the role of the encoder f_e for CogNGen. Doing so would yield the additional advantage that a top-down directed generative model, or decoder $f_g: \mathbf{z}_t \mapsto \mathbf{o}_t$, would be learned jointly along with $f_e: \mathbf{o}_t \mapsto \mathbf{z}_t$. This is motivated by prior work (Ororbia and Kifer 2022) showing that NGC learns a good density estimator of data (from which new samples can be "fantasized"; Ororbia et al. 2020; Ororbia and Kifer 2022). An NGC encoder would, by default, be unsupervised, especially if it is being pre-trained before a task simulation (i.e., the NGC f_e would be trained to predict $\mathbf{o}_t$ given $\mathbf{z}_t$, where $\mathbf{z}_t$ is iteratively crafted by the NGC settling process, and a feedforward model can be trained to amortize the settling process to further reduce computational complexity). Having a decoder allows for visual interpre-

tation of the distributed representations acquired by CogNGen, since a latent vector $\hat{\mathbf{z}}_t$ (such one as produced by the procedural dynamics model) could be run through the underlying top-down directed generative f_g to produce its corresponding instantiation $\hat{\mathbf{o}}_t$ in observation space.

Another advantage of the above formulation is that, if a task-specific encoder f_e for a modality is available (as in Ororbia and Kelly 2022), it may be used alongside or in place of a full NGC encoder circuit. This can simplify and speed up the simulations involving CogNGen, especially if learning a joint perceptual-memory-control system is not the goal, allowing the experimenter to leverage a reliable, stable state representation to design or experiment with various configurations of the CogNGen kernel's other internal subsystems and observe their impact on the task at hand.

Working Memory The *self-recurrent slot buffer* (see Figures 1 & 3) serves as the glue that joins the modules of CogNGen together. Working memory in the CMC can be implemented in a variety of ways (Laird, Lebiere, and Rosenbloom 2017). In ACT-R (Ritter, Tehranchi, and Oury 2019), for example, the mind/brain is understood as consisting of modules connected by buffers, each storing data in a small, finite number of slots. Collectively, the buffers serve as ACT-R's working memory and recurrent slot buffers in CogNGen serve the same purpose, and are inspired by Kruijne et al. (2021)'s memory model. The model stores a finite quantity M_w of projected latent state vectors into a set of self-recurrent memory vector slots. Each memory slot in the buffer is represented by $\mathbf{m}^i \in \mathcal{R}^{M_d \times 1}$ (M_d is the dimensionality of the memory slot). Concretely, the self-recurrent slot buffer operates according to the following:

$$\mathbf{k}_t^i = \mathbf{Q}^i \cdot \mathbf{z}_t, \forall i = 1, ..., M_w \tag{12}$$

$$s^i = \mathbf{s}^i = \frac{1}{|\mathbf{m}^i|} \left(\sum_j \lfloor \mathbf{m}^i - \mathbf{k}_t^i \rfloor_{j,1} + \lfloor \mathbf{k}_t^i - \mathbf{m}^i \rfloor_{j,1} \right) \tag{13}$$

$$\mathbf{m}_t = \left[[\mathbf{m}^1, \mathbf{s}^1], ..., [\mathbf{m}^i, \mathbf{s}^i], ..., [\mathbf{m}^{M_w}, \mathbf{s}^{M_w}] \right] \tag{14}$$

where $\mathbf{Q}^i \in \mathcal{R}^{M_d \times D_z}$ is the ith random projection matrix (sampled from a centered Gaussian distribution), i.e., there is one projection matrix per working memory slot. The match score for any slot i is $\mathbf{s}^i = \mathcal{R}^{1 \times 1}$ (a 1×1 vector) and thus a scalar s^i. The working memory buffers, in effect, calculate the match score between the ith key and ith slot/value, and then return the entire concatenated contents $\mathbf{m}_t$ of working memory (including match scores). In general, other NGC circuits manipulate these buffers, deciding whether or not to store information within them at any time.

On Motor-Control To allow CogNGen to manipulate its world, as well as facilitate internal control (such as modifying a slot buffer), CogNGen is endowed with circuits to drive its actuators. Building on notions of active inference (Botvinick and Toussaint 2012), we leverage variations of NGC circuits called "active NGC" circuits (Ororbia and Mali 2021). Specifically, we design a motor-action model $f_a: \mathbf{z}_t \mapsto (\mathbf{c}_t^{int}, \mathbf{c}_t^{ext})$ (which, in effect, offers some of the functionality provided by the motor cortex) that outputs two

control signals at each time step, i.e., internal control signal $\mathbf{c}_t^{int} \in \mathcal{R}^{A_{int} \times 1}$ and external control signal $\mathbf{c}_t^{ext} \in \mathcal{R}^{A_{ext} \times 1}$. A discrete internal action $a_t^{int} \in \{1, 2, ..., A_{int}\}$ is read as $a_t^{int} = \arg\max_i \mathbf{c}_t^{int}$ while an external action $a_t^{ext} \in \{1, 2, ..., A_{ext}\}$ is read as $a_t^{ext} = \arg\max_j \mathbf{c}_t^{ext}$ (A_{int} is the number of internal actions and A_{ext} is the number of external actions). Action a_t^{ext} impacts the environment that CogNGen is interacting with while action a_t^{int} manipulates coupled memory. At a high-level, any active NGC circuit is trained using a reward signal that is a function of an instrumental term (e.g., sparse problem reward, prior preference circuit), and an epistemic foraging term—a dopamine-like signal produced by procedural memory—to encourage actions that support intelligent exploration.

Procedural Memory   Neuro-behavioral studies find that reward signals are used (by the brain) to evaluate whether or not an action (motor activity) is desirable/undesirable (Rangel, Camerer, and Montague 2008). Action selection is driven by changes in the neural activity of the basal ganglia which estimate the value of the expected reward (Hikosaka, Nakamura, and Nakahara 2006). Motivated by the finding of expected value estimation in the brain, CogNGen's procedural module implements a neural circuit that produces intrinsic (epistemic) signals. At a high level, this machinery facilitates some of the functionality offered by the basal ganglia and procedural memory. Specifically, we implement an NGC dynamics model from which a reward signal is calculated as a normalized form of its VFE at time t. We couple the dynamics model with a short-term memory module, based on MINERVA 2, which adjusts the reward value produced by the dynamics circuit by determining if the currently observed state is familiar or unfamiliar.

Another important aspect of CogNGen's procedural memory includes sub-circuits that generate "gating variables" that dynamically turn on (trigger) or turn off (suppress) sub-networks, i.e., subsets of connected neuronal units within particular neural structures, effectively emulating the basal ganglia's role as a conditional information routing construct (Stocco, Lebiere, and Anderson 2010; Stocco 2018). These sub-circuit structures are learned through various forms of competitive Hebbian learning (which is particular design choice influenced by its ability to prevent catastrophic forgetting from happening within the basal ganglia itself; Ororbia et al. 2022). While this element of CogNGen's procedural memory has not yet been evaluated in the context of the whole system itself, prior effort has shown that complementary basal ganglia-perceptual systems work well in the face of online task-free continual learning (Ororbia 2021; Ororbia et al. 2022). We remark that, while CogNGen focuses on providing task-switching level executive functionality through the basal ganglia/procedural memory, other candidate brain structures that could provide this dynamic neural region suppression/gating include the prefrontal cortex (Rougier and O'Reilly 2002).

Long-Term Memory   For CogNGen, we adopt Collins, Milliken, and Jamieson (2020)'s approach and model both working and declarative memory using MINERVA 2. Our working MINERVA 2, like the slot buffer, is cleared after a task is completed (e.g., a maze is solved), whereas the contents of the declarative MINERVA 2 (which serves as the episodic memory in CogNGen) persist.

Much as is done in (deep) reinforcement learning (RL), in order to improve the stability and convergence of networks trained over many episodes (Mnih et al. 2015), CogNGEN leverages its MINERVA 2 episodic memory by sampling from it, inducing an approximate form of experience replay. This is motivated by early studies of rats where neural replay sequences were detected in the hippocampus (Skaggs and McNaughton 1996) during rest—it was found that "place" cells spontaneously/rapidly fired so as to represent the previous paths traversed by the rats when they had been awake. These "replay" sequences lasted only a fraction of a second but essentially covered several seconds of real-world experience. Similar replay effects have been been detected in human subjects (Kurth-Nelson et al. 2016), providing further neurobiological justification of the replay buffer used in modern-day RL systems. In CogNGen, information is transferred into this MINERVA 2 memory through an intermediate working buffer (which contains partial experience information encountered over time, such as state, action taken, and reward). At various points in time, when CogNGen is not adjusting synapses online, memories are replayed sequentially from the episodic memory and the confabulated patterns are used to induce an extra update to parameters.

Simulation Results

Empirically, CogNGen has already been demonstrated to perform well (Ororbia and Kelly 2022, 2023) on a set of tasks taken from the grid-world environment benchmark known as Mini-GridWorld (an OpenAI Gym extension; Chevalier-Boisvert et al. 2023). The maze solving tasks in the grid-world are shown in Figure 4 with results depicted in Table 1. The mazes provide a useful suite of controlled simulations of varying complexity that test the viability and effectiveness of our initial prototype of the CogNGen system. We have compared CogNGen to several baseline models: a standard deep Q-network (DQN; Mnih et al. 2015), a DQN that equipped with random network distillation (RnD; Burda et al. 2018), and a DQN that learns through the BeBold exploration framework (BeBold DQN-CNT; Zhang et al. 2020). See Ororbia and Kelly (2022, 2023) for simulation environment setup and additional experimental details.

On the Value of a Neuro-mimetic CMC

Although the current form of CogNGen is not without its limitations, there are already many advantages to the cognitive control architecture developed thus far. Notably, when taking in the totality of its vector-symbolic memory and predictive coding elements, CogGNen can be written down as optimizing, in a modular fashion, a single free energy function; concretely, this means that the entire system is minimizing a sum of local functionals, including the vector-symbolic memory components (in terms of their Hopfield energies) and the NGC circuits (in terms of their VFEs). Formally, this global VFE is: $\mathcal{F}(\Theta) = \sum_j \mathcal{F}_j(\Theta_j)$, where Θ_j contains the plastic synapses associated with structure
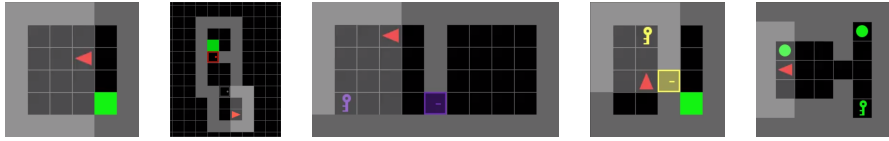
Figure 4: Visualization of Gym Mini-Grid tasks that CogNGen has been empirically demonstrated on. From left to right: random room escape (R6x6), procedural multi-room task (MR), door-unlock (Unl), door-key (DK), and memory task (Mem).

	Average Success Rate					Average Episode Length				
	R6x6	MR	Unl	DK	Mem	R6x6	MR	Unl	DK	Mem
DQN	99.50	0.00	0.00	0.00	40.0	9.31	100.0	100.0	100.0	41.14
RnD	**100.00**	90.00	**100.0**	**100.00**	48.5	3.50	31.46	4.08	3.71	2.78
BeBold DQN-CNT	**100.00**	98.00	**100.0**	**100.00**	48.0	3.98	23.51	4.46	3.93	2.92
CogNGen	**100.00**	**98.50**	**100.0**	**100.00**	**98.5**	3.90	23.41	4.15	5.48	2.96

Table 1: We present results for the average success rate (%) over the last 100 episodes and average episode length (% of maximum/worst-case episode length - closer to 0 is better/more efficient) over the last 100 episodes.
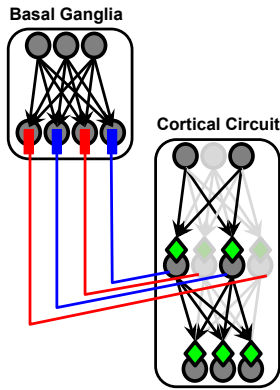

Figure 5: An element of CogNGen's procedural memory includes circuits that implement the basal ganglia for information routing. Depicted is such a sub-circuit producing gating signals to turn off/on neurons in an NGC perceptual circuit.

j (e.g., this could be a memory module within declarative memory or a NGC circuit portion of the motor cortex) and $\mathcal{F}_j$ is its associated (free) energy functional.

Desirably, at the heart of CogNGen is the story of on-line learning and adaptation, allowing one to characterize its global decisions (with respect to tasks) from not only the behavioral level but also the neural dynamics level. In addition, it is the particular form of CogNGen's neural dynamics, which follow the gradient flow of a variational free energy, that not only connect well to Bayesian brain theory (and the grander picture of predictive processing as not only a mechanistic description of neural activity but as a general principle underlying brain structure and function) but also permit flexible design choices such as the encouragement of strong sparsity in the neural activities. In addition, CogNGen is a small piece of evidence that a heterogeneous mixture of learning paradigms—Hebbian adaptation in the memory modules, local error-driven plasticity in the neural generative coding (NGC) circuits, and competitive Hebbian learning for the basal ganglia construct—can yield a coherent platform for crafting brain-motivated structures to instantiate the various core modules of the Common Model of Cognition. Beyond this, the neural dynamics of NGC generalize well to the level of spiking neuronal cells (Ballard, Rao, and Zhang 2000; Ororbia 2023), offering a possible transition between CogNGen's current story of adaptation through coarse-grained rate coded dynamics to finer-grained spiking electro-physiological dynamics.

Future Work

CogNGen is a functional implementation of the Common Model of Cognition and is able to learn a sample of reinforcement learning tasks at a rate comparable to or exceeding competing deep Q-learning networks (see Ororbia and Kelly 2022, 2023). However, CogNGen, as implemented, is only a prototype. Our goal is to develop CogNGen to be a general purpose framework able to learn a wider variety of tasks and to learn different tasks consecutively without catastrophic forgetting. Learning tasks with continuous-valued perceptual-motor environments may require CogNGen to incorporate more complex perceptual and motor modules, for which we would use additional NGC circuits.

The main obstacle in scaling CogNGen to larger tasks, and consecutive sequences of distinct tasks, is memory architecture. In our current work, we are refining CogNGen's memory. The first step is a switch from using the MINERVA 2 memory retrieval (Equation 2) to the more efficient softmax retrieval (Equation 3) of modern Hopfield networks and transformers). The second step is to allow CogNGen to learn to how to control and modulate the contents of the short-term and long-term MINERVA 2 models / Hopfield networks in the same manner that CogNGen learns to control the contents of the slot buffers. At present, the contents of the MINERVA 2 models are hardwired to be histories of recent observations. Allowing CogNGen to control memory

would allow it to pay more attention to some observations than others as well as learn representations for memories.

If in scaling CogNGen to larger tasks or more diverse tasks, we need more long-term memory storage than what is provided by modern Hopfield networks / MINERVA 2, we may explore the episodic / semantic memory distinction and add a separate semantic memory, a model that scales with the number of concepts or categories rather than the number of new episodes (e.g., Kanerva 1988; Kelly et al. 2020a).

Conclusion

CogNGen is a proof-of-concept that the Common Model of Cognition provides an appropriate blueprint for designing a modern, deep neural cognitive architecture capable of learning to complete arbitrary tasks. CogNGen commits to using free energy minimization and Hebbian learning rather than the standard back-propagation of errors (backprop), a choice motivated by: (1) the lack of neuro-scientific evidence for backprop; (2) the recent success of Hebbian-like neuro-circuitry in the attention heads of transformer models; (3) the long history of success of Hebbian learning in computational cognitive modelling; and (4) the aim of designing an architecture that can more easily avoid catastrophic forgetting, allowing it to, in principle, be a more general learner (as is desirable for modelling learning in human and animals). We are excited to push forward this architecture with further refinements and more simulations, and equally excited for other Common Model-inspired approaches to designing intelligent agents using deep learning.

Acknowledgments

References

Baars, B. J. 1993. *A cognitive theory of consciousness*. Cambridge University Press.

Ballard, D. H.; Rao, R. P. N.; and Zhang, Z. 2000. A single-spike model of predictive coding. *Neurocomputing*, 32: 17–23.

Betti, A.; and Gori, M. 2020. Backprop Diffusion is Biologically Plausible. arXiv:1912.04635.

Bi, J.; Zhu, Z.; and Meng, Q. 2021. Transformer in Computer Vision. In *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, 178–188. Fuzhou, China.

Botvinick, M.; and Toussaint, M. 2012. Planning as inference. *Trends in Cognitive Sciences*, 16(10): 485–488.

Burda, Y.; Edwards, H.; Storkey, A.; and Klimov, O. 2018. Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*.

Chalmers, D. J. 2023. Could a Large Language Model be Conscious? arXiv:2303.07103.

Chevalier-Boisvert, M.; Dai, B.; Towers, M.; de Lazcano, R.; Willems, L.; Lahlou, S.; Pal, S.; Castro, P. S.; and Terry, J. 2023. Minigrid & Miniworld: Modular & Customizable Reinforcement Learning Environments for Goal-Oriented Tasks. *CoRR*, abs/2306.13831.

Clark, A. 2015. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Collins, R. N.; Milliken, B.; and Jamieson, R. K. 2020. MINERVA-DE: An instance model of the deficient processing theory. *Journal of Memory and Language*, 115: 104151.

Crick, F. 1989. The recent excitement about neural networks. *Nature*, 337(6203): 129–132.

Demircigil, M.; Heusel, J.; Löwe, M.; Upgang, S.; and Vermet, F. 2017. On a Model of Associative Memory with Huge Storage Capacity. *Journal of Statistical Physics*, 168: 288–299.

Dougherty, M. R. P.; Gettys, C. F.; and Ogden, E. E. 1999. MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, 106: 180–209.

Elman, J. L. 1990. Finding structure in time. *Cognitive Science*, 14(2): 179–211.

French, R. M. 1999. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4): 128–135.

Friston, K. 2005. A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456): 815–836.

Friston, K. 2010. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2): 127–138.

Gayler, R. W. 2003. Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. In Slezak, P., ed., *Proceedings of the Joint International Conference on Cognitive Science*, 133–138. Sydney, Australia: University of New South Wales.

Hebb, D. O. 1949. *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Hikosaka, O.; Nakamura, K.; and Nakahara, H. 2006. Basal ganglia orient eyes to reward. *Journal of Neurophysiology*, 95(2): 567–584.

Hintzman, D. L. 1984. MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers*, 16: 96–101.

Hintzman, D. L. 1986. "Schema abstraction" in multiple-trace memory models. *Psychological Review*, 93: 411–428.

Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8): 2554–2558.

Jamieson, R. K.; Avery, J. E.; Johns, B. T.; and Jones, M. N. 2018. An Instance Theory of Semantic Memory. *Computational Brain & Behavior*, 1(2): 119–136.

Jamieson, R. K.; Crump, M. J. C.; and Hannah, S. D. 2012. An instance theory of associative learning. *Learning & Behavior*, 40: 61–82.

Jamieson, R. K.; and Mewhort, D. J. K. 2011. Grammaticality is inferred from global similarity: A reply to Kinder (2010). *The Quarterly Journal of Experimental Psychology*, 64: 209–216.

Johns, B. T.; Jamieson, R. K.; Crump, M. J. C.; Jones, M. N.; and Mewhort, D. J. K. 2016. The combinatorial power of experience. In Papafragou, A.; Grodner, D.; Mirman, D.; and Trueswell, J. C., eds., *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 1325–1330. Austin, TX: Cognitive Science Society.

Juliani, A.; Arulkumaran, K.; Sasai, S.; and Kanai, R. 2022. On the link between conscious function and general intelligence in humans and machines. *Transactions on Machine Learning Research*.

Juliani, A.; Kanai, R.; and Sasai, S. 2022. The Perceiver architecture is a functional global workspace. In Culbertson, J.; Perfors, A.; Rabagliati, H.; and Ramenzoni, V., eds., *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, 955–961. Toronto, ON: Cognitive Science Society.

Kanerva, P. 1988. *Sparse distributed memory*. Cambridge, MA: The MIT Press.

Kanerva, P. 2009. Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1: 139–159.

Kelly, M. A.; Arora, N.; West, R. L.; and Reitter, D. 2020a. Holographic Declarative Memory: Distributional Semantics as the Architecture of Memory. *Cognitive Science*, 44(11): e12904.

Kelly, M. A.; Ghafurian, M.; West, R. L.; and Reitter, D. 2020b. Indirect associations in learning semantic and syntactic lexical relationships. *Journal of Memory and Language*, 115: 104153.

Kelly, M. A.; Mewhort, D. J. K.; and West, R. L. 2017. The Memory Tesseract: Mathematical Equivalence Between Composite and Separate Storage Memory Models. *Journal of Mathematical Psychology*, 77: 142–155.

Kotseruba, I.; and Tsotsos, J. K. 2018. 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*.

Kruijne, W.; Bohte, S. M.; Roelfsema, P. R.; and Olivers, C. N. 2021. Flexible working memory through selective gating and attentional tagging. *Neural Computation*, 33(1): 1–40.

Kurth-Nelson, Z.; Economides, M.; Dolan, R. J.; and Dayan, P. 2016. Fast sequences of non-spatial state representations in humans. *Neuron*, 91(1): 194–204.

Laird, J. E.; Lebiere, C.; and Rosenbloom, P. S. 2017. A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*, 38(4): 13–26.

Lin, T.; Wang, Y.; Liu, X.; and Qiu, X. 2022. A survey of transformers. *AI Open*, 3: 111–132.

Liu, Y. H.; Smith, S.; Mihalas, S.; Shea-Brown, E.; and Sümbül, U. 2022. Biologically-plausible backpropagation through arbitrary timespans via local neuromodulators. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 17528–17542. Curran Associates, Inc.

Mannering, W. M.; and Jones, M. N. 2021. Catastrophic Interference in Predictive Neural Network Models of Distributional Semantics. *Comput. Brain Behav.*, 4(1): 18–33.

Mashour, G. A.; Roelfsema, P.; Changeux, J.-P.; and Dehaene, S. 2020. Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, 105(5): 776–798.

McCloskey, M.; and Cohen, N. J. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation*, 24(109): 92.

Metz, C. 2023. 'The Godfather of AI' Leaves Google and Warns of Danger Ahead. *New York Times*.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529–533.

Ororbia, A. 2021. Continual competitive memory: A neural system for online task-free lifelong learning. *arXiv preprint arXiv:2106.13300*.

Ororbia, A. 2023. Spiking neural predictive coding for continually learning from data streams. *Neurocomputing*, 544: 126292.

Ororbia, A.; and Kelly, M. A. 2021. Towards a predictive processing implementation of the common model of cognition. *arXiv preprint arXiv:2105.07308*.

Ororbia, A.; and Kifer, D. 2022. The neural coding framework for learning generative models. *Nature communications*, 13(1): 1–14.

Ororbia, A.; Mali, A.; Giles, C. L.; and Kifer, D. 2022. Lifelong neural predictive coding: Learning cumulatively online without forgetting. *Advances in Neural Information Processing Systems*, 35: 5867–5881.

Ororbia, A. G.; and Kelly, M. A. 2022. CogNGen: Constructing the kernel of a hyperdimensional predictive processing cognitive architecture. In Culbertson, J.; Perfors, A.; Rabagliati, H.; and Ramenzoni, V., eds., *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, 1322–1329. Toronto, ON: Cognitive Science Society.

Ororbia, A. G.; and Kelly, M. A. 2023. Maze Learning Using a Hyperdimensional Predictive Processing Cognitive Architecture. In Goertzel, B.; Iklé, M.; Potapov, A.; and Ponomaryov, D., eds., *Artificial General Intelligence*, 321–331. Cham: Springer International Publishing.

Ororbia, A. G.; and Mali, A. 2021. Backprop-free reinforcement learning with active neural generative coding. *arXiv preprint arXiv:2107.07046*.

Ororbia, A. G.; Mali, A.; Giles, C. L.; and Kifer, D. 2020. Continual learning of recurrent neural networks by locally

aligning distributed representations. *IEEE Transactions on Neural Networks and Learning Systems*.

Ororbia, A. G.; Mali, A.; Kifer, D.; and Giles, C. L. 2023. Backpropagation-Free Deep Learning with Recursive Local Representation Alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 9327–9335.

Payeur, A.; Guerguiev, J.; Zenke, F.; Richards, B. A.; and Naud, R. 2021. Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature neuroscience*, 24(7): 1010–1019.

Ramsauer, H.; Schäfl, B.; Lehner, J.; Seidl, P.; Widrich, M.; Adler, T.; Gruber, L.; Holzleitner, M.; Pavlović, M.; Sandve, G. K.; Greiff, V.; Kreil, D.; Kopp, M.; Klambauer, G.; Brandstetter, J.; and Hochreiter, S. 2021. Hopfield Networks is All You Need. arXiv:2008.02217.

Rangel, A.; Camerer, C.; and Montague, P. R. 2008. A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, 9(7): 545–556.

Rao, R. P. N.; and Ballard, D. H. 1999. Predictive coding in the visual cortex: A functional interpretation of some extraclassical receptive-field effects. *Nature Neuroscience*, 2(1): 79–87.

Ritter, F. E.; Tehranchi, F.; and Oury, J. D. 2019. ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(3): e1488.

Roose, K. 2022. An AI-generated picture won an art prize. Artists aren't happy. *The New York Times*.

Rougier, N. P.; and O'Reilly, R. C. 2002. Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cognitive Science*, 26(4): 503–520.

Rudolph, J.; Tan, S.; and Tan, S. 2023. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, 6(1).

Salvatori, T.; Mali, A.; Buckley, C. L.; Lukasiewicz, T.; Rao, R. P.; Friston, K.; and Ororbia, A. 2023. Brain-Inspired Computational Intelligence via Predictive Coding. *arXiv preprint arXiv:2308.07870*.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587): 484–489.

Skaggs, W. E.; and McNaughton, B. L. 1996. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257): 1870–1873.

Steine-Hanson, Z.; Koh, N.; and Stocco, A. 2018. Refining the Common Model of Cognition Through Large Neuroscience Data. *Procedia Computer Science*, 145: 813–820. Papers on the Common Model of Cognition.

Stocco, A. 2018. A biologically plausible action selection system for cognitive architectures: Implications of basal ganglia anatomy for learning and decision-making models. *Cognitive science*, 42(2): 457–490.

Stocco, A.; Laird, J.; Lebiere, C.; and Rosenbloom, P. 2018. Empirical Evidence from Neuroimaging Data for a Standard Model of the Mind. In Rogers, T. T.; Rau, M.; Zhu, X.; and Kalish, C. W., eds., *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, 1094–1099. Austin, TX: Cognitive Science Society.

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

Stocco, A.; Sibert, C.; Steine-Hanson, Z.; Koh, N.; Laird, J. E.; Lebiere, C. J.; and Rosenbloom, P. 2021. Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage*, 235: 118035.

Stokel-Walker, C. 2023. A Professional Artist Was Unfairly Accused Of Using AI. *BuzzFeed News*.

Thomas, R. P.; Dougherty, M. R.; Sprenger, A. M.; and Harbison, J. I. 2008. Diagnostic hypothesis generation and human judgment. *Psychological Review*, 115: 155–185.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L. u.; and Polosukhin, I. 2017. Attention is All you Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30*, 5998–6008. Curran Associates, Inc.

Verma, P.; and Berger, J. 2021. Audio Transformers:Transformer Architectures For Large Scale Audio Understanding. Adieu Convolutions. arXiv:2105.00335.

Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782): 350–354.

Zhang, T.; Xu, H.; Wang, X.; Wu, Y.; Keutzer, K.; Gonzalez, J. E.; and Tian, Y. 2020. Bebold: Exploration beyond the boundary of explored regions. *arXiv preprint arXiv:2012.08621*.