bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A computational examination of the two-streams hypothesis: which pathway needs a longer memory?

Abolfazl Alipour[1,2], John Beggs[2,3], Joshua Brown[1,2], Thomas James[1,2*]

1. Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA
2. Program in Neuroscience, Indiana University, Bloomington, Indiana, USA
3. Department of Physics, Indiana University, Bloomington, Indiana, USA

Abstract

The two visual streams hypothesis is a robust framework that has inspired many studies in the past three decades. One of the well-studied claims of this hypothesis is the idea that the dorsal visual pathway is involved in visually guided motor behavior, and it is operating with a short memory. Conversely, this hypothesis claims that the ventral visual pathway is involved in object classification, and it works using a long-term memory. In this study, we tested these claims by training identical recurrent neural networks to either perform viewpoint-invariant object classification (a task attributed to dorsal stream) or orientation classification (a task attributed to dorsal stream) and measured how much they rely on their memory in each task. Using a modified leaky-integrator echo state recurrent network, we found that object classification requires a longer memory compared to orientation classification. However, when we used long-short-term memory (LSTM) networks, we observed that object classification requires longer memory only in larger datasets. Accordingly, our results suggest that having a longer memory is advantageous in performing ventral stream's tasks more than their dorsal counterparts, as was originally suggested by the two-streams hypothesis.

Keywords

Two-streams hypothesis, dorsal and ventral visual pathway, LSTM, convolution neural network, memory, time scale

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Introduction

The hypothesis that primate vision might be composed of two distinct visual systems was first proposed in 1968 (Ettlinger, 1990; Schneider, 1969; Trevarthen, 1968). In its initial form, this hypothesis suggested that a genicolustriate system is processes object recognition and a tectofugal system processes spatial information. Subsequently, using evidence from lesion studies, Ungerleider and Mishkin suggested that the dissociation between the two pathways also exists at the cortical level where both pathways are fed by axons originating from striate cortex (Mishkin & Ungerleider, 1982). Eventually, an extended two-streams hypothesis was proposed in a seminal paper by Goodale and Milner in 1992 (Melvyn A. Goodale & Milner, 1992). According to their two-streams hypothesis, the ventral visual stream (from occipital to temporal cortex) is heavily involved in object recognition while the dorsal visual stream (occipital to parietal) is involved in visually guided motor behaviors.

Using clinical and experimental evidence, Goodale and Milner argued for a double dissociation in the primate visual system and listed a series of major functional differences between the ventral and dorsal pathways. First, the ventral stream was more involved in conscious perception while dorsal pathway performance was unconscious. Secondly, the ventral stream had a slower processing speed and a longer memory, while the dorsal pathway maintains a faster processing speed and shorter memory (Milner & Goodale, 2008; Norman, 2002). Accordingly, having these processing features was believed to benefit the functions of each pathway.

Among the neuropsychological studies on the differences between dorsal and ventral stream, the differences between memory spans in the two pathways is particularly interesting. On the one hand, patients with damage to their parietal cortex (dorsal stream) were found to have lost their ability to correctly reach towards visual targets (optic ataxia) while they could still recognize the identity of objects due to their intact ventral stream. On the other hand, patient D.F. who had impaired connectivity between her V1 and inferior temporal cortex was found to be unable to recognize objects or faces, (visual form agnosia) while she

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was able to perform reaching movements similar to healthy controls (M. A. Goodale et al., 1991; Milner et al., 1991).

Interestingly, when optic ataxia patients were asked to point to a target location after a 5-second delay, their performance improved in a paradoxical manner while in healthy subjects, the same delay deteriorated their performance (Milner et al., 1999). This suggested that optic ataxia patients are not able to use their dorsal stream in no-delay trials, but they could retrieve some information from their intact ventral stream in delayed trials. In contrast, when patient D.F. was presented with two plaques of different widths, she could not report the width of plaques using her index finger and her thumb. However, when asked to reach for a specific plaque, her grip aperture (distance between index finger and thumb) was like healthy controls and correlated with the width of each plaque. Intriguingly, this patient was not able to perform this visuomotor reaching task after a 2-second delay as if she has 'lost' the information required to complete the task (M. A. Goodale et al., 1994). This suggested that after a delay, stored information in the ventral stream is used for reaching. Further studies by Hu and Goodale using healthy participants consolidated the idea that real time visuomotor control uses a short memory size estimation mechanism as compared to conscious size estimation or motor control after delay which uses long term memory and conscious task engagement (Hu et al., 1999; Hu & Goodale, 2000). These observations lend credit to the idea that the dorsal stream has a shorter memory and it is crucial for visually guided behavior while the ventral stream has a longer memory and it is crucial for object recognition.

Accordingly, we sought to test these hypotheses in the current study. Specifically, we tested if there is a relationship between the tasks ascribed to each of the pathways and the length of optimal memory for each task. In other words, is short term memory beneficial for tasks that are attributed to dorsal stream (e.g., size or orientation classification) and ventral stream functions such as viewpoint-invariant object classification need longer memory?

To this end, we trained simple convolutional neural networks (CNNs) to either perform orientation classification or viewpoint-invariant object classification. Subsequently, we replaced the last layer of the trained CNNs with recurrent

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

networks and trained the new networks to recognize frames of rotating objects or frames of different objects with the same orientation or width. Our results suggest that a longer memory benefits viewpoint object classification, while a shorter memory is helpful for orientation classification tasks.

2 methods

2.1 datasets

We created 16 synthetic objects and rotated each object around its vertical axis (yaw) to generate images from different viewpoints of each object. Subsequently, to create two tasks from the same set of objects, we derived a dataset for a viewpoint invariant object classification task and a dataset for width classification of objects.

To create the viewpoint invariant object classification dataset, we defined 16 object classes where each class had 22 pictures of the same object rotated along its vertical (yaw) axis (16.3 degrees increments in each frame). This dataset contained 352 images in total (16x22).

To avoid the huge number of orientation classes that can be created when an object is rotated along all of its three axes, we decided to create a simplified orientation classification dataset in which we only changed the width of objects by moving them along their yaw axis. Accordingly, we defined six classes where each class contained 16 different objects with the same width. Due to the presence of different vertical orientations with the same width (e.g., 0 and 180 degrees), the same width of one object was repeated twice, and this resulted in 32 images per class and a dataset with 192 images in total (6x32). The width classes were as following:

- Width class No. 1: All objects with an orientation of 0° and 180°
- Width class No. 2: All objects with an orientation of 16° and 164°
- Width class No. 3: All objects with an orientation of 33° and 147°
- Width class No. 4: All objects with an orientation of 49° and 131°

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- Width class No. 5: All objects with an orientation of 65° and 115°
- Width class No. 6: All objects with an orientation of 82° and 98°

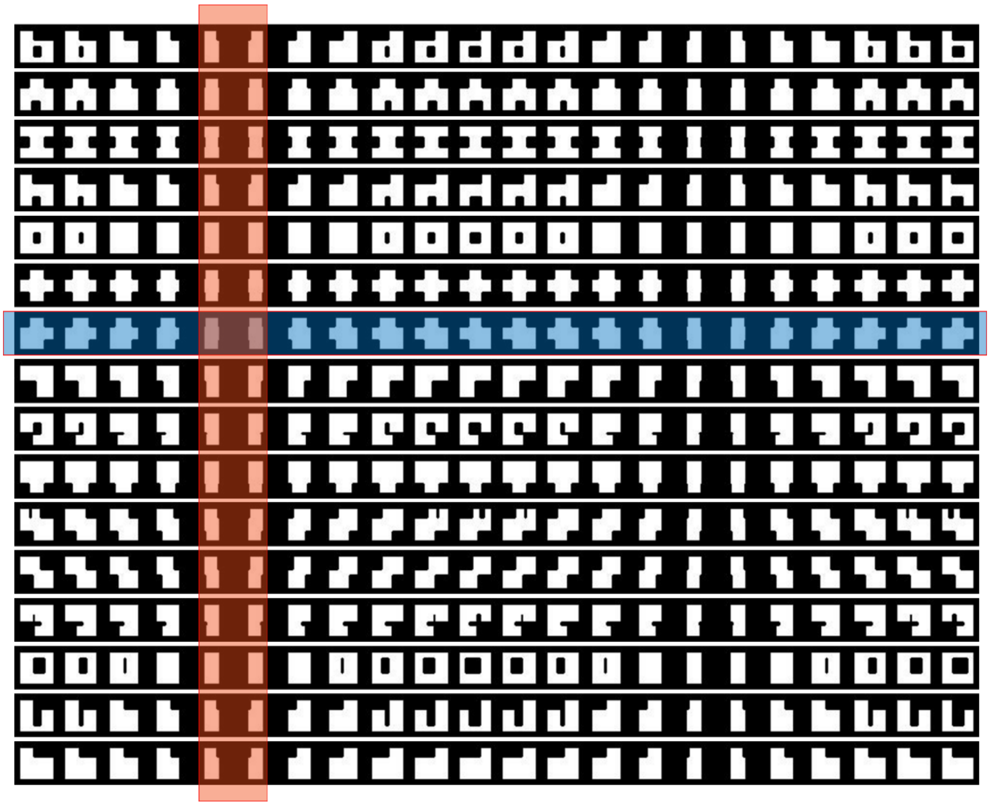


Fig. 1. Structure of the synthetic objects' dataset. Each row in this matrix forms an object class that was used for viewpoint invariant object classification task. The region indicated by blue is an example of an object class that was split into the training set (with a probability of 80%) and test set (with a probability of 20%). Conversely, the orange shaded line indicates images that were put into a width class (32 images) and further split into a training and test set with 80% and 20% probability, respectively. Note that only 6 width classes were derived from this dataset and therefore, some of the columns were not used in width classification task. See supplementary figure 1 for a detailed illustration of width classes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To see if the results were specific to the synthetic dataset, we repeated our experiments using the Columbia Object Image Library (COIL-100) dataset (Nene et al., 1996). This dataset contains images of 100 natural objects, and there are 72 images of the same object from different viewpoints in 5-degree increments in each object class.

In COIL-100 dataset, we derived an orientation classification dataset instead of width classification dataset to see if our results will generalize to more naturalistic stimuli. The orientation classification dataset was derived according to the following procedure: first, 46 objects in the COIL-100 dataset that show ambiguous orientation as they are rotated (bottles, cans, cups) were removed and we only used the remaining 54 objects for both of our datasets.

For orientation classification, we used four simplified orientation classes as following:

- Class 1: All orientations between 0°±15° (5-degree increments) and their mirror orientations (180°±15°) regardless of the object identity.
- Class 2: All orientations of 45°±15° (5-degree increments) and their mirror orientations (225°±15°) regardless of the object identity.
- Class 3: All orientations between 90°±15° (5-degree increments) and their mirror orientations (270°±15°) regardless of the object identity.
- Class 4: All orientations between 135°±15° (5-degree increments) and their mirror orientations (315°±15°) regardless of the object identity.

This resulted in four classes, each with 756 images (a total of 3,024 images). which was used for training networks on orientation classification task. Each image was grayscaled and then downscaled to a 28 by 28 pixels resolution before being used in training.

To make our object classification dataset consistent with the orientation classification, we only used the 54 object that have been used in our orientation classification task for object classification as well. In each of the 54 object classes we had 72 images of the same object from different viewpoints (a total of 3,888 images)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

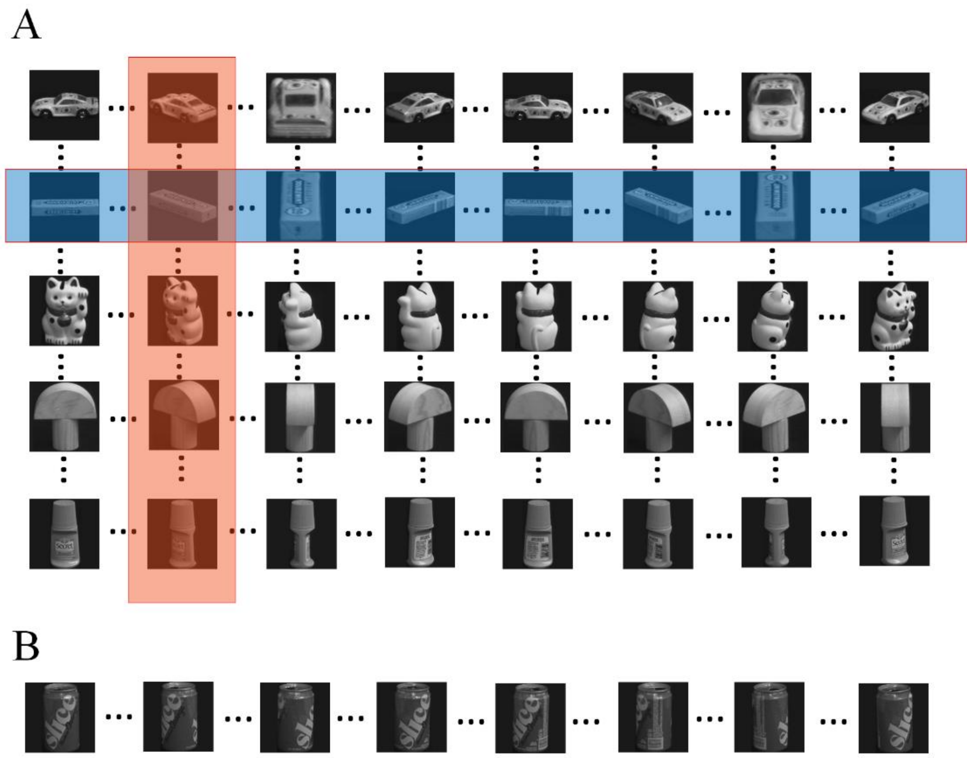


Fig. 2. Example images from the COIL-100 dataset. COIL-100 contains 100 object classes each with 72 images of the same object from different viewpoints (5-degree increments). Objects are shown with 45-degree increments and rotating clockwise. The blue shade covering one row of the image array represents an object class while the light red shade covering an orientation with multiple objects represent central elements of an orientation class. An orientation class was defined as all object images with a central orientation (e.g. 45°**)**, images with ±15° orientations around the central orientation plus all of their mirror orientations (e.g. 225°±15°) B. Some objects (e.g. the soda can) have an ambiguous orientation as they are moved along their vertical axis. These objects were excluded from the dataset. The resulting dataset after removing such objects contained 54 objects in total which was subsequently used for both tasks.

2.2 Networks and training procedure

Given that the convolutional neural networks (CNNs) show a reasonable similarity to primate visual system (Cadieu et al., 2014; Schrimpf et al., 2018; D.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

L. Yamins et al., 2013; D. L. K. Yamins & DiCarlo, 2016) and the fact that both dorsal and ventral pathways receive information from primary visual areas (except for some subcortical inputs to the dorsal pathway), we used CNNs as simple approximations of early visual areas.

To this end, we made two architecturally identical convolutional neural networks (details in figure 1) and trained one of them on the viewpoint invariant object classification dataset and the other one on the width classification dataset (orientation classification in COIL-100). The task of the networks at this stage was to recognize if one image belonged to a specific object class (in case of the object classifier network) or a width/orientation class (in case of width/orientation classifier network). At this stage, networks had to classify individual images.

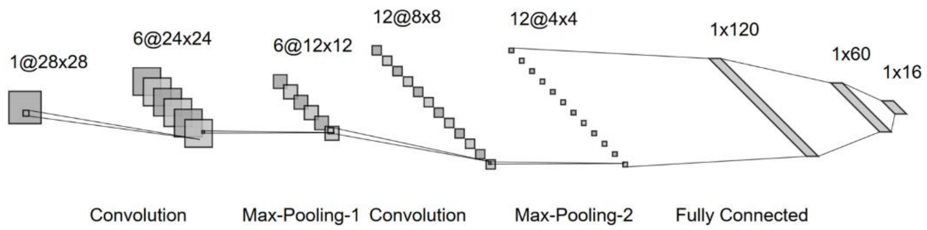


Fig. 3. Network architecture. Our classifier network comprises 2 convolutional layers with max-pooling and 3 fully connected layers. Note that here, fully connected layer is referred to a layer is fully connected to all other nodes in its previous layer and it does not refer to any recurrence within the layer itself. This network was trained on either viewpoint invariant object classification (object classifier) or width/ orientation classification. Subsequently, it was modified by adding recurrent layers to perform sequence learning (see Fig. 4).

Next, to see if temporal information (i.e., memory) in a sequence of images is important for the performance of the network, we froze the weights in the feedforward network and replaced the last layer in each of these two networks with either a leaky-integrator echo state network (LiESN) (experiment 1) or a long-short term memory (LSTM) network (experiment 2). This was done based

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on previous studies showing the effectiveness of transfer learning (Tan et al., 2018). Accordingly, these new networks could learn sequences of stimuli. Hence, their task was to recognize images in a sequence of frames. For viewpoint invariant object recognition, the stimulus was a sequence of frames that contained different viewpoints of the same object. In each training epoch, the network was fed with this sequence to classify the object identity of the frames in the sequence. For width/orientation detection, the stimulus was a sequence of frames that contained different objects with the same width/orientation. Similarly, in each training epoch, the network was fed with a sequence of frames to classify the widths/orientations.

In experiment 1.A, we replaced the last layer of classifiers by a leaky-integrator echo state network and trained them on either viewpoint-invariant object classification or width/orientation classification. The last layer of the viewpoint-invariant object classifier was replaced with an LiESN and trained to recognize images of the same objects from different viewpoints. Similarly, the last layer of the width/orientation classifier network was replaced with an LiESN, and the resulting network was trained to recognize the width/orientation in a series of images containing different objects with similar widths/orientations.

In a standard LiESN, the dynamics of the network with N_U inputs and N_R reservoir units were governed according to:

$$X(t) = (1-a)\boldsymbol{X}(t-1) + a\tanh\big(W_{in}u(t) + \theta + Wx(t-1)\big) \qquad (1)$$

Where X is the state of reservoir with N_R dimensions, t is the time, u is the input signal with N_U dimensions, W_{in} is the input weight matrix (N_U by N_R) and Wx is the reservoir weight matrix (N_R by N_R). The θ is the bias term, and a is the leaking rate (Gallicchio et al., 2017; Jaeger, n.d.; Schaetti et al., 2016).

As seen in Equation 1, to ensure the stability of the outputs in the standard LiESN networks, the leaking rate specifies the dependence on past time steps

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

relative to dependence on the current input by the computation in each step. In other words, when leaking rate goes to zero, the network solely depends on its past state and ignores the current inputs, and when the leaking rate goes to 1, the network solely relies on current inputs and ignores the past states.

To disentangle the role of past states on task performance from the importance of current inputs, we disjointed the memory dependence from current inputs by removing the effect of leaking rate on current inputs:

$$X(t) = (1-a)\boldsymbol{X}(t-1) + \tanh\big(W_{in}u(t) + \theta + Wx(t-1)\big) \quad (2)$$

As seen in Equation 2, the leaking rate a does not affect the results of the $\tanh()$ term of the equation. We will call the a in this equation the forget rate from now on. In experiment 1.a, we observed changes in accuracy and forget rate as we trained the networks to either recognize objects or widths/orientations. We ran 18 training epochs and measured the test accuracy and forget rates after each training epoch. This simulation was repeated 30 times, and the average test accuracies with their corresponding forget rates were obtained for comparison.

In experiment 1.b, we performed a parameter sweep of 50 equally spaced forget rates between 0 and 1 (0.02 steps) to see which task was harmed by larger forget rates. To control for the effect of network size on our results, we performed this parameter sweep on 6 different network sizes (n=20,40,80,160,320,640). The echo state networks were trained for 50 epochs, and their test accuracy was measured afterward.

In experiment 2, we used LSTM networks instead of LiESN networks to see if our results were independent from the type of recurrent networks used in our experiments. LSTM network was implemented based on the original LSTM paper (Hochreiter & Schmidhuber, 1997).

To measure how much of the past information (history, memory) is relevant to the classification performance in an LSTM network, we used the ratio of forget gate to input gate as an indicator of network reliance on the past information. Forget gate value in an LSTM network is a number between zero and 1, and it

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is multiplied by the hidden state value to control the effect of the past information (hidden states) on the current output of the LSTM network. The input gate value in an LSTM network is also a number between 0 and 1, and it is multiplied by the current input of the LSTM network to indicate how much of the current input should affect the cell state of an LSTM network. Accordingly, the ratio of forget gate to input gate can show how much a given network is relying on its memory (hidden states) as compared to its current inputs to generate an output. The larger this ratio, the more reliance there is on the past information (memory) in the LSTM network.

In experiment 2.a, to obtain these ratios with the synthetic dataset, we trained our LSTM network on each of the tasks. Subsequently, we fed the test data to the network and measured the logarithm of the ratio of 'forget gate values to input gate values' in the LSTM network as images were passing through the network. The average logarithm of (forget gate / input gate) for the entire test dataset was calculated for both width/orientation classification and object classification tasks.

In experiment 2.b, we trained the LSTM network using the COIL-100 dataset to see if the small size of the synthetic dataset affected our results. To obtain these ratios in the COIL-100 dataset, we used the same procedure as experiment 2.a.

Like LiESN networks, to control for the effect of network size on our results, we ran all of the experiments on six different network sizes (n=20,40,80,160,320,640). Experiments were performed using PyTorch (*Automatic Differentiation in PyTorch | OpenReview*, n.d.) running on python 3.6.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

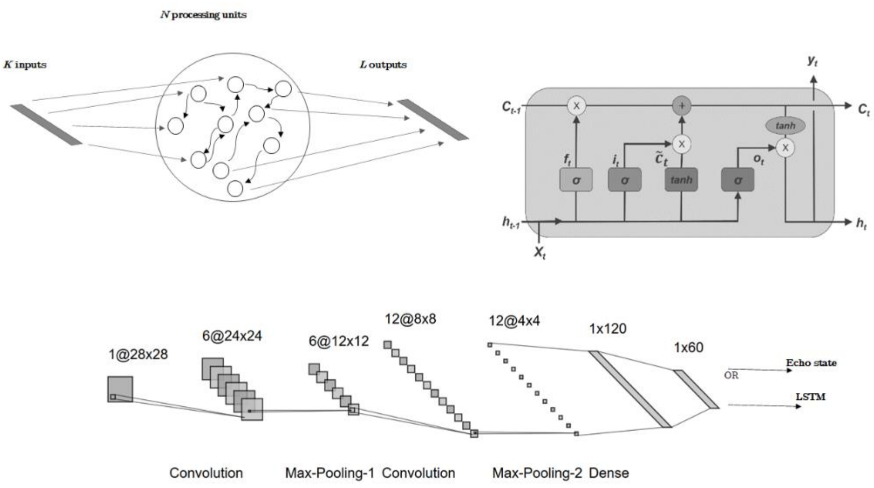


Fig. 4. The architecture of the recurrent networks used in this study. Top right: schematic representation of an LSTM network (from (Zheng et al., 2017), under Creative Commons Attribution License). Top left: Schematic representation of an echo state network. Bottom: The last layer in the classifier was replaced with either an echo state network or an LSTM recurrent network.

3 Results

3.1 Feedforward convolutional networks (without recurrent networks)

We trained the feedforward convolutional network (figure 3.) to either classify the object orientation or object identity in both datasets. After 200 epochs of training with a learning rate of 10^{-4}, we tested the performance of the feedforward networks on the test dataset. Object and width classification accuracy in the synthetic dataset were 56% and 75%, respectively. When trained on COIL-100 dataset, the networks achieved a similar object and orientation classification accuracy of 53% and 68%, respectively. See table 1 and figure 5 for graphical representation and details of the accuracies in each dataset.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Synthetic Objects' Dataset		COIL-100 Dataset	
	Object	Width	Object	Orientation
Accuracy	56%	75%	53%	68%

Table 1. Numerical values of classification accuracies in feedforward networks trained on the synthetic objects and the COIL-100 dataset.

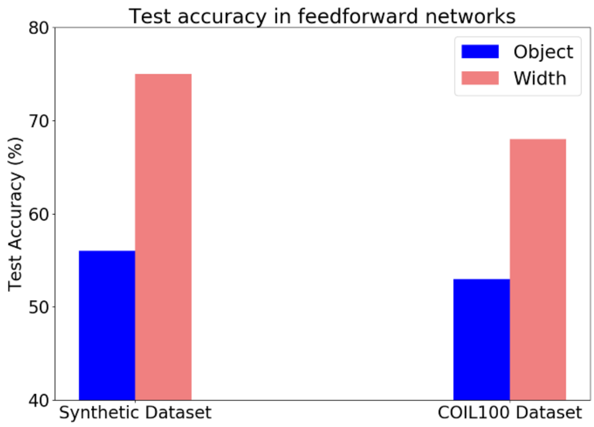


Figure 5. Classification accuracies of the feedforward networks on the synthetic dataset and the COIL-100 dataset. The blue bars show object classification accuracy in the synthetic dataset (left) and the COIL-100 dataset (right). Light red shows accuracies for width classification in synthetic dataset and orientation classification in the COIL-100 dataset.

3.2 Experiment 1.A: Forget rate learning: which task needs more memory?

To see if adding recurrent networks would help the accuracy of the network (compared to purely feedforward networks) and to be able to compare the memory requirements for each task, we replaced the last layer of the feedforward network with an echo state recurrent network (size=200 neurons). The outputs of the echo state network were fed to a new linear layer to classify either objects or orientations (figure 4). Subsequently, we looked at the accuracy and forget rate of 30 networks with different random initializations as they learned to classify objects or orientations. We found that as the accuracy of the viewpoint-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

invariant object classification network increased, the forget rate dropped to lower values compared to the width classification network. These results were consistent regardless of the initial choice of forget rate value (figure 6). Moreover, object classification task required smaller forget rates (longer memory) regardless of the dataset type by showing the same trend in both synthetic objects and COIL-100 dataset. Adding recurrence to the classifiers benefited object classification more than orientation classification in the synthetic dataset but not COIL-100 dataset. Notably, when we looked at the networks initialized by a forget rate of 0.25, we observed that in the synthetic objects' dataset, object classification accuracy increased from 56% to 92 (36% improvement) while the width classification increased from 75% to 94% (only improved by 19%). In the COIL-100 dataset, however, the object classification increased from 53% to 85% (32% improvement), and width classification also increased from 68% to 99% (31% improvement).

Overall, results using this modification suggested that object classification needed smaller forget rate values and hence, a longer memory (figure 6).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

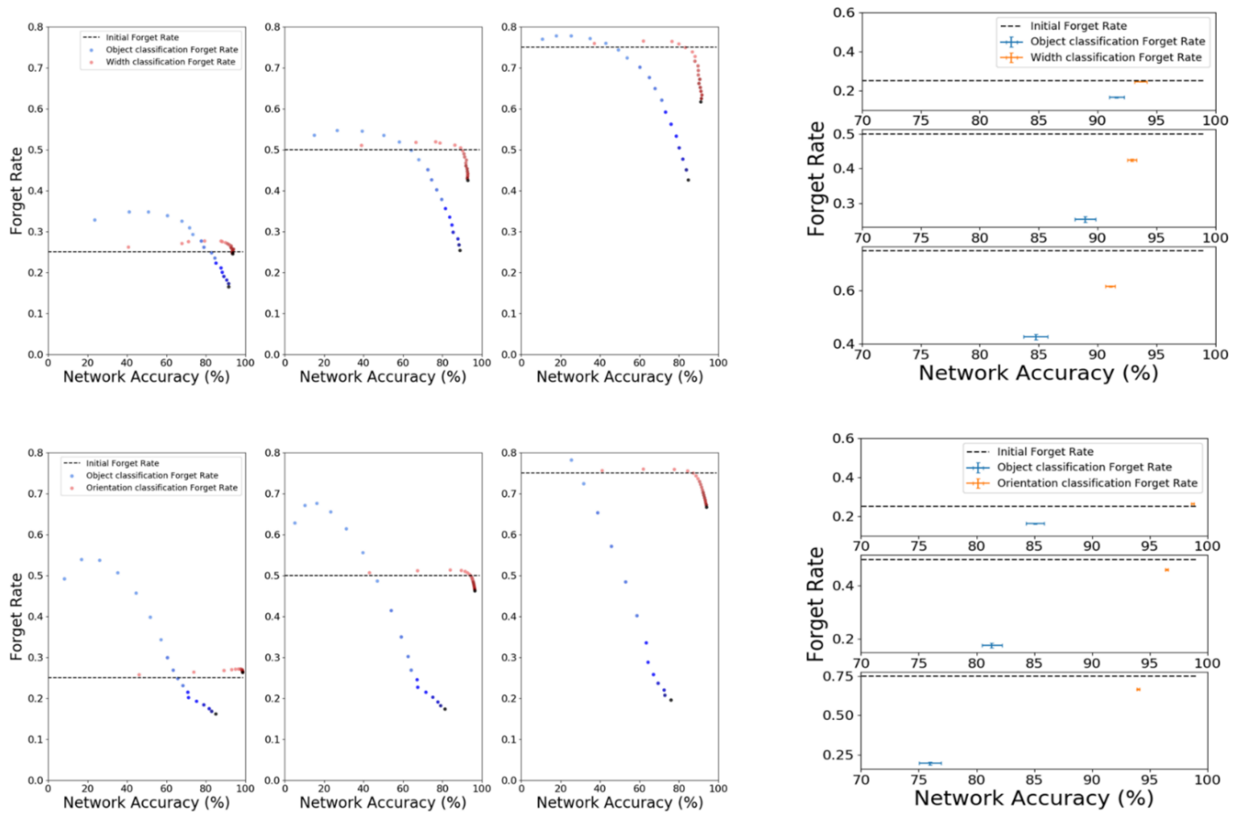


Fig. 6. Simultaneous changes of forget rate and accuracy in two networks as one learns to classify objects and the other learns to classify widths/orientations. Top left: object classification results are shown in blue, and width classification results are shown in red. Darker shades indicate the progression in the learning epochs (from epoch 1 to 18). Dashed lines show initial forget rates. Subpanel on the left: changes in leaky rate and accuracy in a network with initial leaky rate of 0.25. As the network learns the sequence, the accuracy of the network increases in both tasks and the forget rate goes down (longer memory) in both networks. However, the forget rate goes to smaller values (longer memory) in object classification networks. Middle and right subpanels showing the same with different initial values for the forget rate (0.5 and 0.75). Each dot represents the mean forget rate and accuracy value of 30 network realizations. The top plots belong to values obtained from the synthetic dataset. Top right: final forget rate values, the error bars in here, and other figures in this paper are SEMs. The bottom plots show the same results obtained from the COIL100 dataset.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.3 Experiment 1.B: Parameter sweep in ESNs: which task needs more memory?

Building on the results of experiment 1.A, we sought to test if our results (longer memory in object classification) depend on the network size. Additionally, we aimed to perform a parameter sweep for forget rates to see how network accuracies change if we fix the forget rates and only allow output weights change (weights from echo state network to linear classifying layer). To this end, we kept the forget rate constant and let the network train until it reaches a performance plateau after 50 training epochs for each of the 50 different forget rates between 0 and 1 with 0.02 intervals. This simulation repeated 30 times using 30 network realizations. Subsequently, average accuracies of these networks were taken for comparison. The results suggested that overall, parameter sweep does not effectively separate the forget rate requirements in object classification from the forget rate requirements in width classification. Specifically, we observed that in networks trained on the synthetic objects' dataset, both tasks show similar changes in the accuracy as the forget rate changes (figure 7). The results obtained from the COIL-100 dataset were similar and did not show apparent dissociation between the two tasks (supplementary figure 2).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

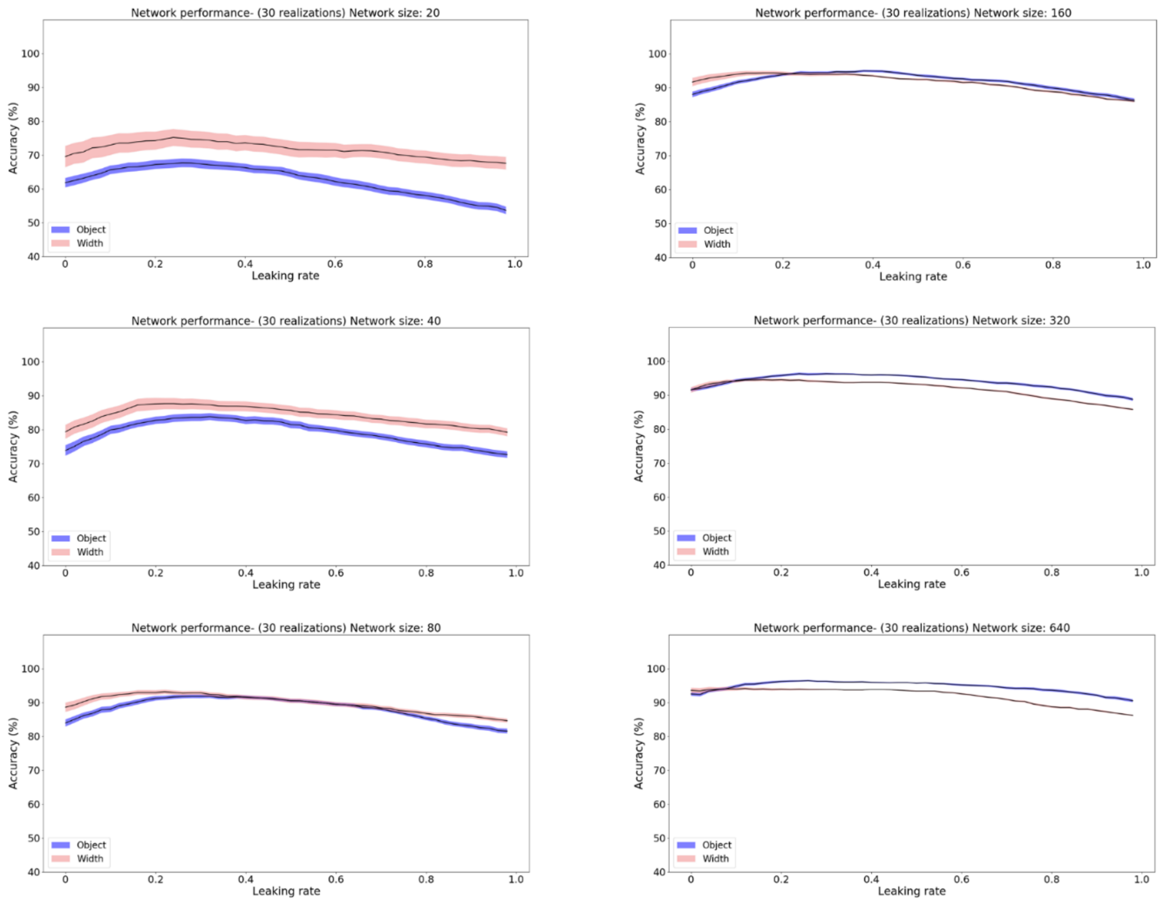


Figure 7. Performance of echo state networks in viewpoint-invariant object classification and width classification tasks with different forget rates. Each subpanel indicates the accuracy of both object classification (blue) and width classification networks (red) with different forget rates (0 to 1, 0.02 intervals) in each network size (20 to 640 neurons). The accuracy of the networks was measured after 50 training epochs so that networks reach a relative performance plateau. Results are the mean accuracy of 30 different network realizations, and shaded areas are standard error of the mean (SEM).

3.4 Experiment 2.A: Which task requires more memory? LSTM networks trained on synthetic objects' dataset

In experiment 2.A, we used LSTM networks instead of echo state networks to measure memory dependence of each task in a different type of recurrent network. Similar to our echo state network experiments, when feedforward

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

networks were trained on each task, their weights were frozen, and their last layer was replaced by an LSTM network and a linear classifier. After training, we fed the network with sequences of test images again and extracted the logarithm of forget gate to input gate ratio as each single test image was passed through the network. We found that the average logarithm of forget to input gate ratio is higher in *width* classification compared to viewpoint invariant object classification. To investigate this effect more, we changed the learning rate from the initial value (10^{-3}) to 10^{-4} to see if the accuracy and logarithm of ratios are sensitive to the learning rate. We observed that even though the logarithm of ratios changes, the overall trend stays the same, and *width* classification has higher input to forget rate ratio (longer memory) compared to object classification. An interesting trend was the reduction of the logarithm of forget to input gate ratio as network size increased which may point to an inverse relationship between network size and memory dependence. See figure 8 for more detail.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

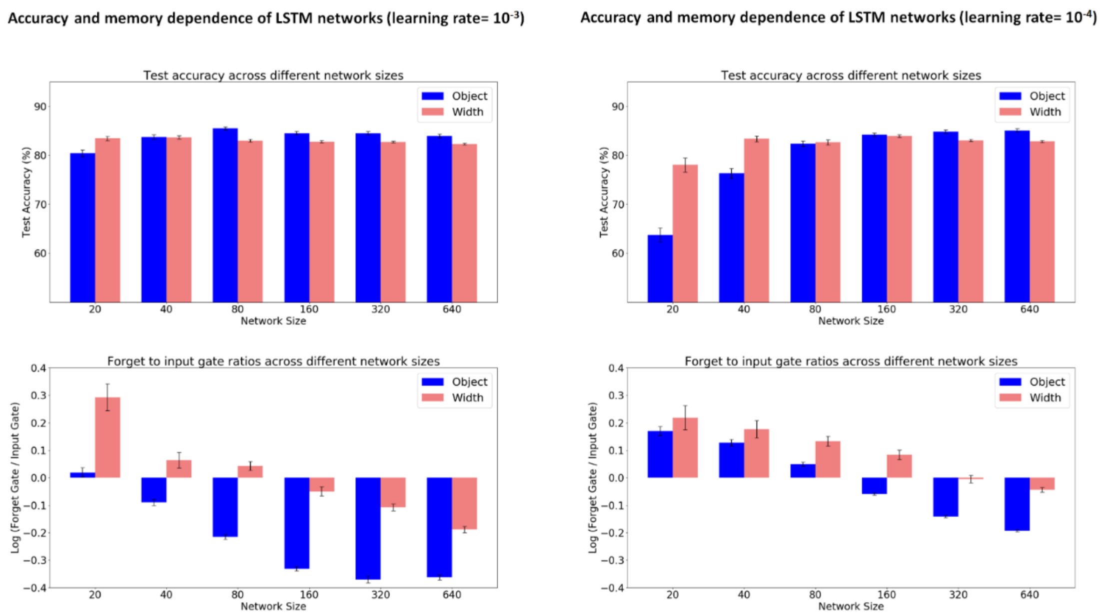


Fig. 8. Difference between the forget to input gate ratios (as measures of memory) in width and object classification tasks obtained from networks with different sizes trained on synthetic objects' dataset. Top left: mean accuracy of networks in each task as the network size changes (20 to 640 neurons). Blue bars are accuracies in object classification tasks, and light red bars are accuracies in width classification tasks (learning speed=10^{-3}), values are averaged across 30 network realizations. Bottom left: Bars show the average $\boldsymbol{log}(\frac{\textbf{forget gate}}{\textbf{input gate}})$ for 30 networks trained on either object classification (blue) or width classification (light red). Error bars show the standard error of the mean (SEM). Top right and bottom right show the same values for networks trained with a learning rate of 10^{-4}.

3.5 Experiments 2.B: Which task requires more memory? LSTM networks trained on COIL-100 dataset.

All parameters in experiment 2.B were the same as experiment 2.A, but we used the COIL-100 dataset instead of our synthetic dataset to train the network. In contrast to the results we obtained from the synthetic dataset, the average logarithm of forget to input gate ratio is higher when the network is performing viewpoint invariant object classification compared to orientation classification.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additionally, the object classification networks kept a higher forget to input gate ratio in all sizes. This trend remained the same regardless of the learning rate. See figure 9 for details. Interestingly, the overall forget to input gate ratios showed a decreasing trend as we increased the network size in here as well. See the discussion section for a possible explanation for this observation.

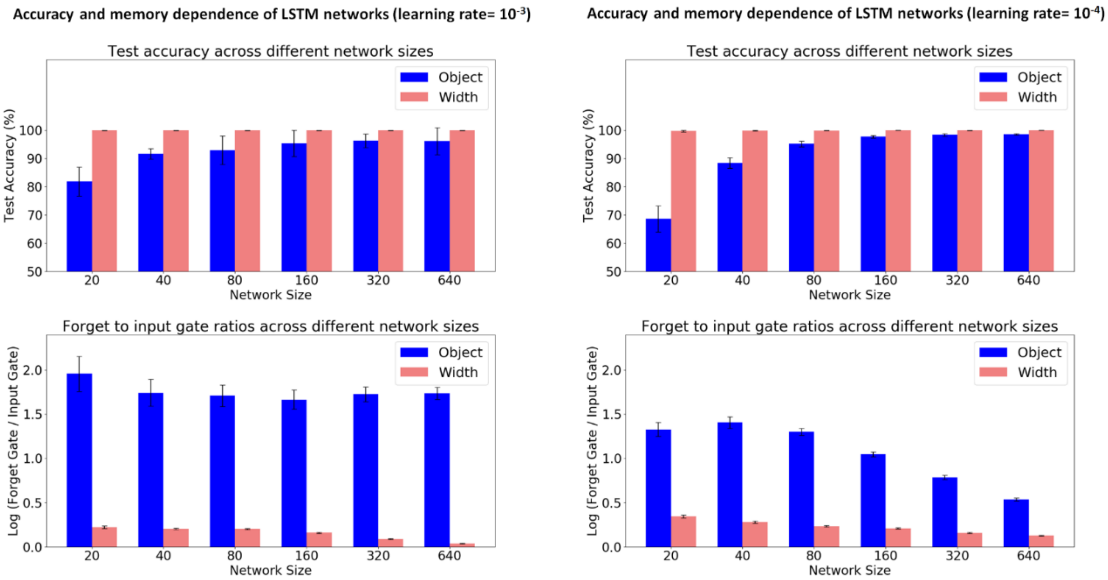


Fig. 9. Difference between the forget to input gate ratios (as measures of memory) in orientation and object classification tasks in networks with different sizes trained on COIL-100 dataset. Top left: mean accuracy of networks in each task as the network size changes (20 to 640 neurons). Blue bars are accuracies in object classification tasks, and light red bars are accuracies in orientation classification tasks (learning speed=10^{-3}); values are averaged across 30 network realizations. Bottom left: Bars show the average $\boldsymbol{log}(\frac{\textbf{forget gate}}{\textbf{input gate}})$ for 30 networks of the same size trained on either object classification (blue) or orientation classification (light red). Error bars show the standard error of the mean (SEM). Top right and bottom right show the same values for networks trained with a learning rate of 10^{-4}.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4 Discussion

In this study, we have shown that in echo state networks, learning to classify objects requires a longer memory compared to width/orientation classification. Additionally, our results indicated that in LSTM networks, object classification requires a longer memory only when the network is trained on larger datasets (3-4 thousand images) and not in small datasets (200-400 images).

One natural first response to these results would be to think that differences in memory lengths in these tasks are probably caused by the length of image sequences in each class of the dataset, i.e., object classification networks might have longer memory simply because they have been learning longer sequences of images. However, we deliberately used longer sequences in width/orientation classes of both datasets. Notably, there were 32 images in each width class of the synthetic objects' dataset (compared to 22 in the object classes). In the COIL-100 dataset, there were 756 images in each orientation class compared to 72 in each object class.

The second concern with the results is the fact that the total number of images in each task is different. Therefore, the reliance on memory reflects the amount of information that has to be learned in each task. Specifically, in the synthetic objects' dataset, there were 192 images in the orientation classification task, while there were 352 images in the viewpoint invariant object classification task. Similarly, in COIL-100 dataset, there were 3,888 images in the object classification task, while there were only 3,024 images in the orientation classification task. However, if this were the case, we would not see the width classification networks trained on synthetic objects' dataset to have higher memory reliance compared to object classification networks (192 images compared to 352 images, see figure 7 for LSTM results). Moreover, if the number of images in each task was a primary factor in determining accuracy, orientation classification in networks trained on the synthetic objects' dataset would become significantly better than the orientation classification of networks trained on the COIL-100 dataset (3,024 compared to 194 images). However, the orientation classification accuracies in both echo state networks and LSTMs trained on either

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the datasets are similar (figure 6 showing echo state networks, and figure 8 and 9 showing LSTMs).

One interesting effect that we observed in this study was the relationship between memory and size of the network. Both echo state networks and LSTMs indicated an inverse relationship between network size and their need for larger memory spans (figure 7, 8, and 9). While the relationship between network architecture and memory capacity in echo state networks has been explored elsewhere (Gallicchio et al., 2018), the exact relationship between network size and memory capacity deserves further investigation.

From a broader perspective, our results were in line with clinical findings in patients and the two-streams hypothesis for primate vision. Meanwhile, It is noteworthy to mention that some studies called initial assumptions of the two-streams hypothesis into question (Hesse & Schenk, 2014; Konen & Kastner, 2008; Rogers et al., 2009), (see (Schenk & Hesse, 2018) for a critical review). Thus, our results cannot generalize to the entire dorsal pathway. In particular, it is essential to note that according to Kravitz et al., the dorsal stream itself is composed of three different sub-pathways: the parieto-prefrontal, the parieto-medial temporal, and the parieto-premotor paths (Kravitz et al., 2011). Specifically, the parieto-prefrontal path is involved in spatial working memory and eye movement control, the parieto-medial temporal pathway is critical for spatial navigation and spatial long term memory, and the parieto-premotor pathway is heavily involved in visually controlled movements such as reaching and grasping (Kravitz et al., 2011). Due to the nature of the tasks we used here, our results are most relevant to the parieto-premotor pathway.

The short-term nature of memory in the dorsal pathway has been extensively studied using both psychophysical and imaging studies and generated mixed results. For example, Cant et al. showed that while naming of objects can be primed, a grasping movement cannot be primed by previous grasping movements (Cant et al., 2005), supporting the short-term nature of the visuomotor control representations. Similarly, Jax and Rosenbaum reported that while priming in a visually guided obstacle avoidance task with short delays between the priming stimulus and actual task was possible, this effect went away with delays that

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were longer than a second (Jax & Rosenbaum, 2007). Meanwhile, the idea that in delayed motor-controlled tasks, the source of information switches from dorsal to ventral pathway was challenged by some studies (Schenk & Hesse, 2018). For example in a study by Himmlbach et al., a bilateral optic ataxia patient (IG) showed strong activity in regions around his lesion in the dorsal pathway in both immediate and delayed reaching tasks (Himmelbach et al., 2009). Moreover, the famous patient D.F. showed that she could perform delayed visually guided movement as good as controls when the environmental cues (allocentric information) were *not* available (Hesse & Schenk, 2014). This study lent credit to the idea that the dorsal pathway can still keep information related to visually guided behavior for long delays (>2seconds), and it is the contextual information that becomes available after a delay.

The short memory span of the dorsal stream could be understood from the perspective of its inputs as well. Magnocellular inputs mostly innervate the dorsal pathway whereas the ventral pathway is more innervated by parvocellular inputs (Merigan & Maunsell, 1993). The main difference between these two types of inputs is that the magnocellular cells are better at the classification of higher temporal frequencies, whereas parvocellular cells are more suitable for higher spatial frequencies. Additionally, magnocellular cells are 20 milliseconds faster in terms of their response latency to stimuli (Bullier & Nowak, 1995). The faster dynamics of the dorsal pathway is in line with what we found in our study.

Since similar two-stream dissociation is suggested in other sensory modalities such as somatosensation (Dijkerman & de Haan, 2007; James & Kim, 2010) or audition (Hickok & Poeppel, 2007; Rauschecker, 2018), the relationship between short term memory and motor control tasks might even go beyond vision. However, our current datasets and tasks are limited to vision, and specific tasks for each modality would be required to find out if such a relationship holds for other sensory modalities as well.

Another issue is the fact that the majority of the computational models of the visual cortex are based on the ventral stream (e.g., (Lotter et al., 2016, 2020)), while models incorporating both ventral and dorsal pathways are less common (but see (O'Reilly et al., 2017, 2020) as an example incorporating both streams). We believe that newer models of the visual cortex should incorporate

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the tasks relevant to both streams. This will demonstrate if different tasks (e.g. object recognition vs. motor control) require incompatible computational paradigms that need different circuitries.

5 Limitations

Our current study comes with several shortcomings. First, we could have added a motor control task to our width classification task to be able to make a direct comparison between our model and parietal-premotor pathway.

Second, since the dorsal pathway receives information from subcortical sources such as superior colliculus and these sources are mostly innervated by magnocellular cells with higher sensitivity to lower spatial frequencies, we could have used larger sized kernels for convolutional layers that we used to approximate dorsal pathway. However, doing so would render comparison of recurrent networks a bit less straight forward.

6 Conclusions

In the present study, we have shown that there is a close relationship between length of memory in recurrent networks and how they perform in object classification or width/orientation classification tasks. While having a longer memory span benefits object classification performance (as a ventral stream task) in echo state networks, in LSTM networks, such effect is present only when the network is trained on larger dataset.

7 Acknowledgment

This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU was also supported in part by Lilly Endowment, Inc. The authors acknowledge the Indiana University Pervasive Technology Institute for providing Carbonate HPC

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

resources that have contributed to the research results reported within this paper (Stewart et al., 2017).

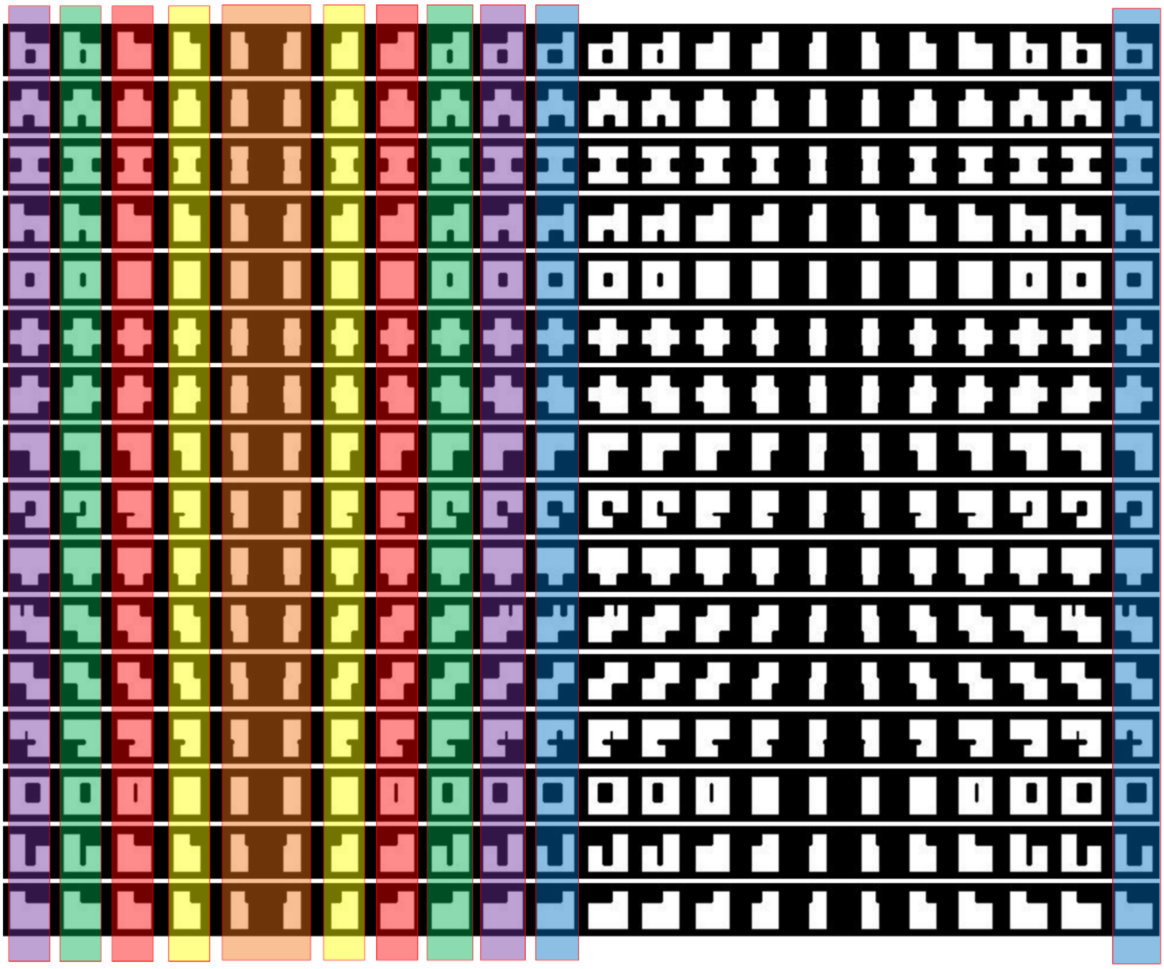
8 Supplementary Material

8.1 Code

The code used in this study is available at (https://github.com/Abolfazl-Alipour).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8.2 Supplementary figures



Width class No. 1: All objects with an orientation of 0° and 180°

Width class No. 2: All objects with an orientation of 16° and 164°

Width class No. 3: All objects with an orientation of 33° and 147°

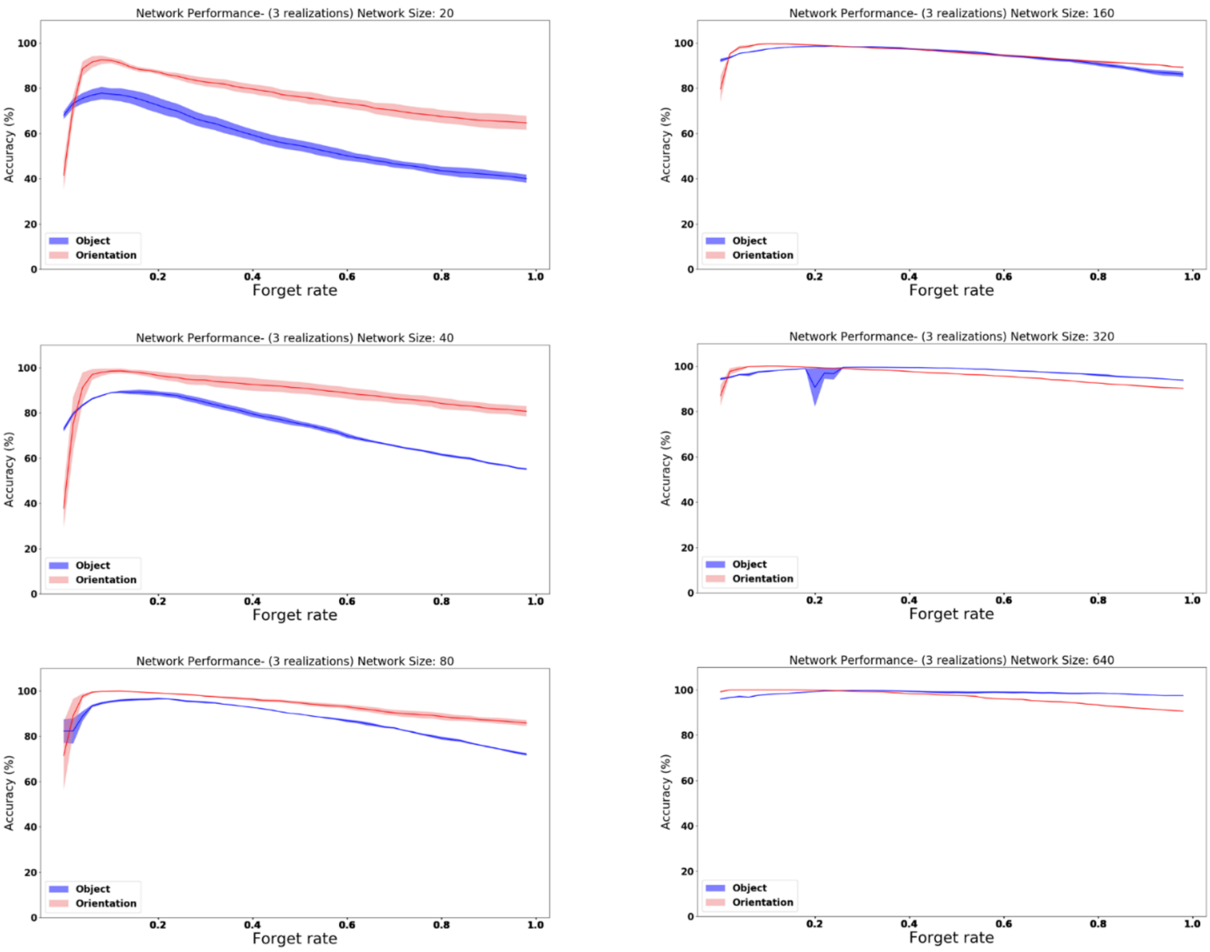
Width class No. 4: All objects with an orientation of 49° and 131°

Width class No. 5: All objects with an orientation of 65° and 115°

Width class No. 6: All objects with an orientation of 82° and 98°

Supplementary Fig. 1. Width classes derived from the synthetic dataset. Each with class is shown by a different color.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Fig. 2. Performance of echo state networks in viewpoint-invariant object classification and width classification tasks trained on COIL-100 dataset. Each subpanel indicates the accuracy of both object classification (blue), and orientation classification networks (red) with different forget rates (0 to 1, 0.02 intervals) in each network size (20 to 640 neurons). The accuracy of the networks was measured after 50 training epochs so that networks reach a relative performance plateau. Results are the mean accuracy of 3 different network realizations, and shaded areas are standard error of the mean (SEM).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9 References

Automatic differentiation in PyTorch | OpenReview. (n.d.). Retrieved December 14, 2019, from https://openreview.net/forum?id=BJJsrmfCZ

Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: New vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*(4), 497–503. https://doi.org/10.1016/0959-4388(95)80011-5

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Computational Biology*, *10*(12), e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Cant, J. S., Westwood, D. A., Valyear, K. F., & Goodale, M. A. (2005). No evidence for visuomotor priming in a visually guided action task. *Neuropsychologia*, *43*(2), 216–226. https://doi.org/10.1016/j.neuropsychologia.2004.11.008

Dijkerman, H. C., & de Haan, E. H. F. (2007). Somatosensory processes subserving perception and action. *Behavioral and Brain Sciences*, *30*(2), 189–201. https://doi.org/10.1017/S0140525X07001392

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ettlinger, G. (1990). “Object Vision” and “Spatial Vision”: The Neuropsychological Evidence for the Distinction. *Cortex*, *26*(3), 319–341. https://doi.org/10.1016/S0010-9452(13)80084-6

Gallicchio, C., Micheli, A., & Pedrelli, L. (2017). Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, *268*, 87–99. https://doi.org/10.1016/j.neucom.2016.12.089

Gallicchio, C., Micheli, A., & Silvestri, L. (2018). Local Lyapunov exponents of deep echo state networks. *Neurocomputing*, *298*, 34–45. https://doi.org/10.1016/j.neucom.2017.11.073

Goodale, M. A., Jakobson, L. S., & Keillor, J. M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia*, *32*(10), 1159–1178. https://doi.org/10.1016/0028-3932(94)90100-7

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*(6305), 154–156. https://doi.org/10.1038/349154a0

Goodale, Melvyn A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Hesse, C., & Schenk, T. (2014). Delayed action does not always require the ventral stream: A study on a patient with visual form agnosia. *Cortex*, *54*, 77–91. https://doi.org/10.1016/j.cortex.2014.02.011

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Himmelbach, M., Nau, M., Zündorf, I., Erb, M., Perenin, M.-T., & Karnath, H.-O. (2009). Brain activation during immediate and delayed reaching in optic ataxia. *Neuropsychologia*, *47*(6), 1508–1517. https://doi.org/10.1016/j.neuropsychologia.2009.01.033

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, Y., Eagleson, R., & Goodale, M. A. (1999). The effects of delay on the kinematics of grasping. *Experimental Brain Research*, *126*(1), 109–116. https://doi.org/10.1007/s002210050720

Hu, Y., & Goodale, M. A. (2000). Grasping after a Delay Shifts Size-Scaling from Absolute to Relative Metrics. *Journal of Cognitive Neuroscience*, *12*(5), 856–868. https://doi.org/10.1162/089892900562462

Jaeger, H. (n.d.). *The "echo state" approach to analysing and training recurrent neural networks – with an Erratum note*. 48.

James, T. W., & Kim, S. (2010). Dorsal and Ventral Cortical Pathways for Visuo-haptic Shape Integration Revealed Using fMRI. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory Object Perception in the Primate*

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brain (pp. 231–250). Springer. https://doi.org/10.1007/978-1-4419-5615-6_13

Jax, S. A., & Rosenbaum, D. A. (2007). Hand path priming in manual obstacle avoidance: Evidence that the dorsal stream does not only control visually guided actions in real time. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 425–441. https://doi.org/10.1037/0096-1523.33.2.425

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*(2), 224–231. https://doi.org/10.1038/nn2036

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008

Lotter, W., Kreiman, G., & Cox, D. (2016). *Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning*. https://arxiv.org/abs/1605.08104v5

Lotter, W., Kreiman, G., & Cox, D. (2020). A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature Machine Intelligence*, *2*(4), 210–219. https://doi.org/10.1038/s42256-020-0170-9

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Merigan, W. H., & Maunsell, J. H. R. (1993). How Parallel are the Primate Visual Pathways? *Annual Review of Neuroscience*, *16*(1), 369–402. https://doi.org/10.1146/annurev.ne.16.030193.002101

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. https://doi.org/10.1016/j.neuropsychologia.2007.10.005

Milner, A. D., Paulignan, Y., Dijkerman, H. C., Michel, F., & Jeannerod, M. (1999). A paradoxical improvement of misreaching in optic ataxia: New evidence for two separate neural systems for visual localization. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1434), 2225–2229. https://doi.org/10.1098/rspb.1999.0912

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., Bettucci, D., Mortara, F., Mutani, R., Terazzi, E., & Davidson, D. L. W. (1991). PERCEPTION AND ACTION IN 'VISUAL FORM AGNOSIA.' *Brain*, *114*(1), 405–428. https://doi.org/10.1093/brain/114.1.405

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*(1), 57–77. https://doi.org/10.1016/0166-4328(82)90081-X

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nene, S. A., Nayar, S. K., & Murase, H. (1996). *Object image library (COIL-100*.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*(1), 73–96. https://doi.org/10.1017/S0140525X0200002X

O'Reilly, R. C., Russin, J. L., Zolfaghar, M., & Rohrlich, J. (2020). Deep Predictive Learning in Neocortex and Pulvinar. *ArXiv:2006.14800 [q-Bio]*. http://arxiv.org/abs/2006.14800

O'Reilly, R. C., Wyatte, D. R., & Rohrlich, J. (2017). Deep Predictive Learning: A Comprehensive Model of Three Visual Streams. *ArXiv:1709.04654 [q-Bio]*. http://arxiv.org/abs/1709.04654

Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*, *98*, 262–268. https://doi.org/10.1016/j.cortex.2017.10.020

Rogers, G., Smith, D., & Schenk, T. (2009). Immediate and delayed actions share a common visuomotor transformation mechanism: A prism adaptation study. *Neuropsychologia*, *47*(6), 1546–1552. https://doi.org/10.1016/j.neuropsychologia.2008.12.022

Schaetti, N., Salomon, M., & Couturier, R. (2016). Echo State Networks-Based Reservoir Computing for MNIST Handwritten Digits

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Recognition. *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*, 484–491. https://doi.org/10.1109/CSE-EUC-DCABES.2016.229

Schenk, T., & Hesse, C. (2018). Do we have distinct systems for immediate and delayed actions? A selective review on the role of visual memory in action. *Cortex*, *98*, 228–248. https://doi.org/10.1016/j.cortex.2017.05.014

Schneider, G. E. (1969). Two Visual Systems. *Science*, *163*(3870), 895–902. https://doi.org/10.1126/science.163.3870.895

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Schmidt, K., Yamins, D. L. K., & DiCarlo, J. J. (2018). Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? *BioRxiv*, 407007. https://doi.org/10.1101/407007

Stewart, C. A., Welch, V., Plale, B., Fox, G., Pierce, M., & Sterling, T. (2017). *Indiana University Pervasive Technology Institute. Bloomington, Indiana*.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A Survey on Deep Transfer Learning. In V. Kůrková, Y. Manolopoulos, B.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hammer, L. Iliadis, & I. Maglogiannis (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2018* (pp. 270–279). Springer International Publishing. https://doi.org/10.1007/978-3-030-01424-7_27

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*(4), 299–337. https://doi.org/10.1007/BF00422717

Yamins, D. L., Hong, H., Cadieu, C., & DiCarlo, J. J. (2013). Hierarchical Modular Optimization of Convolutional Networks Achieves Representations Similar to Macaque IT and Human Ventral Stream. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 3093–3101). Curran Associates, Inc. http://papers.nips.cc/paper/4991-hierarchical-modular-optimization-of-convolutional-networks-achieves-representations-similar-to-macaque-it-and-human-ventral-stream.pdf

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244

Zheng, H., Yuan, J., & Chen, L. (2017). Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.30.321299; this version posted September 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Importance Evaluation. *Energies*, *10*(8), 1168. https://doi.org/10.3390/en10081168