Recurrent neural networks that learn multi-step visual routines with reinforcement learning

Sami Mollard[1], Catherine Wacongne[1,2], Sander M. Bohte[3,4], Pieter R. Roelfsema[1,5,6,7]

Affiliations:
[1]Department of Vision & Cognition, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA, Amsterdam, The Netherlands.

[2]AnotherBrain, Paris, France.

[3]Machine Learning Group, Centrum Wiskunde & Informatica, Science Park 123, 1098XG, Amsterdam, The Netherlands

[4]Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1090GE Amsterdam, Netherlands

[5]Laboratory of Visual Brain Therapy, Sorbonne Université, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Institut de la Vision, Paris F-75012, France

[6]Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, VU University, Amsterdam, The Netherlands.

[7]Department of Neurosurgery, Academic Medical Center, Amsterdam, The Netherlands

Abstract

Many cognitive problems can be decomposed into series of subproblems that are solved sequentially by the brain. When subproblems are solved, relevant intermediate results need to be stored by neurons and propagated to the next subproblem, until the overarching goal has been completed. We will here consider visual tasks, which can be decomposed into sequences of elemental visual operations. Experimental evidence suggests that intermediate results of the elemental operations are stored in working memory as an enhancement of neural activity in the visual cortex. The focus of enhanced activity is then available for subsequent operations to act upon. The main question at stake is how the elemental operations and their sequencing can emerge in neural networks that are trained with only rewards, in a reinforcement learning setting.

We here propose a new recurrent neural network architecture that can learn composite visual tasks that require the application of successive elemental operations. Specifically, we selected three tasks for which electrophysiological recordings of monkeys' visual cortex are available. To train the networks, we used RELEARNN, a biologically plausible four-factor Hebbian learning rule, which is local both in time and space. We report that networks learn elemental operations, such as contour grouping and visual search, and to execute sequences of operations, solely based on the characteristics of the visual stimuli and the reward structure of a task.

After training was completed, the activity of the units of the neural network resembled the activity of neurons in the visual cortex of monkeys solving the same tasks. Relevant information that needed to be exchanged between subroutines was maintained as a focus of enhanced activity and passed on to the subsequent subroutines. Our results demonstrate how a biologically plausible learning rule can train a recurrent neural network on multistep visual tasks.

Introduction

Many everyday tasks can be decomposed into subroutines that are executed sequentially. If you want to find your way on a map, for instance, you first search where your goal and actual position are, and then you find path between those two positions. Ullman (1984) proposed that such complex visual tasks can be solved using visual routines, neuronal programs consisting of a sequence of atomic mental steps that he called elemental operations. Curve tracing, visual search, cuing, matching or contour grouping are examples of such elemental operations (Roelfsema, 2005; Roelfsema et al., 2000; Zylberberg et al., 2011).

There are at least two critical conditions for the successful implementation of visual routines: (1) the correct elemental operations must be chosen and applied in the appropriate order, and information must flow between subroutines so that the output of one subroutine can be used as the input for the next one. Those two conditions are straightforward to implement in a computer program, but it remains unclear how variables are bound to subroutines in the brain. Several studies provided insights into the neural basis of elemental operations and visual routines. For example, Jolicoeur et al. (1986) focused on curve tracing as an example elemental operation. In their task, subjects had to determine if two dots were on the same or different curves. They found that curve-tracing is a serial mental operation because the reaction time increases when subjects trace longer curves (see Fig. 1A for an example curve-tracing stimulus).

Previous studies started to gain insight in how elemental operations are implemented in the visual cortex. The neural activity during elemental operations can be divided in several phases (Lamme & Roelfsema, 2000; Roelfsema et al., 2007). The first phase is dominated by feedforward processing. Basic features are registered in low level visual areas and more

complex features, such as object category are registered in higher areas (Roelfsema, 2006; Vanrullen, 2007). Later phases are dominated by recurrent processing, enabled by feedback and horizontal connections, allowing the early visual cortex to act as a cognitive blackboard (Roelfsema & de Lange, 2016). Higher areas may write to the blackboard in early visual cortex using feedback connections, to enhance the representation of task-relevant features and suppress the representation of distracting information. Other higher visual areas can then read from the blackboard utilizing feedforward connections to selectively process features that are task relevant.

Studies in monkeys carrying out attention-demanding tasks have supported the blackboard analogy. For example, Roelfsema et al. (1998) trained monkeys in a curve-tracing task while recording the activity of neurons in the primary visual cortex (area V1). The animals were trained to make an eye movement toward the end of a target curve that was connected to the fixation point, while ignoring another distractor curve (Fig. 1A). The researchers found that neurons with a receptive field (RF) on the target curve enhanced their activity compared to neurons whose RF fell on distractor curves. Later studies demonstrated that the enhanced activity spreads gradually over the V1 representation of the curve (Pooresmaeili & Roelfsema, 2014). The latency of the response enhancement was short at the start of the curve, and it occurred later for neurons with RFs farther along the curve. These findings demonstrate the serial nature of the contour-grouping process. Related findings have been reported for visual search, during which the representation of items that are searched for is enhanced during a delayed phase of the neuronal responses in the monkey visual cortex (Chelazzi et al., 1998, 2001; Desimone & Duncan, 2003; Mirpour & Bisley, 2012; Sato et al., 2001; Zhou & Desimone, 2011) .

To examine how elemental operations can be combined into visual routines, Moro et al. (2010) and Roelfsema et al. (2003) recorded from neurons in the visual cortex of monkeys trained to perform more complex visual tasks. The first task was a ‘search-then-trace’ task (Fig. 1B). In this task, the fixation point could be one of two colors. The monkey had to register the color of the fixation point and to search for a circular marker with the same color, which demarcated the start of the target curve. The monkey then traced the target curve to its other end and planned an eye movement to a circle at that position (blue dot in Fig. 1B). The search and trace operations were performed mentally, while the monkey kept its gaze on the fixation point. The eye movement was made only at the end of the trial and the monkey received a reward when he made an eye movement to the circle at the end of the target curve. The second composite task was a trace-then-search task (Fig. 1C). Now the order of the two subroutines was reversed: the monkey first had to trace the target curve connected to the fixation point and register the color of a marker at the end of this curve. The monkey then then searched for a larger circle with the same color that was the target for an eye movement.

The neurophysiological studies demonstrated that it is possible to monitor the order and progress of the successive subroutines that are executed by the monkey by recording from area V1. The findings also illustrate how information can be transferred from one elemental operation to the next. For example, in the search-then-trace task, the initial visual search enhanced the representation of the target marker as the beginning of the target curve (at a latency of 159ms, light blue circle in Fig. 1B) so that the tracing operation could spread activity along this curve, beginning at this marker (after 229ms, light yellow in Fig. 1B). In the trace-then-search task, the trace operation enhanced the responses elicited by the target marker (at 180ms, Fig. 1C) so that its color could be registered as the target color for the subsequent search. The visual search caused an enhanced response at the circular eye movement target at a latency of 267ms (Fig. 1C).

An important finding is that the enhancement of activity in the visual cortex makes the outcome of an elemental operation available as a focus of enhanced activity that may correspond to the working memory of that feature so that it can be used as the input for the

next elemental operation (Van Kerkoerle et al., 2017). These findings also illustrate the role of early visual cortex as a cognitive blackboard (Roelfsema & de Lange, 2016). Reading and writing to the cognitive blackboard in early visual areas permits the exchange of information between elemental operations that are part of a visual routine. For example, in the search-then-trace task, the visual search operation tags the location of the target marker with enhanced activity in area V1, which corresponds to a write operation on the cognitive blackboard. This focus of enhanced activity marks the beginning of the tracing operation, starting at this position.

At a psychological level of description, the neuronal response modulations that occur during the recurrent processing phase correspond to shifts of attention, which is directed to image elements that are represented with enhanced neuronal activity (Roelfsema & Houtkamp, 2011). Indeed, the sequence of mental operations is associated with a sequence of attention shifts. Visual search for a target shape or color in an array of stimuli leads to a shift of attention to the item that needs to be found (Kim & Cave, 1995). Similarly, when subjects mentally trace a curve, they gradually spread attention across all of its contour elements (Houtkamp et al., 2003).

Hence, there is substantial evidence that visual routines are implemented in the visual cortex through the propagation of enhanced activity during a recurrent processing phase. However, it remains unknown what architectures and plasticity rules permit the learning of such routines in the brain. To investigate how curve-tracing can be learned, Brosch et al (2015) developed RELEARNN. RELEARNN is a biologically plausible learning rule that uses two factors to gate Hebbian plasticity. The first factor is a reward-prediction error that is central in many approaches to biologically inspired learning (Sutton & Barto, 2018) and that is thought to be made available throughout the brain by the release of neuromodulators such as dopamine (Schultz, 2002; Schultz et al., 1997). The second factor used by RELEARNN is a feedback signal initiated by the selected action, which propagates through a dedicated set of recurrent connections in an “accessory network” that highlights the synapses that have had most influence on this action and that gates the plasticity. The combination of these two factors gives rise to a learning rule that is equivalent to the Almeida-Pineda algorithm that can be used to train recurrent networks (Almeida, 1987; Pineda, 1987), while all the information needed to control plasticity is available locally at every synapse. Importantly, RELEARNN is based on trial-and-error learning and does not require the “teacher” of error-backpropagation, because the only feedback that the network receives is whether it obtains a reward.

Interestingly, Brosch et al. (2015) found that the strategy that the network learned to trace curves resembled the neurophysiological results in the visual cortex of monkeys. The network learned to spread enhanced neuronal activity over the target curve. However, Brosch et al. (2015) only tested curve-tracing and it remains unclear if neural networks can learn to stitch elemental visual operations into visual routines using a biologically plausible plasticity rule that is based on trial-and-error learning. In the present study we therefore set out to examine learning of visual routines by neural networks with feedforward, feedback and horizontal connections. We asked whether the network (1) would learn multiple elemental operations, (2) whether it would learn to execute the elemental operations sequentially and in the correct order and (3) how the network ensured the transfer of information from one elemental operations to the next.

We used RELEARNN (Brosch et al., 2015) to train the networks on three visual routines for which neurophysiological data in monkeys exist: a curve-tracing task, a search-then-trace and a trace-then-search task (Moro et al., 2010). We found that the networks learned all three tasks. Interestingly, the activation of units of the networks resembled to spatiotemporal patterns of activity observed in the visual cortex of monkeys. The networks learned to propagate enhanced activity along curves and to enhance activity in retinotopic areas to transfer information between subroutines. Our results show how recurrent neural

networks trained with a biologically plausible learning rule can learn to propagate enhanced activity to solve complex visual tasks.

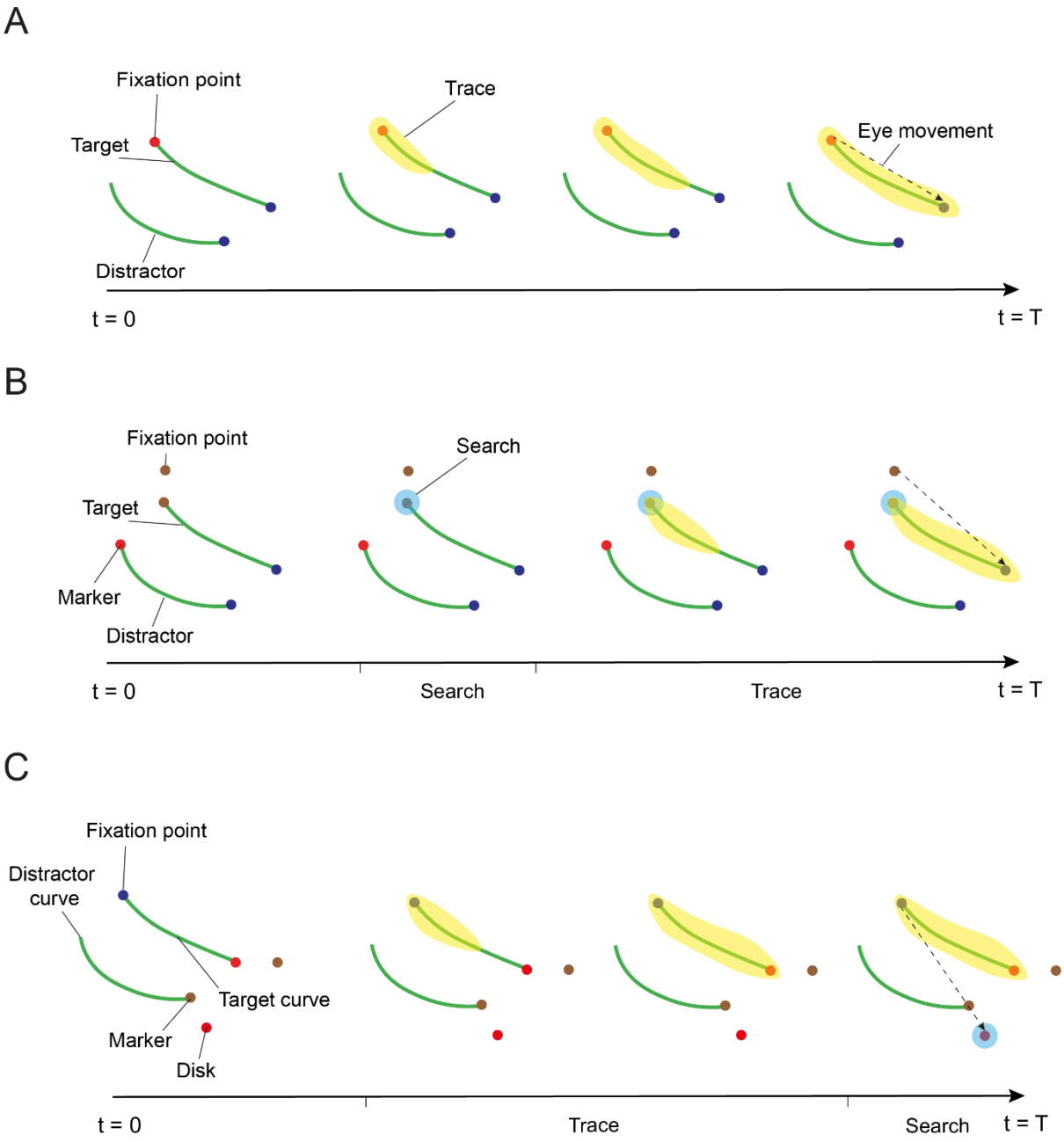

Fig. 1. Example stimuli for the three tasks. A, Trace task. The monkey makes an eye movement toward the blue dot connected to the red fixation point. The representation of the target curve in the visual cortex is enhanced because extra neuronal activity spreads over this curve (yellow). **B,** Search-then-trace task. The monkey searches for a marker with one of two colors and then traces the curve that starts at this marker to its other end to make an eye movement to the blue dot. This task requires visual search followed by curve-tracing. In the visual cortex, the search operation first labels the target marker with enhanced activity (light blue circle). Its position can be used as the starting position of the tracing operation which propagates enhanced activity over the target curve (light yellow)**. C,** Trace-then-search task. The monkey first traces the target curve connected to the fixation point and identifies the color at the end of this curve. It then has to search for a disk with the same color, which is the target for an eye movement. In the visual cortex, enhanced activity first propagates across the target curve and identifies the target color (trace operation, light yellow), which is used during the subsequent search (light blue circle).

The tasks for the model

We trained networks on three tasks. The first task is a version of the curve tracing task (Fig. 2A). This task requires only one of the elemental operations, which is curve tracing (Ullman, 1984). To solve it, the model had to learn to group a set of connected pixels on a grid of 15 by 15 pixels that started with a red pixel. The model had to locate the end of the curve that was composed of green pixels, but the last pixel was blue. The model had to select the blue pixel at the end of the target curve for an eye movement.

The second task was the search-then-trace task (Fig. 2B). We cued the color that was the target of search by activating one of two additional input units, which coded for brown and red (Fig. 2). The model had to locate the pixel with the same color in the visual stimulus and trace the curve starting at this pixel to its other end to locate a blue pixel that was the target for an eye movement.

The third task was the trace-then-search task (Fig. 2C). Now the target curve started with a blue pixel. The model had to identify the color at the other end of the target curve and to select the same color at a location outside the pixel frame. Specifically, it could choose between a red and brown pixel.

In the previous experiments with the monkeys, the trial started when the animal directed its gaze to the fixation point. However, in our modeling work, we started each trials by presenting the full stimulus, irrespective of the model's gaze position.

The search-then-trace and the trace-then-search tasks can be solved by visual routines, composed of visual search and curve tracing as elemental operations, but in different orders. These tasks also illustrate how the output of one operation needs to be carried over as input for the next operation. For example, the output of the search operation of the search-then-trace task is the location of the marker with a specific color, which is the starting point for the subsequent tracing operation.

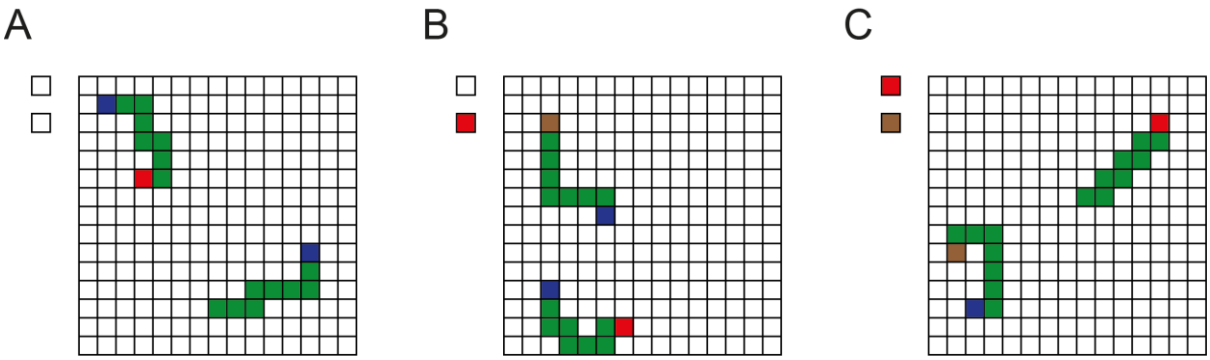

Fig. 2. Example stimuli for the three tasks for the model. A, Trace task. The task is to make an eye movement to the blue pixel of the curve starting with a red pixel. **B,** Search-then-trace task. The model searches for a target marker (here red, as cued on the left of the array) and has to make an eye movement to the blue pixel at the other end of the curve starting with this marker. **C,** Trace-then-search task. The model traces the curve starting with the blue pixel to identify a colored marker at the other end (here brown), which cued the target color that the model should select at the left of the grid. Each trial started with the full stimulus in view.

Model

We will first describe the network architecture and RELEARNN, the biologically plausible reinforcement learning rule. We will then describe the network simulations and compare them to neuronal activity in the visual cortex of monkeys that were trained on the same tasks.

Structure of the network

We trained convolutional neural networks with an input layer, two hidden layers and an output layer. The input layer (layer 0) consisted of a grid of 15x15 input units and there were two extra input units outside the grid. There were four units at every grid location, one for each of the possible colors of the pixels (red, brown, blue and green). In each layer there was a group of feedforward units that only propagated information to the next layer and to other feedforward units. These feedforward units also gated the activity of units in another, recurrent group with a RF at the same position (Fig. 3, “horizontal gating”). Units in the recurrent group could propagate information to the next or previous layer and to units with different receptive fields within the same layer. As a result, units in the recurrent group could be modulated by activity outside their RF. The gating by units in the feedforward group with the same RF (Fig. 3) ensured that the response modulation could not spread to recurrent units that are not active because they represent features that are not present in the image.

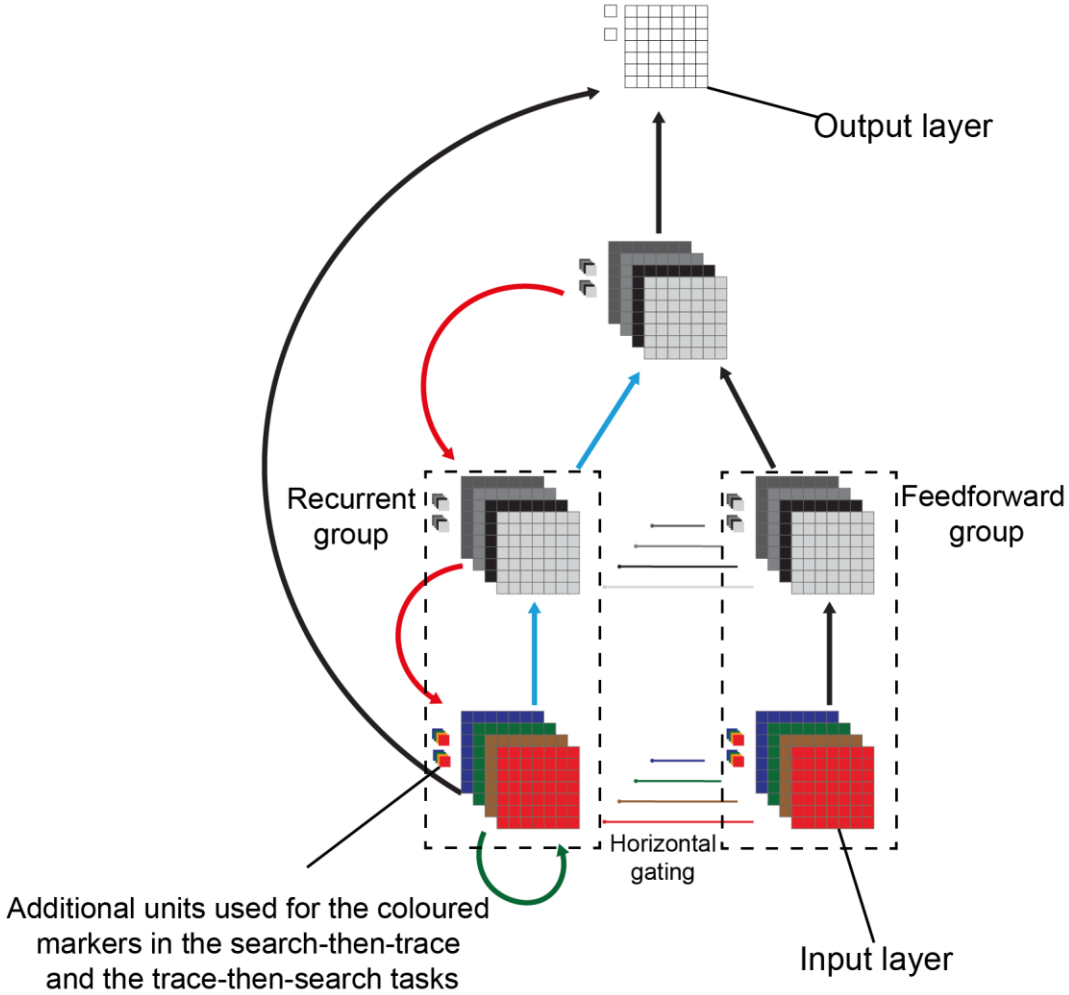

Fig. 3. Architecture of the network. The network comprises one input layer, two hidden layers and one output layer. The input and hidden layers have four features each. In each layer, units belong to the feedforward or to the recurrent group. The activity of units in the recurrent group is gated by input of neurons in the feedforward group with the same RF so that they cannot participate in the spread of enhanced activity in case the corresponding feedforward unit is inactive.

The activities of neurons in the feedforward group in layer *i* at timestep *t* were determined by activities of feedforward neurons in the lower layer at the same timestep through connections

V. If *X* represents the activity of neurons in the feedforward group, *Y* the activity of neurons in the recurrent group, *i* the layer index, then the activity of neurons was determined by:

$$X^l(t) = \sigma\left(V \otimes X^{l-1}(t)\right) \tag{1}$$

$$Y^l(t) = \begin{cases} \sigma\left(\varphi(X^0(t)) \odot (U^1 \otimes Y^1(t-1) + H^0 \otimes Y^0(t-1))\right), & if\ l = 0 \\ \sigma\left(\varphi\left(X^l(t)\right) \odot \left(U^{l+1} \otimes Y^{l+1}(t-1) + T^{l+1} \otimes Y^{l-1}(t)\right)\right), & otherwise \end{cases} \tag{2}$$

Here σ is a nonlinear activation function and φ a gating function (see below), ⊙ represents an element-wise product between the activity of feedback and feedforward units and ⊗ indicates a convolution.

In the recurrent group, neurons in the input layer receive horizontal input from the same layer at the previous timestep through connections *H* (green terms in eq. 2) and units in the higher layers receive feedforward input from the preceding layer at the same timestep through connections *T* (blue terms). The horizontal connections in the input layer carry color information and facilitate the learning of the search component in the search-then-trace and the trace-then-search tasks.

For connections between hidden layers, we used a 3x3 convolutional kernel, and enforced that neurons could only make connections to their Von Neumann neighbourhood (4-neighbourhood). As is common for convolutional neural networks, the kernels were replicated using weight sharing, a property not present in biological neural networks (Pogodin et al., 2021). However, the networks could also learn the tasks without weight sharing. Specifically, without weight sharing the networks learnt the curve-tracing task after an average of 140,000 trials as opposed to 19,000 trials when we used weight sharing. We used weight sharing to reduce computation time but we note that our results are likely to generalize to biologically plausible learning schemes without weight sharing.

The last hidden layer was fully connected to the output layer and the recurrent input layer with weights linking units with the same RF initialized with positive values and between units with different RFs initialized with negative values to bias the winner-takes-all mechanism in the last layer, which stabilized learning.

The activity of units in the retinotopically organized output layer was determined as:

$$Q = W^3 \otimes X^3(T) + W^0 \otimes X^0(T) \tag{3}$$

where *T* is the timestep at which the network reached a stable state.

Units in the output layer were associated with a saccade to one of the pixels of the stimulus grid and they learned to represent the also called Q-value (Sutton & Barto, 2018), which is the expected reward when the networks selected this position for an eye movement.

During training, the model chose the eye movement with the highest Q-value with probability $1 - \epsilon$. With probability ϵ the network explored other actions and sampled a random action *a* from the Boltzmann distribution P_B:

$$P_B(a) = \frac{exp(Q_a)}{\sum_k exp(Q_k)} \tag{4}$$

ϵ was set to 0.05 across tasks and networks.

We trained the networks with RELEARNN (Brosch et al., 2015), which is a biologically plausible implementation of Q-learning that uses three phases. Upon presentation of the stimulus, the activity of the input neurons remains constant and activity propagates through the recurrent connections of the network. We selected the action as described above once the activity of the units settled, but we did not wait longer than 50 timesteps. When an action was selected, an "attentional" signal originating from the winning action was propagated through an accessory network to determine the influence of each neuron on the selected action. The network then gets a reward *r* from the environment and computes a reward prediction error δ that is determined by

$$\delta = r - Q_a \tag{5}$$

Where Q_a is the Q-value of the selected action *a*. The reward prediction error δ is broadcasted to the whole network by a neuromodulatory signal, such as dopamine (Schultz et al., 1997) and is thus available at all synapses. δ and the "attentional" feedback signal are combined to update the weights of the network, as will described below in the RELEARNN section.

Activation and gating functions

A popular activation function to train recurrent neural networks is the ReLU function because it alleviates the problem of the vanishing gradient (Glorot et al., 2011). However, ReLU functions offer no guarantees that the network will reach a stable state (Simard et al., 1989), which is a necessary condition for learning rules such as RELEARNN. On the other hand, squashing nonlinearities like sigmoid or tanh functions guarantee that the network reaches a stable state (Liao et al., 2018), but may suffer from the vanishing gradient problem, which might hurt performance (Linsley et al., 2020). To guarantee that our network reaches a stable state while mitigating the vanishing gradient problem, we used a ReLU-like function with a slope that decreases after a threshold:

$$\sigma(x) = \begin{cases} max(0, x), & x \leq 6 \\ log(1.5 \cdot x + 1) + c, & otherwise \end{cases} \tag{6}$$

Where c is set to $6 - log(10)$ to ensure that σ is continuous. We will describe below how the networks learn to label relevant pixels with a response enhancement. We observed that this process was hampered if the activation function was bounded from above when curves were long. In that situation, units coding for all pixels reached the bound and the networks could no longer differentiate between the target and the distractor curve. The activation function of Eq. 6 ensured that the network differentiated between long target and distractor curves.

The horizontal gating of feedback units by feedforward units, was also governed by a non-linear function. The activity of neurons of the feedforward units was passed through a continuous approximation of a step-function:

$$\varphi(x) = \frac{100x}{\sqrt{1 + (100x)^2}} \tag{7}$$

Here $\varphi(x)$ is non-negative because x is non-negative (see Eq. 2 for the role of $\varphi(x)$).

RELEARNN

A popular algorithm to train recurrent neural network is backpropagation trough time (Werbos, 1988, 1990). The algorithm needs to memorize the consecutive states of every unit because this information is needed to update the weights. It is non-biological because it is neither local in time nor in space (Lillicrap & Santoro, 2019). We therefore used RELEARNN (Brosch et al., 2015), a learning rule that is local in time and in space and equivalent to backpropagation in a reinforcement learning paradigm (Almeida, 1987; Pineda, 1987). RELEARNN uses a separate accessory "attentional" network with one accessory unit for every regular network unit (Fig. 4). On each trial, the output layer chooses one of the possible actions. The accessory units represent the influence of the corresponding regular units on this chosen action. Once an action has been chosen, the winning unit in the output layer activates accessory units and thereby initiates the propagation of activity in the accessory network. The strength of a connection from unit A to unit B in the accessory network is proportional to the strength of the connection from unit B to unit A in the regular network so that the activity in the accessory network (dotted lines in Fig. 4) equals the influence of regular units on the chosen action, once the accessory network converged to a stable state (Fig. 4, see Brosch et al., 2015 for details). Accessory units corresponding to regular units with a strong influence on the Q-value of the chosen action are very active and those that correspond to regular units that did not influence this Q-value are silent. The network of accessory units can thereby assign credit (or blame) to the regular units that are responsible for the outcome of the action. A change in the synapses between the regular units with highly active accessory units has a pronounced influence on the Q-value of the chosen action.

We distinguish between three phases on a trial (Fig. 4). In the first phase, the stimulus is presented as a pattern of activity across the input neurons. Activity propagates trough the feedforward, horizontal and feedback connections among regular feedforward and recurrent units until the network converges to a stable state. At the end of this phase a presynaptic neuron *i* and a postsynaptic neuron *j* will have activity x_i^∞ and y_j^∞. At the start of the second phase, an action is selected, which is represented by one of the regular units in the output layer. The corresponding unit in the motor layer of the accessory network is activated and activity is propagated within the accessory network until a stable state is reached (in our experience this phase can be terminated after a fixed number of timesteps; Liao et al., 2018). At the end of the second phase, neurons *i* and *j* in the accessory network have activity x_i^{acc} and y_j^{acc}, which is proportional to the influence of the corresponding regular units on the chosen action value (see Brosch et al., 2015 for a proof):

$$x_i^{acc} \propto \frac{\partial Q_a}{\partial x_i^\infty} \tag{8}$$

In the third phase, the network receives a reward if the correct action was selected, and no reward in case of an erroneous response. The reward prediction error is broadcasted through the network using a neuromodulator and weights are modified according to a four-factor Hebbian learning rule:

$$\Delta w_{ij} = \delta \cdot x_i^\infty \cdot y_j^{acc} \cdot \sigma'\left(y_j^\infty\right) \tag{9}$$

Note that the weight update of RELEARNN is local in space, as weights are modified using only pre- and postsynaptic information, and local in time, because it is not necessary to memorize the activities of the units at every time step.

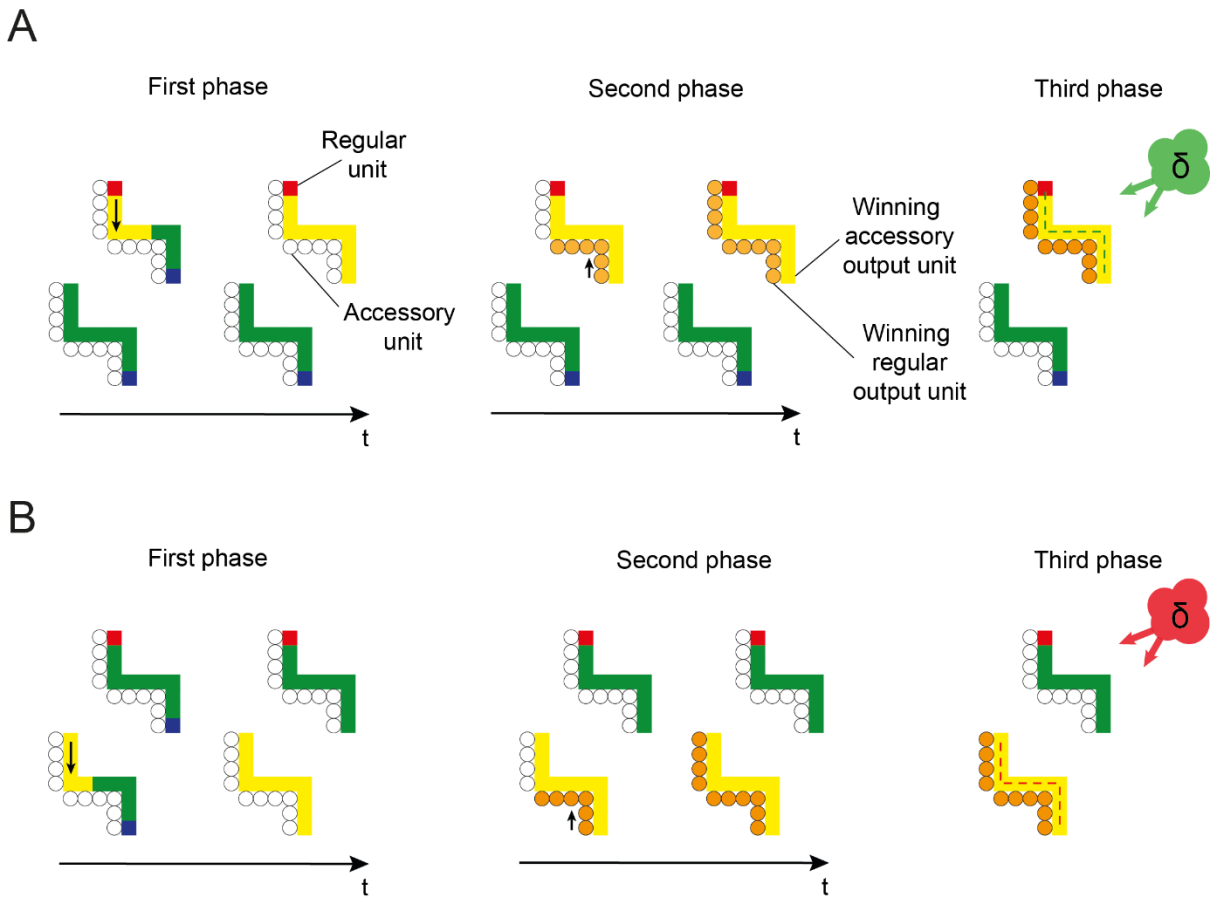

Fig. 4. RELEARNN for an example stimulus. A. Correct choice by the network. In the first phase, activity propagates in the regular network that has both feedforward and recurrent connections (squares) until this network reaches a stable state. In the second phase, a winning unit is selected and activates the corresponding unit of the accessory network (small circles). From there, activity propagates in the accessory network (small orange circles) to tag the connections that influence the Q-value of the chosen action. After a few timesteps, activity in the accessory unit x_i^{acc} becomes proportional to the influence of the activity of the corresponding regular unit x_i^{∞} on the chosen output unit Q_a. In the third phase, a reward is given if the action was correct, or not in case of an error, and a neuromodulator (green cloud, δ) broadcasts the reward prediction error to the network. Weights are changed according to a four-factor Hebbian learning rule. (green connections between the units.) **B. Incorrect choice by the network**. In this case, the reward prediction error is negative (red cloud) and the weights are changed accordingly.

Results

Curve-tracing task

We trained the networks with a curriculum to perform the curve-tracing task, a strategy that is also used to train monkeys. We first presented a stimulus with one blue and one red adjacent pixel, which were placed randomly on the grid, to train the network to select blue pixels as target for an eye movement. If the network achieved an accuracy of 85% during a test phase with 500 trials, during which we fixed the weights and switched exploration off, the network was trained with a target and a distractor curve with a length of three pixels. When the network achieved an accuracy of 85%, we added one pixel to the curves until the length of the curves was 9 pixels. Using this curriculum, we trained 24 networks and judged that they had learned the task if they achieved an accuracy larger than 85% for curves that were 9 pixels long. All networks learned the task, within an average of 19,000 trials.

We next aimed to gain insight into the strategy of the networks. The curves were drawn randomly on each trial (see Fig. 5A for an example stimulus). The number of possible curves grows exponentially with their length, and we therefore conjectured that it might be impossible to memorize all possible input patterns (Roelfsema et al., 1999). Instead, we hypothesized that the networks learned a general rule, which could be based on applying an elemental curve-tracing operation. To test this hypothesis, we exposed the networks to curves that were longer than those shown during training. When a network achieved an accuracy larger than 85% for curves of length N, we switched off exploration and tested curves with lengths ranging from N+1 to N+4. We found that the networks' accuracy was above chance levels for curves that were longer than the curves in the training set (Fig. 5B). For example, networks trained on curves up to a length of 9 pixels generalized to curves with a length of 13 pixels (red in Fig. 5B) ($p<10^{-6}$, Wilcoxon signed-rank test). This generalization emerged during training, indicating that networks learned to trace connected pixels as a generalized elemental operation (Roelfsema et al., 1999).

We next examined the activity of units of trained networks and compared it to the activity in the visual cortex of monkeys that had been trained on a similar curve tracing task (Roelfsema et al., 1998). We first compared the activity of networks units whose receptive field fell on the target curve to the activity of units whose receptive field fell on the distractor curve. By design, the activity of the units of the feedforward group is not modulated during curve-tracing, whereas units of the recurrent group could learn to propagate enhanced activity (Fig. 5A,C). Indeed, the activity of units whose RF fall on the target curve is enhanced compared to units whose RF fall on the distractor curve.

Fig. 5A illustrates the flow of activity in the recurrent input layer of one of the networks for an example stimulus. At t=0, the stimulus is presented and the feedforward flow of activity is complete. Feedforward units with a pixel of the appropriate color in their receptive field are active but the recurrent units do not yet discriminate between the target and distractor curve. After a few timesteps, enhanced activity starts to spread from the red cue at the beginning of the target curve. However, the end of the curve is not yet labeled and the network doesn't have enough information to make the appropriate eye movement. After 40 timesteps, however, the enhanced activity has spread over the entire representation of the target curve to reach the blue eye movement target, which is selected by the network.

This simulation result resembles neuronal activity in the visual cortex of monkeys solving a similar curve-tracing task. Fig. 5D shows the average activity of V1 neurons in the visual cortex of monkeys elicited by a target curve (orange) and a distractor curve (blue). Initially, the neurons were activated by the appearance of a contour element in their RF, and this feedforward response was the same for the target and distractor curves. After a delay of 130ms, recurrent processing caused the V1 representation of the target curve to be enhanced over the representation of the distractor curve (Roelfsema et al., 2003).

We quantified the magnitude of the response enhancement across 345,000 stimuli and the 24 networks that were trained successfully with a modulation index (Roelfsema et al., 1998), which is the difference between the response evoked by the target and distractor curve divided by the average response:

$$MI = \frac{Target - Distractor}{(Target + Distractor)/2} \tag{10}$$

A positive modulation index signifies an enhancement of activity evoked by the target curve relative to that evoked by the distractor curve. We observed that the recurrent units representing target pixels had a higher activity than those representing distractor pixels. The mean modulation index was 1.18, corresponding to an activity enhancement of 292% (Fig. 5E). This result is similar to what has been observed in monkeys. However, the increase in

the representation of the target curve in area V1 of monkeys was weaker (Fig 5F, see also Roelfsema et al., 1998), although the V1 modulation index distribution in monkeys includes neurons that were not modulated by curve-tracing, resembling feedforward units of the network.

In the visual cortex of monkeys, the enhanced activity spread iteratively, starting at the red cue at the beginning of the curve until it reached the end (Fig 5H). Accordingly, units in the model representing pixels close to the beginning of the target curve enhanced their activity before units representing pixels that were farther away. To quantify the time-point of modulation, we used the moment when the difference between the activity elicited by the target and distractor curve became larger than 70% of its maximum (Fig. 5G). An analysis of modulation latencies across units confirmed that the onset of modulation was significantly later for pixels farther from the fixation point (7 pixels away), compared to pixels closer to the fixation point (2 pixels away) ($p<10^{-15}$, Mann-Whitney U test).

In monkeys, the spread of activity among neurons in the visual cortex is accompanied by an enhancement of the activity of neurons coding for the target curve in the frontal eye fields, which is an area involved in the planning and generation of eye movements (Khayat et al., 2009; Pooresmaeili et al., 2014). Interestingly, the spread of enhanced activity was also evident in the model's output layer, which represents the upcoming eye movement (Fig. S1). Hence, the model had learned a strategy of spreading enhanced neuronal activity along the entire target curve in multiple layers. In the model, spreading is iterative and relies on horizontal and feedback connections, which makes the process generalizable across stimuli and curve lengths, in spite of the limited network depth. The results are also in line with the results by Jolicoeur et al. (1986) showing that the reaction times of human observer in a curve-tracing task increase linearly with the length of the curve that needs to be traced.

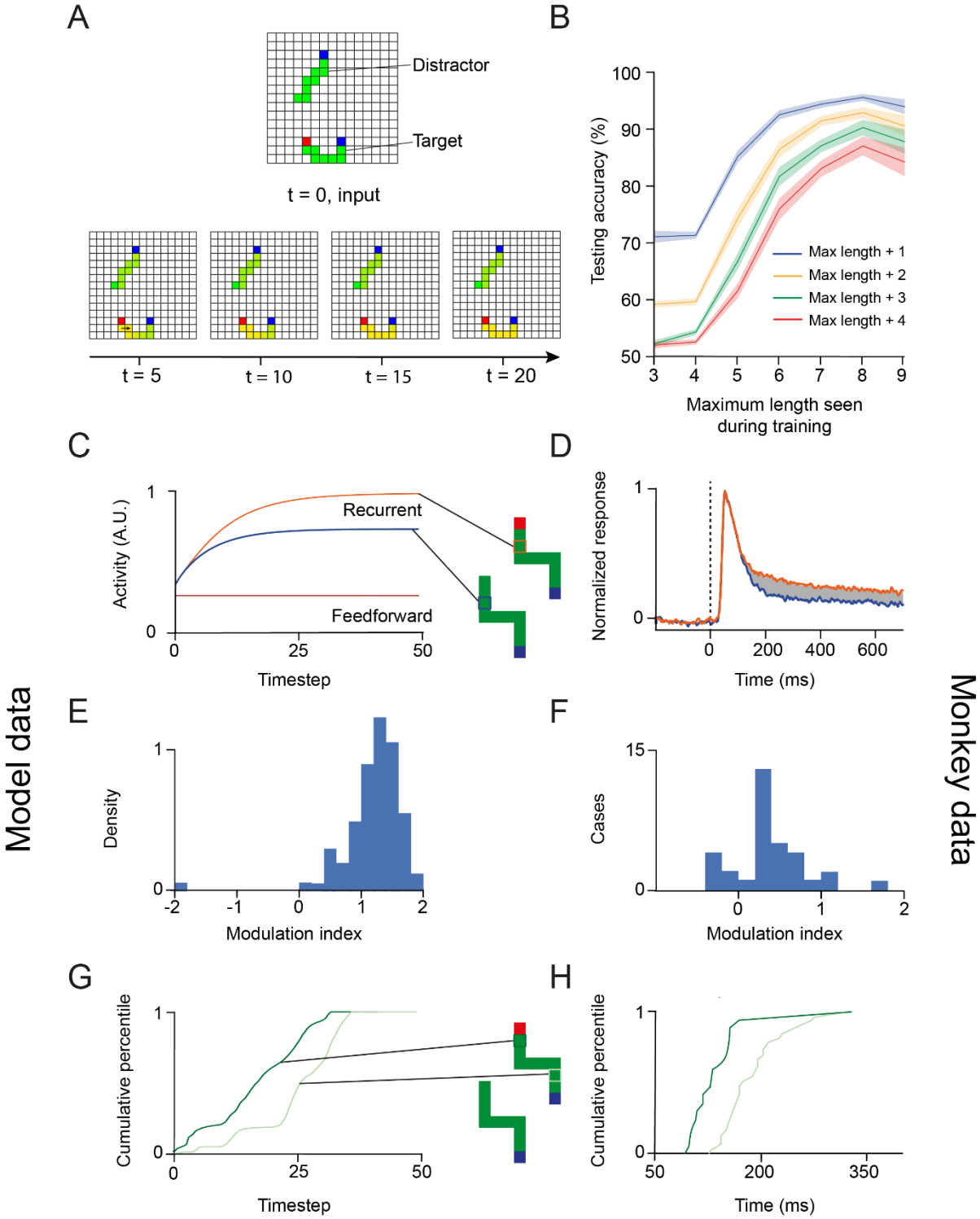

Fig. 5. Propagation of enhanced activity across the representation of the target curve during curve-tracing. A. Upper, example stimulus presented to one of the networks. The target curve starts with a red pixel. Lower, activity of recurrent units in the input layer across time. The yellow color denotes an increase in activity. Note the spread of enhanced activity over the representation of the target curve, starting at the red pixel. **B**. Testing accuracy for curves of length up to N+4 pixels where N is the maximum length used during training. At the beginning of training, the model does not generalise to longer curves. At the end of training, a model trained with curves up to 9 pixels long, generalized to curves with up to 13 pixels ($p<10^{-6}$, curves with 13 pixels, Wilcoxon signed-rank test). **C**. Activity of an example unit in the recurrent group elicited by the target (orange) or distractor curve (blue), and activity of the corresponding unit in the feedforward group (brown). Note that the activity elicited by the target curve is enhanced compared to that elicited by the distractor curve**. D**. Average activity of neurons in area V1 of the visual cortex of monkeys during a curve tracing task, when their RF fell on the target curve (orange) or on the distractor curve (blue). Adapted from (Roelfsema et al., 2003). **E.** Distribution of the modulation index across recurrent units of the neural networks. A positive value indicates an enhanced response to the target curve. **F**. Distribution of modulation index in area V1 of the visual cortex of monkeys (from Roelfsema et al., 1998) **G**. Distribution of the modulation latency across units of the network. The onset of modulation is delayed for units representing pixels that are farther (7 pixels away), compared to pixels that are closer (2 pixels away) to the beginning of the curve ($p<10^{-15}$, Mann-Whitney U test). **H**. Distribution of the modulation latency across recording sites in monkeys performing the curve-tracing task, adapted from Pooresmaeili & Roelfsema, 2014. Dark green represents RF that were close to the fixation point, and light green represents RF that were farther from the fixation point.

How does the network learn to trace, i.e. to group pixels into curves, when the only feedback after each trial is the presence or absence of a reward? To understand the learning process enabled by RELEARNN, we examined the activity of the accessory network units that influence learning. Figure 6A illustrates how the activity propagates among the accessory units that represent the selected curve, starting from the selected eye movement and then across the representation of other pixels of this curve, i.e. in the opposite direction of the activity propagation. This strength of this accessory signal represents the influence of the units of the regular network on the selected action. If the network happens to select the target curve, the connections between regular units are strengthened in proportion to the reward-prediction error δ. If the network erroneously selects the distractor curve, the credit assignment signal ensures that that the recurrent connections among units representing this curve are weakened because δ is negative. The unique and distinguishing feature of the target curve is that it starts with a red pixel, and the credit assignment signal of the accessory network ensures that the representation of this red pixel becomes the source of activity propagation in well-trained networks during learning.

An analysis of the spread of activity within the accessory network also provides insight in the role of the curriculum. If the network has been trained with curves up to a certain length N, and is tested with curves that have N+1 pixels, the credit assignment signal reaches the fixation point if the model selects the appropriate eye movement and the relevant connections increase in strength (Fig. 6B, left). However, when we presented curves with N+5 pixels, the signal at the beginning of the curve did not discriminate between the target and distractor curves and learning did not occur (Fig. 6B, right). Hence, training was more efficient with a curriculum in which we gradually increased the length of the curves.

Let us briefly summarize how the networks learn to solve the curve-tracing task. Upon presentation of the stimulus, the network starts with a feedforward processing phase which does not discriminate between the target and the distractor curve. This early activity pattern has been called ‘base representation’ in previous work (Roelfsema, 2006). During the later, recurrent processing phase, units learn to propagate enhanced activity along the representation of the target curve, starting at the fixation point with its unique color. This enhancement of activity cannot spill over from the target curve to nearby curves because feedforward neurons that respond to image locations without pixels are silent. The activation of these feedforward neurons is necessary before the recurrent units activate. In other words, the gating process of the base representation ensures that enhanced activity can only spread along the representations of connected pixels.

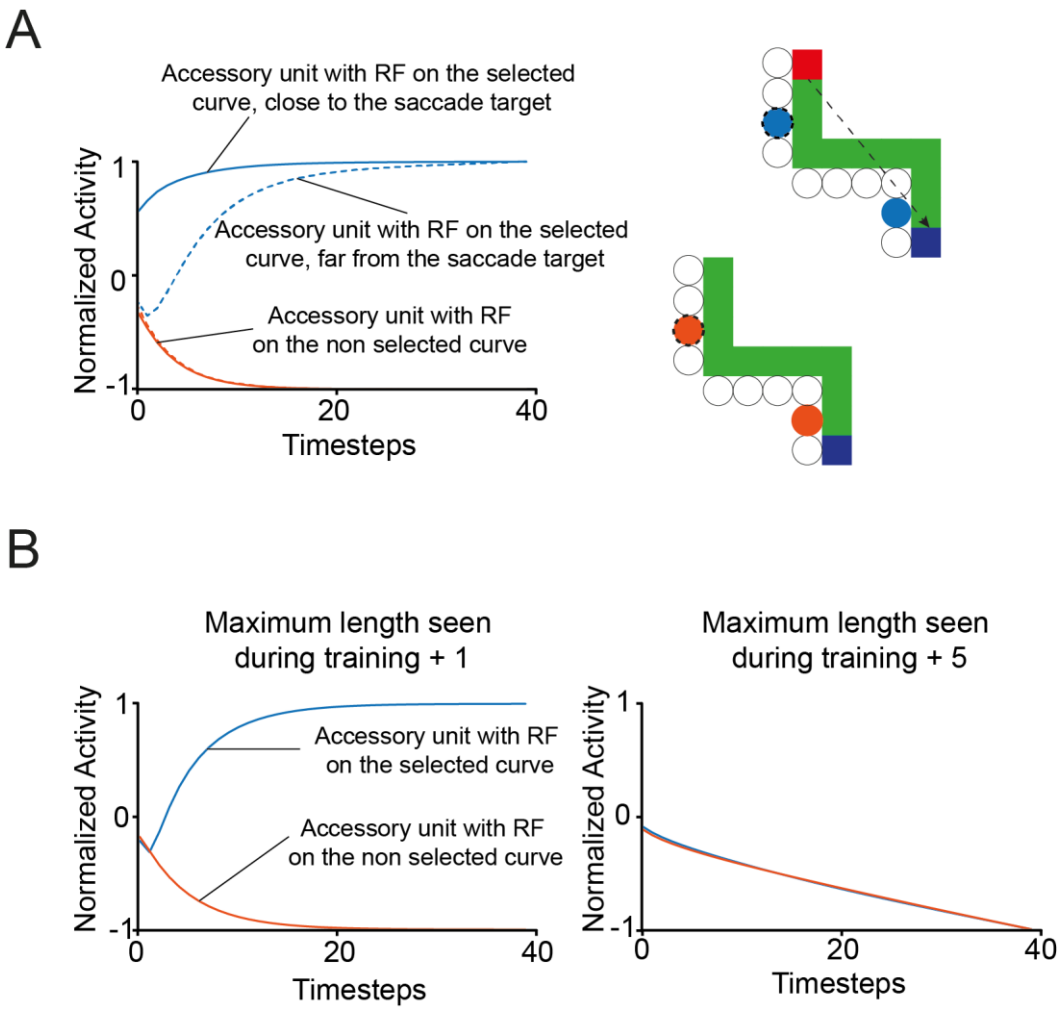

Fig. 6. Activity in the accessory network. A. Activity of units in the accessory network whose RFs fall on the selected curve (blue traces) or the non-selected one (orange traces), at different distances from the blue pixel that is target of the eye movement (continuous and dotted traces show the activity of accessory units representing pixels nearer and farther to the saccade target, respectively). Hence, the credit assignment signal propagates in the opposite direction compared to the enhanced activity, starting from the selected eye-movement target. This credit assignment signal is absent from the representation of the distractor curve. **B.** Activity of units corresponding to the beginning of the selected and non-selected curves in the accessory network, for curves that were one (left panel) or five pixels longer (right) than the curves used during training. If the length of the curve was similar to that in the curriculum, the credit assignment signal propagates to the beginning of the selected curve (red fixation point on correct trials) and training is effective. However, if the curves are much longer, the credit assignment signal does not all other pixels of the selected curve and training fails.

Search-then-trace task

Having established that the networks learn to trace curves, we explored more complex multistep visual routines that require the succession of multiple elemental operations, and the transfer of information between them. The search-then-trace task required a visual search for a colored marker followed by curve-tracing. In the task for the model, the color of relevant markers was brown or red and the color if the target marker was cued by activating one of two input units (color cue in Fig. 7A). The network had to trace the target curve starting at this marker and to select an eye movement to the blue pixel at the other end of this curve.

We trained the network using a curriculum that was similar to the one used for the trace task. First, we presented a stimulus with a blue pixel connected to a red pixel randomly placed on the grid. If the network achieved 85% accuracy during a 500 trials test phase with weights fixed and no exploration, we presented a target and a distractor curve with a length of

three pixels, where one of the two curves started with a brown pixel and the other one with a red pixel. If the network achieved 85% accuracy during the test phase, we added one pixel to the curves and repeated this until the curves were 9 pixels long.

We trained 24 networks and 79% of them learned the task within an average of 40,000 trials. Just as in the curve-tracing task, the networks learned a solution in which they generalized across different shapes of the curves and for both marker colors. Trained networks generalized to curves that were 4 pixels longer than the ones presented during training (Fig. 7B, red curve, $p<10^{-4}$, Wilcoxon signed-rank test).

How does the network learn to transfer information between elemental operations such that the output of an operation can be used at the input of the next one? In the visual cortex of monkeys, this information transfer appears to take place by the persistence of enhanced neuronal activity in early visual cortex between the elemental operations. In the search-then-trace task, the visual search for the marker with the cued color enhances the activity of neurons with a RF at this marker (Fig. 1A), as if it is ‘written' to the cognitive blackboard represented in early visual cortex. The subsequent tracing operation can ‘read’ this position and use it as the start for the curve-tracing operation.

Interestingly, the model discovered a similar method to transfer information between the successive search and trace operations. In the model, recurrent processing started with an enhancement of the representation of the target marker in the recurrent input layer (Fig. 7A). After training, the horizontal connections in the recurrent input layer between the color cue units and the retinotopic units of a same color are strengthened compared to those between the other colors, and the enhanced activity propagated through those connections to the retinotopic unit with the same color as the cue. This strategy solved the search operation. As in the trace operation, the gating by the feedforward units of color cue units and retinotopic units ensured that the enhanced activity due to the search operation is confined to units with an active element in their receptive field.

After the search operation, the enhanced activity propagated across the representation of the target curve (Fig. 7A, green traces in Fig. 7C). As in the curve-tracing task, the time course of activity propagation in the output layer was similar to that in the visual layers of the model (Fig. S1). We next examined the time course of enhanced activity of the units across 260,000 stimuli and the 19 networks that learned the task. This analysis revealed that the early enhancement of the activity of units with a RF on the target marker (red in Fig. 7E), followed by the propagation of enhanced activity across the target curve (green in Fig. 7E) was a general feature. Fig. 7F shows similar results in area V1 of the visual cortex of monkeys performing a search-then-trace task (Moro et al., 2010). Also in the monkey visual cortex, the marker with the relevant curve is first labeled with enhanced activity, before activity spread across the V1 representation of the target curve.

These results, taken together, indicate how RELEARNN, a biologically plausible reinforcement learning rule, can train networks to execute a visual routine composed of multiple elemental operations and to use persistent enhanced activity to transfer information from one element operation to the next.

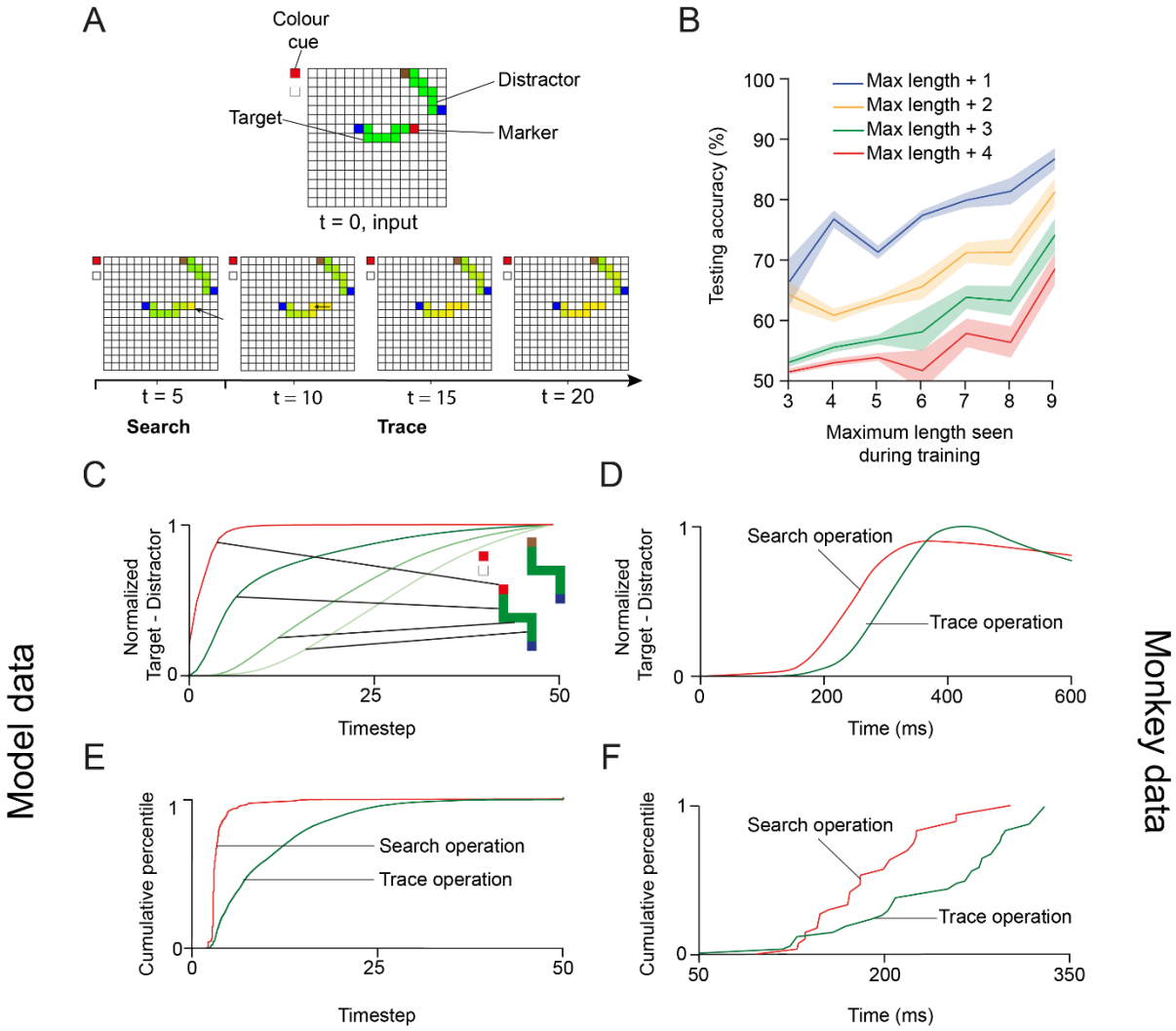

Fig. 7. Search-then-trace task. A. Example stimulus shown to one of the networks. Upper panel, visual stimulus. Lower panel, yellow shading shows the propagation of enhanced activity among recurrent units of the input layer, starting at the representation of the red marker, which is highlighted as the result of the search operation. From here, the enhanced activity spread along the curve (trace operation). **B.** We tested how well the models generalized to curves that were longer than those presented during training. Generalization was better for networks that had been trained on longer curves (x-axis). E.g. networks trained on curves up to a length of 9 pixels generalized to curves with 13 pixels ($p<10^{-6}$, Wilcoxon signed-rank test). **C**. Normalized response enhancement for the target marker and target curve. Each curve is normalized by its maximum over time. First the activity of the unit with a RF at the location of the target marker was enhanced (search operation, red curve). Thereafter, enhanced activity propagated across the target curve connected to it (trace operation, green curves). **D**. In the visual cortex of monkeys, the representation of the target marker is enhanced (red) before the enhanced activity spreads over the V1 representation of the target curve (green; adapted from Moro et al. 2010). **E**. Distribution of the latency of the response enhancement across 260,000 stimuli and 19 networks. The latency of the modulation related to the search operation was shorter than that related to curve-tracing ($p<10^{-15}$, Mann-Whitney U test). **F**. Distribution of the latency of response enhancements across V1 neurons in monkeys solving the search-then-trace task (adapted from Moro et al., 2010).

Trace-then-search task

We next used the trace-then-search task to examine if the network could learn a different visual routine composed of the same elemental operations that need to be applied in the opposite order. First, the agent has to trace the curve connected to the fixation point, which was now blue. This target curve ended with a marker that was either red or brown which was the color that the model had to search because the model had to select the pixel with this color on the side of the grid (Fig. 1C).

We adapted the curriculum by first training the networks on simple curve-tracing tasks in which the target curve started with a blue pixel, and the two curves ending with a red and a brown marker. When the networks had learned to trace curves with a length of 9 pixels, we added the two, colored pixels outside the grid and required the model to select the pixel with the same color as the target marker for an eye movement.

Of the 24 networks trained on this task 66% learned it, after an average of 37,000 trials. As in the previous tasks, the networks learned a general solution, which did not depend on the position or color of the marker or on the shape of the curve. Trained networks also performed the task for curves that were 4 pixels longer than the curves of the curriculum (Fig. 8B, $p=1.4\cdot10^{-4}$ for curves with a length of 13 pixels, Wilcoxon signed-rank test).

Trained networks first spread activity along the target curve until its end was reached (Fig. 8A,C) and the target color could be identified. The enhanced activity then spread to colored eye movement target, as the outcome of the search operation. Hence the networks learned to first perform a trace operation, followed by a search operation. As in the previous tasks, the time course of modulation in the output layer followed a similar pattern as in the input layer (Fig. S1). The progression of activity was also evident when we examined the activity across 230,000 stimuli in the 16 networks that were trained successfully. The green trace in Figure 8E illustrates the time-point in which the activity of the representation of the last, colored pixel of the curve was enhanced, which corresponds to the output of the curve-trace operation (green curve). Thereafter, the representation of the colored eye-movement target was enhanced, corresponding to the realization of the search operation (red curve in Fig. 8E). The results are qualitatively similar to those in the visual cortex of monkeys trained on the trace-then-search tasks, where the enhanced neuronal activity first spread along the target curve (Fig. 8F, green trace), before the V1-representation of the colored disk that was the target of the eye movement was enhanced (Fig. 8F, red curve).

These results indicate that RELEARNN can train networks to carry out elemental operations in different orders depending on the task at hand, which is remarkable for a biologically inspired learning rule in which the only feedback that the networks receives is a reward upon a correct choice.

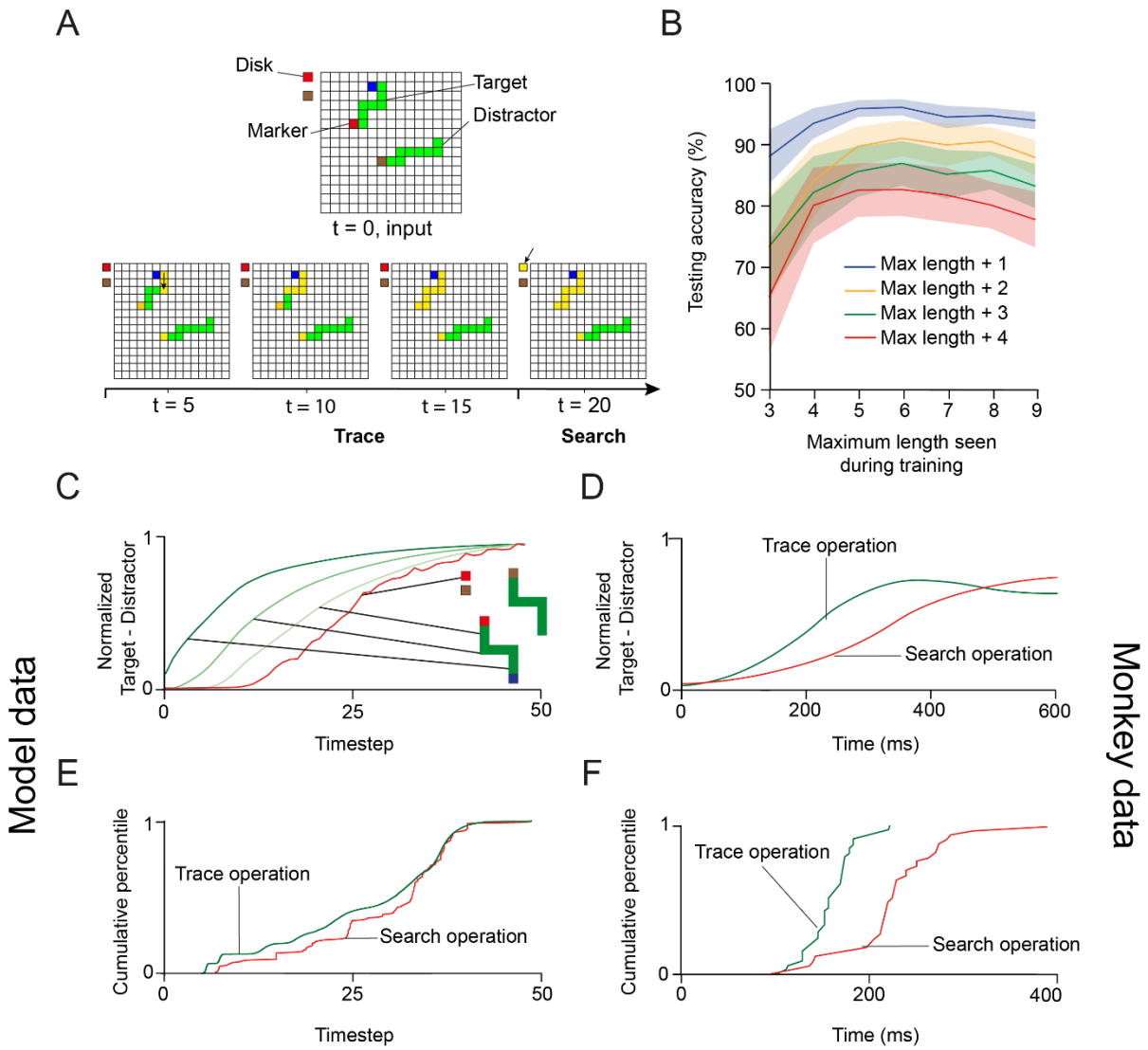

Fig. 8. Model performance in the trace-then-search task. A. Example stimulus shown to one of the networks. Upper, an example stimulus. Lower, the spread of enhanced activity is shown in yellow. It first spreads over the curve starting at the blue cue and reaches the target marker at the other end, cuing the color that needed to be selected during the the search operation. **B.** Testing accuracy for curves of length up to N+4 pixels where N is the maximum length in the curriculum. The generalization performance improved when the network learned to trace longer curves ($p=1.5\cdot10^{-4}$ for curves of 13 pixels, Wilcoxon signed-rank test). **C**. Normalized response enhancement for target pixels, averaged across units. Each curve is normalized by its maximum over time. First the curve connected to the fixation point is labeled with enhanced activity (trace operation, green curves) and then the units that represent the correct eye movement target, i.e. with the same color as the target marker, enhanced their activity (search operation, red trace). **D**. In the visual cortex of monkeys, the response enhancement also first labels the segments of the target curve (green trace), before it labels the position of the eye movement target (red trace; adapted from Moro et al., 2010). **E**. Distribution of the modulation latency across model units (230,000 stimuli and 16 networks). The response modulation of trace operation precedes that of the search operation ($p=1.5\cdot10^{-5}$, Mann-Whitney U test). **F**. Distribution of the modulation latency across recording sites in monkeys solving the search-then-trace task (adapted from Moro et al., 2010).

Discussion

We studied how visual reasoning tasks can be learned by trial and error, using a biologically plausible learning rule. We took inspiration from previous work on visual routines, which are built from elemental operations (Ullman, 1984). Studies of the visual cortex revealed that elemental operations, like curve-tracing and visual search, are associated with the enhancement of neuronal activity elicited by target elements over distractor elements. Specifically, curve-tracing is associated with the spread of enhanced activity across a target curve and visual search results in an activity enhancement of the representation of the target of search. We found that the neural networks trained to trace curves or to carry out visual routines relying on search and curve-tracing developed a similar strategy. They learned to successively label task-relevant image elements with enhanced neuronal activity. We also studied how these networks learned to transfer information between one elemental operation and the next. We observed that the strategy of labeling image elements with enhanced activity provides a mechanism for information transfer, because the focus of enhanced activity that was the result of one operation could be used to start the next elemental operation. Note that this mechanism is equivalent to the binding of parameters to subroutines in a computer program.

The model architecture included a feedforward group of units that is only sensitive to input from lower layers, and a recurrent group that also receives feedback connections from higher layers and horizontal connections from units in the same layer. Non-zero activity of units in the feedforward group allows the corresponding recurrent units to become active, implementing a form of gating. This gating process prevents that recurrent activity can spill over to units that represent the blank screen, helping with the selectivity of the propagation of activity among units representing connected image elements. A comparable division of labor has been observed in the visual cortex, where some neurons are influenced by curve-tracing and other neurons are not (Pooresmaeili et al., 2010). The neurons that are and are not sensitive to recurrent processing are clustered in the different layers of cortex, as has been observed during curve-tracing and figure-ground segregation tasks (Self et al., 2013; Van Kerkoerle et al., 2017). Neurons that are most sensitive to the feedforward input are enriched in layers 4 and 6, whereas the influence of figure-ground segregation and visual attention are more pronounced in the superficial layers 2 and 3 and in layer 5 (Self et al., 2013; Van Kerkoerle et al., 2017). Thus, the feedforward group could correspond to neurons in layers 4 and 6 while the recurrent group could correspond to neurons in layer 2, 3 and 5. This view is supported by recent experimental evidence for the existence of separate channels for feedforward and feedback processing, suggesting that these influences might be mediated by partially non-overlapping neuronal circuits (Semedo et al., 2022)

We used RELEARNN to train the networks. RELEARNN is a biologically plausible learning rule that is equivalent to the Almeida-Pineda algorithm but is local both in space and in time so that all the information that is necessary for the synaptic changes are available at the synapse. The learning rule relies on two networks. The first network is active during inference when activity propagates through feedforward and recurrent connections, during the processing of the stimulus. The second network is an accessory network, dedicated to help with the learning process. The activity is initiated by the action that is selected in the output layer and it then propagated through the network, emphasizing the synapses that have a strong influence on the activity of the winning output unit.

We have used two networks, composed of regular recurrent units and of accessory units, which are presumably both related to the role of visual attention at a psychological level of description. In a study in area V4 of the visual cortex of monkeys, Steinmetz et al. (2014) used a task in which the animals were cued to make an eye movement to a search target but to make an eye movement to an opposite location in space. Some V4 neurons enhanced their

activity when their RF was on the target of search, which would correspond to the regular recurrent units of the present study. Other V4 neurons enhanced their activity if their RF was on the target for the eye movement, which may correspond to the accessory units that play a role in learning. The present work suggests that the propagation of enhanced activity by the regular units and the propagation of credit assignment signals by accessory units, in the opposite direction, might be separable functions of the attentional networks. We do realize, however, that the distinction between these two forms of attention is speculative and that much remains to be learned about the flow of attentional selection signals in brain circuits.

When the activity in the regular and accessory networks is combined with the reward prediction error, which can be mediated by a globally released neuromodulator such as dopamine, all the information for the synaptic update is available locally at the synapse. In previous work RELEARN was shown to train networks in the curve-tracing task (Brosch et al., 2015) and here we showed that it is powerful enough to train networks on more complex, multistep tasks.

It is of interest that the networks developed strategies like those observed in the visual cortex of monkeys. During curve-tracing, recurrent units in the lower and higher layers propagated enhanced activity along the representation of the target curve. This finding resembles how curve-tracing is implemented in the visual cortex of monkeys, where the neurons with RFs on the target curve successively enhance their activity, starting at the beginning of the curve. The spreading process continues until the response enhancement reaches the end of the curve. These responses enhancements occur in multiple cortical areas including the frontal cortex, where the correct eye movement is selected (Pooresmaeili et al., 2014). Furthermore, the model implemented visual search by enhancing the activity of units coding for the relevant color, which then ultimately led to an enhanced representation of the location where this color was found, just as is observed in the visual cortex (Chelazzi et al., 1998, 2001; Desimone & Duncan, 2003; Mirpour & Bisley, 2012; Sato et al., 2001; Zhou & Desimone, 2011). Finally, this propagation of enhanced activity was used by the model to transfer information from one elemental operation to the next like in the visual cortex of monkeys.

The present study builds on results from previous studies that used artificial neural networks to solve sequential visual tasks. Marić & Domijan (2019) configured a model to trace curves and they also combined feedforward and recurrent network units. They examined the activity propagation along curves, but they did not examine learning of the curve-tracing task. In their network, tracing could proceed faster for target curves that were farther from a distractor curve than for target curves that were near to a distractor. This result reproduced psychophysical and neurophysiological data showing that tracing speed increases if there is a larger distance between the curves (Jolicoeur et al., 1991; Pooresmaeili & Roelfsema, 2014). Such a scale dependence could rely on the hierarchical organization of the visual cortex where neurons in higher areas have larger receptive fields. If curves are far apart, activity can propagate between neurons at higher levels with larger RFs, whereas if the curves are nearby, activity could propagate at a lower level, where the spatial resolution is higher, but at the cost of increased processing times (Roelfsema & Houtkamp, 2011).

Linsley et al. (2019) trained recurrent neural network on a pathfinder task, which resembles curve-tracing, but now the curves consists of shorter, disconnected line elements. Their network had to determine whether two dots were on the same path. The authors demonstrated that a recurrent unit, called h-GRU (Mély et al., 2018), which activates based on feedforward connections but also integrates horizontal input, was useful to solve the task. Unlike the present study, the authors used backpropagation through time, which is a non-local learning rule that requires a teacher. Instead, in the present work we used RELEARNN, which is a biologically plausible learning rule. In addition, we analyzed the strategies that emerged after training and trained networks on visual routines, in which curve-tracing was combined

with visual search. Although we focused on curve-tracing, future modeling studies may also examine other tasks in which the surfaces of 2D objects need to be grouped in perception. For example, Jeurissen et al. (2016) found that the time it takes to determine whether two dots are on the same surface increases with the distance between the dots. The same principle that enables our network to trace curves could be expanded to enable networks to group 2D image regions, which are important in real-life scenarios, and the principles of region labeling might even be generalized to more complex situations that require semantic segmentation.

In conclusion, the new model architecture illustrates the versatility of networks that spread enhanced activity to solve elemental operations and to glue them together into visual routines, ensuring the transfer of the relevant parameters. The similarity between the network activity and neurons in the visual cortex indicate that the approach provides insight into how the circuits in the visual brain can configure themselves during learning, based on biological plausible learning rules that combine attentional and neuromodulatory signals. Future studies could expand this approach to further elucidate how feedforward, feedback and horizontal connections are used to enhance the cognitive capabilities of brain networks.

Methods

Data Analysis

We trained 24 networks for each task. For the further analysis, we only kept networks that learned the tasks for curves up to 9 pixels long. We froze the weights after training and set the exploration rate ϵ to 0. We presented 15,000 randomly generated stimuli to each network and analyzed the activaty of the units for these stimuli, including only correct trials in the analysis.

Parameters

All networks were initialized and trained with the same parameters. They are summarized in the following table

Parameter	Value
Learning rate δ	0.02
Exploration rate ϵ	0.05
Weights between hidden layers	Sampled from U(0,0.1)
Weights between the last hidden and the output layers, same receptive field	Sampled from U(0,0.1)
Weights between the last hidden and the output layers, different receptive field	Sampled from U(-0.01,0)

Here U(x,y) refers to the uniform distribution on the interval [x,y].

Code availability

All the code used to train the networks and to analyze the data is available on the following GitHub address: https://github.com/samimol/Visual-Routines

Acknowledgements

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3, Task 3.7), an ERC grant (101052963 "NUMEROUS"), an NWO Crossover grant 17619 "INTENSE", "DBI2" (a Gravitation program of the Dutch Ministry of Science) and NWO-OCENW.KLEIN.178.

We acknowledge the use of Fenix Infrastructure resources, which are partially funded from the European Union's Horizon 2020 research and innovation programme through the ICEI project under the grant agreement No. 800858.

References

Almeida, L. B. (1987). A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. *Proceedings of the IEEE First International Conference on Neural Networks*, *II*, 609–618.

Brosch, T., Neumann, H., & Roelfsema, P. R. (2015). Reinforcement Learning of Linking and Tracing Contours in Recurrent Neural Networks. *PLoS Computational Biology*, *11*(10), e1004489. https://doi.org/10.1371/journal.pcbi.1004489

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940. https://doi.org/10.1152/JN.1998.80.6.2918/ASSET/IMAGES/LARGE/JNP.DE17F22.JPEG

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of Neurons in Macaque Area V4 During Memory-guided Visual Search. *Cerebral Cortex*, *11*(8), 761–772. https://doi.org/10.1093/CERCOR/11.8.761

Desimone, R., & Duncan, J. (2003). Neural Mechanisms of Selective Visual Attention. *Https://Doi.Org/10.1146/Annurev.Ne.18.030195.001205*, *18*, 193–222. https://doi.org/10.1146/ANNUREV.NE.18.030195.001205

Glorot, X., Bordes, A., & Bengio, Y. (2011). *Deep Sparse Rectifier Neural Networks* (pp. 315–323). JMLR Workshop and Conference Proceedings. https://proceedings.mlr.press/v15/glorot11a.html

Houtkamp, R., Spekreijse, H., & Roelfsema, P. R. (2003). A gradual spread of attention during mental curve tracing. *Perception & Psychophysics*, *65*(7), 1136–1144. https://doi.org/10.3758/BF03194840

Jeurissen, D., Self, M. W., & Roelfsema, P. R. (2016). Serial grouping of 2D-image regions with object-based attention in humans. *ELife*, *5*(JUN2016). https://doi.org/10.7554/ELIFE.14320

Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, *14*(2), 129–140. https://doi.org/10.3758/BF03198373

Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual Curve Tracing Properties. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 997–1022. https://doi.org/10.1037/0096-1523.17.4.997

Khayat, P. S., Pooresmaeili, A., & Roelfsema, P. R. (2009). Time course of attentional modulation in the frontal eye field during curve tracing. *Journal of Neurophysiology*, *101*(4), 1813–1822. https://doi.org/10.1152/JN.91050.2008/ASSET/IMAGES/LARGE/Z9K0040993900005.JPEG

Kim, M.-S., & Cave, K. R. (1995). Spatial Attention in Visual Search for Features and Feature Conjunctions. *Source: Psychological Science*, *6*(6), 376–380.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Liao, R., Xiong, Y., Fetaya, E., Zhang, L., Yoon, K., Pitkow, X., Urtasun, R., & Zemel, R. (2018). *Reviving and Improving Recurrent Back-Propagation*. https://github.com/

Lillicrap, T. P., & Santoro, A. (2019). Backpropagation through time and the brain. In *Current Opinion in Neurobiology* (Vol. 55, pp. 82–89). Elsevier Ltd. https://doi.org/10.1016/j.conb.2019.01.011

Linsley, D., Ashok, A. K., Govindarajan, L. N., Liu, R., & Serre, T. (2020). *Stable and expressive recurrent vision models*. https://github.com/c-rbp.

Linsley, D., Kim, J., Veerabadran, V., Windolf, C., & Serre, T. (2019). *Learning long-range spatial dependencies with horizontal gated recurrent units*.

Marić, M., & Domijan, D. (2019). Neural dynamics of spreading attentional labels in mental contour tracing. *Neural Networks*, *119*, 113–138. https://doi.org/10.1016/J.NEUNET.2019.07.016

Mély, D. A., Linsley, D., & Serre, T. (2018). Complementary surrounds explain diverse contextual phenomena across visual modalities. *Psychological Review*, *125*(5), 769–784. https://doi.org/10.1037/REV0000109

Mirpour, K., & Bisley, J. W. (2012). Anticipatory Remapping of Attentional Priority across the Entire Visual Field. *The Journal of Neuroscience*, *32*(46), 16449. https://doi.org/10.1523/JNEUROSCI.2008-12.2012

Moro, S. I., Tolboom, M., Khayat, P. S., & Roelfsema, P. R. (2010). Neuronal activity in the visual cortex reveals the temporal order of cognitive operations. *Journal of Neuroscience*, *30*(48), 16293–16303. https://doi.org/10.1523/JNEUROSCI.1256-10.2010

Pineda, F. J. (1987). Generalization of back-propagation to recurrent neural networks. *Physical Review Letters*, *59*(19), 2229–2232. https://doi.org/10.1103/PhysRevLett.59.2229

Pogodin, R., Mehta, Y., Lillicrap, T. P., & Latham, P. E. (2021). Towards Biologically Plausible Convolutional Networks. *Advances in Neural Information Processing Systems*, *17*, 13924–13936. https://arxiv.org/abs/2106.13031v2

Pooresmaeili, A., Poort, J., & Roelfsema, P. R. (2014). Simultaneous selection by object-based attention in visual and frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6467–6472. https://doi.org/10.1073/PNAS.1316181111/-/DCSUPPLEMENTAL

Pooresmaeili, A., Poort, J., Thiele, A., & Roelfsema, P. R. (2010). Separable Codes for Attention and Luminance Contrast in the Primary Visual Cortex. *Journal of Neuroscience*, *30*(38), 12701–12711. https://doi.org/10.1523/JNEUROSCI.1388-10.2010

Pooresmaeili, A., & Roelfsema, P. R. (2014). A growth-cone model for the spread of object-based attention during contour grouping. *Current Biology*, *24*(24), 2869–2877. https://doi.org/10.1016/j.cub.2014.10.007

Roelfsema, P. R. (2005). Elemental operations in vision. *Trends in Cognitive Sciences*, *9*(5), 226–233. https://doi.org/10.1016/j.tics.2005.03.012

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*, 203–227. https://doi.org/10.1146/ANNUREV.NEURO.29.051605.112939

Roelfsema, P. R., Bohte, S., & Spekreijse, H. (1999). *Algorithms for the detection of connectedness and their neural implementation*.

Roelfsema, P. R., & de Lange, F. P. (2016). Early Visual Cortex as a Multiscale Cognitive Blackboard. In *Annual review of vision science* (Vol. 2, Issue 1, pp. 131–151). https://doi.org/10.1146/annurev-vision-111815-114443

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, and Psychophysics*, *73*(8), 2542–2572. https://doi.org/10.3758/s13414-011-0200-0

Roelfsema, P. R., Khaya, P. S., & Spekreijse, H. (2003). Subtask sequencing in the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5467–5472. https://doi.org/10.1073/pnas.0431051100

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature 1998 395:6700*, *395*(6700), 376–381. https://doi.org/10.1038/26475

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research*, *40*(10–12), 1385–1411. https://doi.org/10.1016/S0042-6989(00)00004-3

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different Processing Phases for Features, Figures, and Selective Attention in the Primary Visual Cortex. *Neuron*, *56*(5), 785–792. https://doi.org/10.1016/J.NEURON.2007.10.006/ATTACHMENT/9D72C394-0631-4A87-B220-8A8887E0869E/MMC2.AVI

Sato, T., Murthy, A., Thompson, K. G., & Schall, J. D. (2001). Search efficiency but not response interference affects visual selection in frontal eye field. *Neuron*, *30*(2), 583–591. https://doi.org/10.1016/S0896-6273(01)00304-X

Schultz, W. (2002). Getting Formal with Dopamine and Reward. *Neuron*, *36*(2), 241–263. https://doi.org/10.1016/S0896-6273(02)00967-4

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Self, M. W., van Kerkoerle, T., Supèr, H., & Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Current Biology : CB*, *23*(21), 2121–2129. https://doi.org/10.1016/J.CUB.2013.09.013

Semedo, J. D., Jasper, A. I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C. K., Kohn, A., & Yu, B. M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nature Communications 2022 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41467-022-28552-w

Simard, P. Y., Ottaway, M. B., & Ballard, D. H. (1989). Analysis of Recurrent

Backpropagation. *UR Research*, 103–112. https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=5516

Steinmetz, N. A., & Moore, T. (2014). Eye movement preparation modulates neuronal responses in area V4 when dissociated from attentional demands. *Neuron*, *83*(2), 496–506. https://doi.org/10.1016/J.NEURON.2014.06.014

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction, Second Edition. In *The Lancet* (Vol. 258, Issue 6685). https://doi.org/10.1016/S0140-6736(51)92942-X

Ullman, S. (1984). Visual routines. *Cognition*, *18*(1–3), 97–159. https://doi.org/10.1016/0010-0277(84)90023-4

Van Kerkoerle, T., Self, M. W., & Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications 2017 8:1*, *8*(1), 1–14. https://doi.org/10.1038/ncomms13804

Vanrullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology*, *3*(1–2), 167–176. https://doi.org/10.2478/V10053-008-0022-3

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, *1*(4), 339–356. https://doi.org/10.1016/0893-6080(88)90007-X

Werbos, P. J. (1990). Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE*, *78*(10), 1550–1560. https://doi.org/10.1109/5.58337

Zhou, H., & Desimone, R. (2011). Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron*, *70*(6), 1205–1217. https://doi.org/10.1016/J.NEURON.2011.04.032

Zylberberg, A., Dehaene, S., Roelfsema, P. R., & Sigman, M. (2011). The human Turing machine: A neural framework for mental programs. In *Trends in Cognitive Sciences* (Vol. 15, Issue 7, pp. 293–300). https://doi.org/10.1016/j.tics.2011.05.007

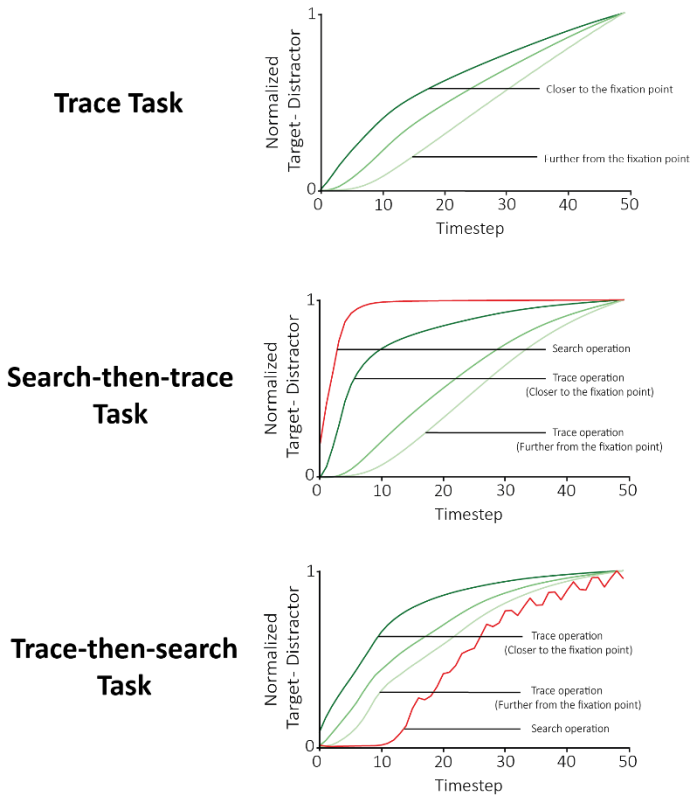

Fig. S1. **Time course of activity modulation in the output layer.** In the output layer, the enhancement of activity of units with RFs on the target curve followed the same time course as neurons in the lower layers. Hence, the time required for spreading activity in the visual layers predicts the reaction time. These results are compatible with the reaction time measurements in humans by Jolicoeur et al. (1986). They demonstrated that the reaction time increases in proportion to the length of the curve that needs to be traced.