The Perceiver Architecture is a Functional Global Workspace

Arthur Juliani (arthur_juliani@araya.org)
ARAYA Inc., Tokyo, Japan

Ryota Kanai (kanair@araya.org)
ARAYA Inc., Tokyo, Japan

Shuntaro Sasai (sasai_shuntaro@araya.org)
ARAYA Inc., Tokyo, Japan

Abstract

Global Workspace Theory (GWT) has become a prominent functional account of cognitive access in humans and other primates. In the decades since its proposal, there have been a number of computational models developed to study the hypothetical dynamics of the global workspace, most of which are hand-designed to reflect the expectations of the theory. Here we examine a recently successful general deep learning architecture, the Perceiver, as a potential theoretical candidate for the global workspace. We find that despite being developed in an unrelated context, the Perceiver meets a number of theoretical requirements of the global workspace. More importantly, it demonstrates empirical behavior consistent with that expected by GWT in both attentional control and working memory tasks drawn from the cognitive science literature. Taken together, this evidence suggests that the Perceiver and related models may be a useful tool for studying the global workspace and its potential realization in both artificial and biological agents.

Keywords: Global Workspace; Neural Networks; Working Memory; Attentional Control

Introduction

Of the many proposed accounts of cognitive access in humans and other primates, Global Workspace Theory (GWT) has enjoyed some of the most lasting influence (Baars, 1988; Mashour, Roelfsema, Changeux, & Dehaene, 2020). This is due thanks both to its elaborations over time (Dehaene, Kerszberg, & Changeux, 1998a; Dehaene & Changeux, 2005), as well as the empirical evidence collected which supports the theory (Dehaene & Changeux, 2011; Van Vugt et al., 2018). Taken at the highest level, GWT proposes that cognitively accessible information is represented in a network of sustained activity between brain regions referred to as the global workspace. This workspace interacts with a set of otherwise independent information processing modules within the brain, acting as both an information sharing hub as well as a central processing point.

The idea that cognitive access is mediated via a global workspace is supported by a variety of neuroscientific evidence (Mashour et al., 2020). Key to the dynamics of the global workspace are the concepts of ignition and broadcasting. In ignition, information present within one module of the brain is amplified via recurrent processing to the point of crossing a critical threshold where it then enters the global workspace. Information sustained within the global workspace is then broadcast to contextually relevant modules (Dehaene & Changeux, 2011). Research has demonstrated that prefrontal cortex activation through an ignition-like event is required for conscious report of a stimuli, whereas high level visual cortex activation alone is insufficient (Van Vugt et al., 2018). Information passing the computational criteria to cause ignition and broadcasting through the global workspace is intimately connected to the related concept of attentional selectivity.

The global workspace is also implicated in tasks requiring the maintenance and manipulation of abstract information (Baars, 1988). As such, there is a strong connection between broadcasting in the global workspace context and what is typically referred to as working memory (Mashour et al., 2020). In this way, the global workspace serves as a high-level computational description both of cognitive access as well as the executive functions typically associated with frontal cortex in mammals. This connection has seen recent empirical support as well in experimental animal research (Panichello & Buschman, 2021).

Starting from the initially abstract description of the global workspace, there have been a number of attempts to develop more concrete computational models of its hypothetical dynamics (Dehaene et al., 1998a; Dehaene, Sergent, & Changeux, 2003; Whyte & Smith, 2021). Taking the properties of the global workspace as the starting point, many of these models have been hand-designed in order to reflect the desired properties in the model. What has been explored less often is the possibility that already popular machine learning architectures may well map onto the theoretical properties of the global workspace.

One compelling area of research to examine are recent developments in the field of deep learning, where there have been a number of advancements driven by novel neural network architectures and training procedures (LeCun, Bengio, & Hinton, 2015). In particular, recent success has been driven by the development and application of multi-headed attention mechanisms, which enable the sequential processing of dynamic units of information (Vaswani et al., 2017). In this work, we analyze one such recently successful attention-based deep learning architecture, the Perceiver, as a potential theoretical candidate for the global workspace. The Perceiver (Jaegle, Gimeno, et al., 2021), and its more recent instantiation, PerceiverIO (Jaegle, Borgeaud, et al., 2021), consist of a set of cross-attentional and self-attentional operations which utilize a shared workspace, and a variable number of input

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

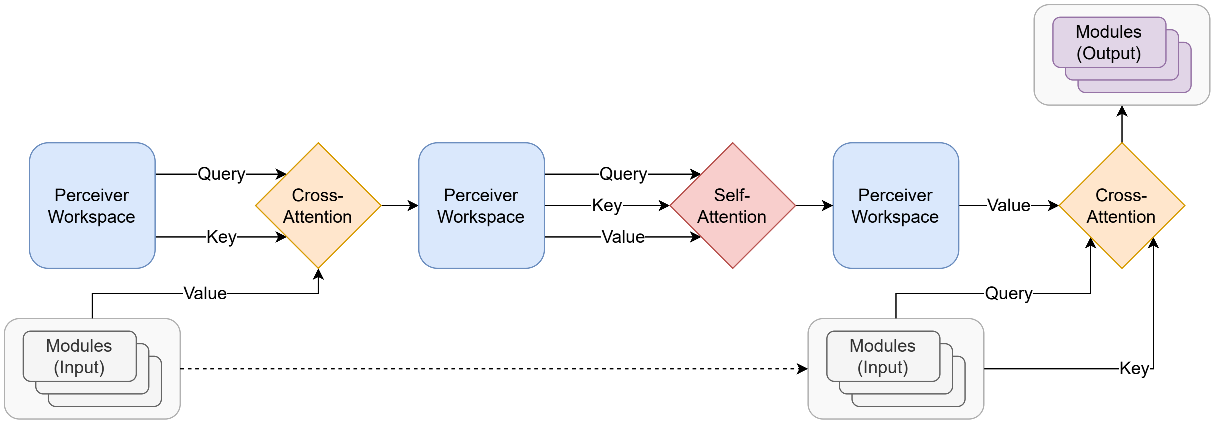

Figure 1: Diagram of the Perceiver architecture within the context of GWT. Information from a set of modules are read into the workspace using a cross-attention mechanism. The contents of the workspace are then maintained and manipulated using a self-attention mechanism. Finally, the information from the workspace is then read back out to the modules utilizing a second cross-attention mechanism. This process is repeated at each time-step of computation within a sequential task.

and output modules. Along with achieving impressive empirical performance on a number of challenging tasks within the machine learning literature, it is also theoretically and functionally similar to another recently proposed architecture which takes explicit inspiration from the global workspace (Goyal et al., 2021). It is this latter connection which motivates our deeper analysis of its connection to GWT.

If we examine the theoretical requirements of the global workspace, we find that the ability to read and write between a dynamic set of modules (ignition and broadcasting), the ability to apply selective attention to that process (attentional control), and the ability to maintain and manipulate information within the workspace over time (working memory) are all core requirements of any potential computational model. Despite being developed in an independent context and with unrelated goals in mind, the Perceiver meets all of the theoretical requirements of the GWT. It support reading and writing from a dynamic set of modules, utilizes attentional mechanisms at multiple levels of processing, and supports the robust maintenance and manipulation of information within its memory storage.

In addition to theoretical considerations, we consider an empirical evaluation of Perceiver from the perspective of the expectations of GWT. While Perceiver has been implemented and validated in a number of large-scale machine learning problems (Jaegle, Gimeno, et al., 2021), it has not been examined in this more restrictive domain as a model of cognitive access. Here specifically, we examine n-back and cue-recall tasks inspired by the cognitive science literature of attentional control and working memory (Rosen & Engle, 1997; Kane, Conway, Miura, & Colflesh, 2007). We find that in all behavioral tasks Perceiver behaves consistently with the expectations of the GWT, while various ablations of the model fail at one or more of these tasks, suggesting that both cross-attention and self-attention are necessary for any valid computational model of the global workspace.

Given both the theoretical alignment as well as the empirical validation, we believe that Perceiver and the related models such as (Goyal et al., 2021), can serve as potentially useful new tools to better understand GWT and its potential realization in both artificial and biological agents. While this work explores theoretical and behavioral expectations of these models, we hope that future work can elucidate complementary results with respect to representational properties and underlying neural dynamics of these models. Of particular interest would be to analyze the population dynamics of these networks as they compare to neural data, demonstrating not only a theoretical and behavioral connection, but a representational one as well.

Theoretical Comparison

We can derive from the description of the GWT provided by (Baars, 1988; Dehaene, Kerszberg, & Changeux, 1998b) a set of criteria by which to judge a candidate neural network implementation of the theory. Here we consider four specific abstract computational properties of the GWT and analyze the extent to which the Perceiver and its potential variants meet these criteria.

Table 1: Table comparing theoretical properties of different candidate model architectures.

Criteria	**FF-ID**	**FF-SA**	**CA-ID**	**CA-SA**
Dynamic Modules	✗	✗	✓	✓
Selective Attention	✗	✗	✓	✓
WM (Maintain)	✓	✓	✓	✓
WM (Manipulate)	✗	✓	✗	✓

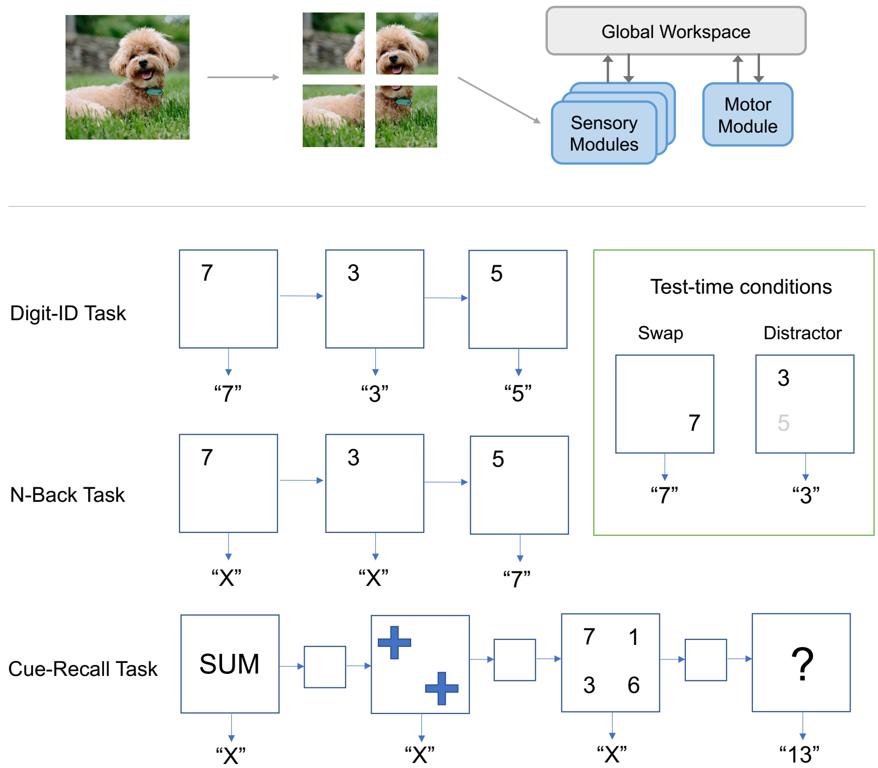

Figure 2: **Top**: Diagram of modules used for behavioral tasks. Each time-step an image is divided into four quadrants, which are each represented as separate modules to the workspace. **Bottom**: Diagram of the various behavioral tasks and their variants. Boxes correspond to images presented to models during a trial. Bottom arrows correspond to expected behavioral outputs from model.

Global Workspace Properties

The first property we consider is the ability to interact with a dynamic set of modules. While the neural anatomy of the brain is largely fixed, what counts as a “module” within the GWT is dynamically determined based on population activity across multiple brain regions at a given time. As such, a neural network model of the global workspace should be able to take input from an unordered set of modules which potentially changes over time.

The second property we consider is whether the model has the capacity for selective attention over the set of modules, corresponding to so-called ignition and broadcasting. While there has been some work to dissociate the more general cognitive phenomena of attention from conscious access (Koch & Tsuchiya, 2007), proponents of GWT suggest that information entering the workspace, and thus becoming cognitive available for broadcast can be thought of as a kind of high-level attention selection process (Mashour et al., 2020). Within the GWT, attentional gating takes place both during the read phase of processing where the workspace reads from the set of modules as well as during the write phase where information from the global workspace is broadcast to a set of relevant modules.

The third and fourth properties under consideration are whether the global workspace possesses the capacity to maintain and manipulate information over time. These capacities can be seen as together being largely consistent with the concept of working memory, and is distinguished from other forms of short term memory in that it is amenable to conscious manipulation, as well as requires conscious effort to maintain information in the face of potential distractor information competing to enter the workspace.

Candidate Model Architectures

We compare the Perceiver neural network architecture to three ablation models derived from the Perceiver which have aspects of the full model removed or altered. We select these ablated models in order to demonstrate the necessity of the full architecture in meeting the criteria of the global workspace.

We begin our analysis with the full Perceiver architecture, which utilizes both cross-attention operations for reading and writing, as well as self-attention operations for recurrent processing. We refer to this model as **CA-SA**. Each of the proposed variants modifies one or more of these at-

Table 2: Table comparing the the various behavioral tasks according to their ability to test for the presence of the global workspace criteria defined above.

Behavioral Task	Dynamic Modules	Selective Attention	WM (Maintain)	WM (Manipulate)
Digit ID	✗	✗	✗	✗
Digit ID (Swap)	✓	✗	✗	✗
Digit ID (Distractor)	✗	✓	✗	✗
N-Back	✗	✗	✓	✗
N-Back (Swap)	✓	✗	✓	✗
N-Back (Distractor)	✗	✓	✓	✗
Cue-Recall	✓	✓	✓	✓

tentional mechanisms. Instead of cross-attention for input and output, we propose models which utilize a concatenation between modules and the hidden state of the workspace as input to a multi-layer perceptron. Likewise, output from the model can be computed using another multi-layer perceptron. We refer to this architecture as **FF-SA** when paired with a self-attention recurrence. Secondly, we consider replacing the self-attention recurrence operation with a simple identity operation between input and output. We refer to this as **CA-ID** when paired with cross-attention, and **FF-ID** when paired with the simpler read/write operation.

We can first compare these candidate architectures according to their capacity to handle dynamic sets of modules. Due to the requirement of a fixed input space at each time-step, the FF-ID and FF-SA models are unable to support dynamic ensembles of modules in either the read or write context. This is the case both for sets of modules which vary in their size from time-step to time-step as well as sets of modules which vary in their order from time-step to time-step. In contrast, the CA-ID and CA-SA architectures both utilize cross-attention operations over the set of input and output modules. This enables support for both dynamically sized and dynamically ordered sets of modules at each time-step.

We next compare these architectures according to their capacity to support selective attention over the contents of the set of modules. The FF-ID and FF-SA models have limited capacity for selective attention, as it must be the result of fixed learned weight matrices which are content agnostic. In contrast, because the CA-ID and CA-SA architectures utilize cross-attention in the reading and writing process, the contents of each module can be selectively attended to or completely ignored based on the behavioral needs of the task.

Finally, we compare the capacity of each of these architectures to support both the maintenance and manipulation of information in working memory. Considering the FF-ID and CA-ID models, we find that while information is retained between time-steps, thus supporting maintenance, there is no explicit process by which information can be manipulated over time, a key aspect of working memory. As such, these models capacity for working memory is only partial. In contrast, the FF-SA and CA-SA architecture utilizes an explicit self-attention mechanism between reading and writing to and from the set of modules. The result is that these models are capable of both maintaining and manipulating information present within the workspace. Table provides a summary of the various theoretically expected capabilities of each model architecture. In the following section, we consider the extent to which the models empirical performance matches these expectations.

Empirical Comparison

Our theoretical examination suggests that the full Perceiver architecture (CA-SA) is best suited as a neural network implementation of the global workspace, and that removing or altering either of its attention mechanisms will reduce this functionality. In this section, we employ a set of behavioral tasks to empirically validate this theoretical expectation. The suite of behavioral tasks was developed to address the four identified properties of a global workspace: the ability to handle dynamic sets of modules, the ability to selectively admit information into the workspace, and the ability to maintain and manipulate that information over time in the workspace. Concretely, we utilize tasks based on n-back and cue-recall paradigms drawn from the cognitive science literature (Rosen & Engle, 1997; Kane et al., 2007).

Methods

We utilize a total of seven different behavioral tasks, based on modifications of the n-back and cue-recall paradigms. Each task is separated into a series of learning trials, followed by a series of test-time trials. The models are trained from data collected during the learning trials, and evaluated based on their performance during the test-time trials. Depending on the task, the nature of the test-time trials may differ in key aspects from the learning trials. The ability of each model to adapt to the test-time conditions determines its performance. Table provides an overview of each of the behavioral tasks, and which properties of the global workspace they are designed to evaluate.

In the *Digit-ID* task, a series of randomly selected digits valued between 0 and 9 are presented in a single quadrant of the screen in sequential order for the duration of a trial. The goal of the task is to provide as output the value

Table 3: Table comparing performance of each of the model architectures on the various behavioral tasks. Each cell presents the mean and standard deviation of the model in a given task computed over five separate experiments.

Model	Digit ID	Digit ID (Swap)	Digit ID (Dist)	N-Back	N-Back (Swap)	N-Back (Dist)	Cue-Recall
FF-ID	.97 (.00)	.06 (.04)	.75 (.12)	.94 (.00)	.22 (.00)	.51 (.11)	.03 (.03)
FF-SA	.98 (.00)	.08 (.03)	.83 (.10)	.96 (.00)	.21 (.01)	.32 (0.1)	.02 (.00)
CA-ID	.98 (.00)	.98 (.00)	.91 (.05)	.87 (.12)	.71 (.23)	.32 (.01)	.56 (.02)
CA-SA	.97 (.00)	.93 (.03)	.74 (.08)	.96 (.02)	.69 (.26)	.72 (.17)	.56 (.03)

of the currently-presented digit at each time-step. In *Digit-ID (Swap)* the location of the digit presentation is changed between the learning trials and the test-time trials. In *Digit-ID (Distractor)* an additional distractor digit is presented in a separate quadrant of the screen during test-time that was not present during training. This additional digit is presented in a lighter color than the true target digit.

In the *N-Back* task the same digit presentation as in the *Digit-ID* tasks is used, with the exception that the goal is to output at each time-step the value of the digit presented two time-steps in the past. For the initial two time-steps a null-token is expected as the output. In the *N-Back (Swap)* condition the position of the digits changes to a separate quadrant during the test-time trials. In the *N-Back (Distractor)* task an additional distractor digit is presented in a separate quadrant of the screen during test-time that was not present during training.

Finally, in the *Cue-Recall* task, the model is presented with a series of images for each trial whose contents vary depending on the time-step of the trial. The first image presented is a visual symbol denoting the trial type, which can consist of either "SUM," "MIN," or "MAX." There is then a blank image presented, followed by between one and four cues denoted by a "+" sign in the four quadrants of the screen. There is then another blank image followed by a set of digits between 0 and 9 presented in the four quadrants. After a final blank image, there is then a prompt to provide a behavioral response corresponding to the denoted mathematical operation applied to the set of cued digits. Between the training and testing trials the set of possible cue combinations is changed. See Figure 2 for a schematic of each of these behavioral tasks.

Given each of the behavioral tasks, the training procedure for the neural network models is as follows. Each task consists of a sequence of 64*x*64 images provided to each model. Each model produces a behavioral output at each time-step, along with a reconstruction of the image input. In each model, the image is initially pre-processed by a convolutional neural network (CNN) (LeCun, Bengio, et al., 1995), and the output of the CNN is split into four separate "visual modules." Each visual model also possesses a learned positional encoding (Vaswani et al., 2017). An additional "motor module" is used to provide a learned encoding for the behavioral output.

All models are trained using gradient descent to minimize mean-squared error on the image reconstruction as well as the cross entropy loss of the target behavioral output. All models utilize hidden layers of size 256, and a representational capacity of 512 units for the workspace itself. The workspace capacity is evenly divided between four sub-units. All models are trained with the same learning rate of $1e^{-4}$ and an L2 regularization loss of $2e^{-5}$ using the Adam optimizer (Kingma & Ba, 2014), and same architecture, with the exception of the workspace module itself. This results in small but insignificant differences in total parameter sizes between the models. Each model was trained for a total of ten epochs on the training set of trials, and evaluated using a separate test set of trials. We repeated each experiment seven times per model in order to perform statistical analysis of relative model performance.

Results

We begin our analysis with the simplest task requiring none of the stated properties of the global workspace, *Digit-ID*. Unsurprisingly, we find that all four models are able to perform this task with near-perfect accuracy ($Mean > 0.95$). With a validation of the basic functioning of each model, we can examine model performance according to the four criteria of interest for a global workspace. We first consider the ability of the model to maintain information within the workspace, as evaluated by the *N-Back* task. Here we find no significant differences between the four models (one-way ANOVA $p > 0.05$), with all models solving the task with a high level of accuracy ($Mean > 0.85$).

We then consider the ability of the model to manage a dynamic set of modules, as evaluated by the *Digit-ID (Swap)* and *N-Back (Swap)* tasks. Here find that the models utilizing cross attention (CA-ID and CA-SA) both significantly outperform the models without cross attention (FF-ID and FF-SA) in both of the tasks (independent t-tests $p < 0.01$). This result demonstrates that the cross-attention mechanism provides a strong inductive bias towards permutation invariance in the model architectures.

We next consider the third criteria of the global workspace, the ability to selectively attend to information present in the set of modules. To examine this, we utilized the *Digit-ID (Distractor)* and *N-Back (Distractor)* tasks. With the *Digit-ID (Distractor)* task, we find mixed results. Here the best performing model, CA-ID, significantly outperforms the CA-SA and FF-ID models ($p < 0.5$), but not the FF-SA model ($p = 0.09$). In the case of the *N-Back (Distractor)* task, we find that the best performing model, CA-SA, outperforms the

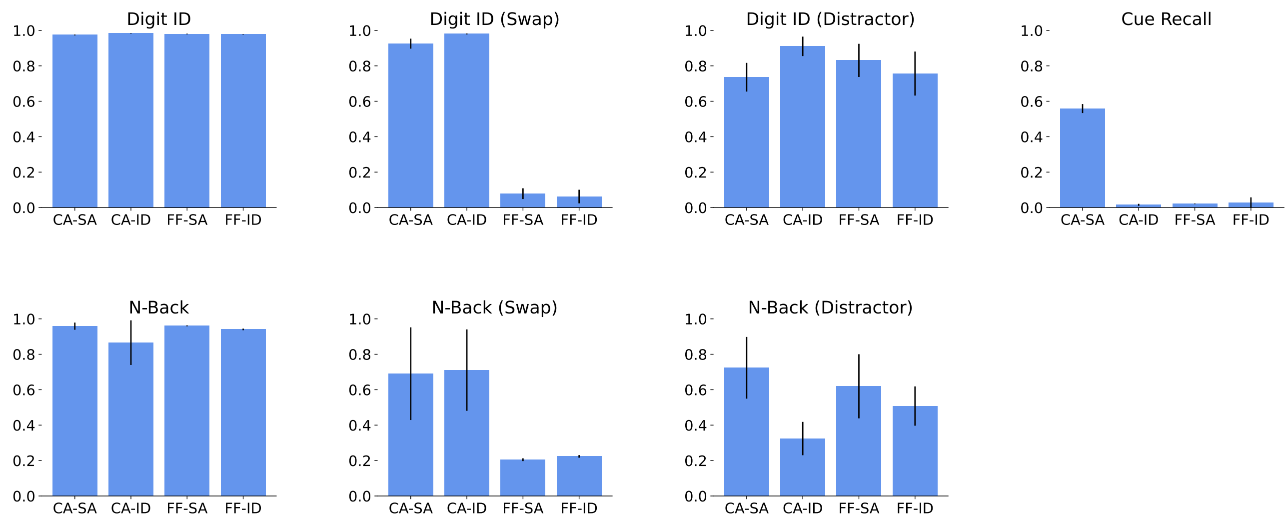

Figure 3: Bar plots displaying mean performance of each model on the various behavioral tasks. Error bars correspond to standard deviations. Values computed over seven separate experiments.

CA-ID and FF-ID models ($p < 0.05$), but did not significantly outperform the FF-SA model ($p > 0.05$). These results suggest that the presence of self-attention was a significant advantage in the task, and that cross-attention, when paired with self-attention provided additional benefit.

The last task we consider, *Cue-Recall* evaluates all four global workspace criteria simultaneously. As such, it is the most comprehensive task of a global workspace like mechanism. Here we find that the CA-SA model performs with greater than 50% accuracy ($Mean = 0.56$), while also significantly outperforming the other three models (independent t-tests $p < 0.01$), none of which achieve an accuracy level above chance. This suggests that both cross-attention and self-attention are necessary for this task. The full results of the experiments are presented in Table and Figure 3.

Discussion

In this work we analyzed the properties of the Perceiver architecture in light of its ability to meet the criteria of the global workspace, as outlined in (Baars, 1988) and (Dehaene et al., 1998b). We identified the ability to handle a dynamic set of modules, the ability to selectively attend to information in those modules, and the ability to maintain and manipulate information within the workspace as four key criteria for evaluation. We then compared the Perceiver to three potential variants derived from specific architectural changes and found that from both a theoretical and empirical perspective, the full Perceiver architecture best meets the criteria, thus serving as a potential model of the global workspace. In particular, we found that the cross-attention mechanisms for input and output, as well as the self-attention for recurrent processing were essential to demonstrating behavior which would be expected from a global workspace. Given Perceiver's strong empirical performance on a diverse set of large-scale learning problems (Jaegle, Borgeaud, et al., 2021), there is perhaps a case to be made that the general principles behind the global workspace can indeed support a complex set of reasoning abilities in humans and machines (Mashour et al., 2020).

Beyond the Perceiver architecture, there are many other opportunities for modern deep learning architectures which implements aspects of GWT. The recently proposed global workspace module from Goyal et al. is architecturally similar to the Perceiver, and we expect it to display similar advantages. Indeed, this is perhaps not surprising, given that the model is explicitly motivated by the global workspace architecture. Not addressed here is the larger question regarding the kind of representational space induced by a global workspace, a topic recently addressed by (VanRullen & Kanai, 2021). In the wider field of computer science, there have also been proposals for first-principles approaches to global workspace like architectures (Blum & Blum, 2021), as well as efforts to integrate global workspace models into larger cognitive architectures (Juliani, Arulkumaran, Sasai, & Kanai, 2022). One interesting future avenue of research would be to move beyond behavioral tasks and examine the learned representations and neural dynamics of the Perceiver and other candidate neural network architectures. Of particular interest would be the extent to which these too match the expectations of the global workspace, and recent neuroscience work inspired by it (Van Vugt et al., 2018).

Acknowledgments

We would like to thank Leonardo Barbosa, Yasuo Kabe, Motoshige Sato, and Kai Arulkumaran for their helpful feedback during the process of developing the experiments presented in this paper.

This work was supported by JST, Moonshot R&D Grant Number JPMJMS2012.

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Blum, M., & Blum, L. (2021). A theoretical computer science perspective on consciousness. *Journal of Artificial Intelligence and Consciousness*, *8*(01), 1–42.

Dehaene, S., & Changeux, J.-P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol*, *3*(5), e141.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998a). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, *95*(24), 14529–14534.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998b). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, *95*(24), 14529–14534.

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*, *100*(14), 8520–8525.

Goyal, A., Didolkar, A., Lamb, A., Badola, K., Ke, N. R., Rahaman, N., … Bengio, Y. (2021). Coordination among neural modules through a shared global workspace. *arXiv preprint arXiv:2103.01197*.

Jaegle, A., Borgeaud, S., Alayrac, J.-B., Doersch, C., Ionescu, C., Ding, D., … others (2021). Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv:2107.14795*.

Jaegle, A., Gimeno, F., Brock, A., Zisserman, A., Vinyals, O., & Carreira, J. (2021). Perceiver: General perception with iterative attention. *arXiv preprint arXiv:2103.03206*.

Juliani, A., Arulkumaran, K., Sasai, S., & Kanai, R. (2022). On the link between conscious function and general intelligence in humans and machines. *arXiv preprint arXiv:2204.05133*.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the n-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in cognitive sciences*, *11*(1), 16–22.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798.

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601–605.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211.

VanRullen, R., & Kanai, R. (2021). Deep learning and the global workspace theory. *Trends in Neurosciences*.

Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, *360*(6388), 537–542.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Whyte, C. J., & Smith, R. (2021). The predictive global neuronal workspace: A formal active inference model of visual consciousness. *Progress in Neurobiology*, *199*, 101918.