Psychological Review

The Emergence of Polychronization and Feature Binding in a Spiking Neural Network Model of the Primate Ventral Visual System

Akihiro Eguchi, James B. Isbister, Nasir Ahmad, and Simon Stringer

Online First Publication, June 4, 2018. http://dx.doi.org/10.1037/rev0000103

American Psychological Association

Psychological Review

0033-295X/18/$12.00

http://dx.doi.org/10.1037/rev0000103

The Emergence of Polychronization and Feature Binding in a Spiking Neural Network Model of the Primate Ventral Visual System

Akihiro Eguchi, James B. Isbister, Nasir Ahmad, and Simon Stringer
Oxford University

We present a hierarchical neural network model, in which subpopulations of neurons develop fixed and regularly repeating temporal chains of spikes (polychronization), which respond specifically to randomized Poisson spike trains representing the input training images. The performance is improved by including top-down and lateral synaptic connections, as well as introducing multiple synaptic contacts between each pair of pre- and postsynaptic neurons, with different synaptic contacts having different axonal delays. Spike-timing-dependent plasticity thus allows the model to select the most effective axonal transmission delay between neurons. Furthermore, neurons representing the binding relationship between low-level and high-level visual features emerge through visually guided learning. This begins to provide a way forward to solving the classic feature binding problem in visual neuroscience and leads to a new hypothesis concerning how information about visual features at every spatial scale may be projected upward through successive neuronal layers. We name this hypothetical upward projection of information the "holographic principle."

Keywords: neural networks, polychronization, binding problem, STDP, spiking network

Many early neural network models of brain function assumed that neurons transmit information exclusively through modulation of their mean firing rates. These "rate-coded" models represented only the current average firing rate of each neuron and did not explicitly represent the timings of the action potentials or "spikes" emitted by cells. However, in modern literature, the precise timing of spikes has been proposed to strongly contribute to encoding in the brain (Akolkar et al., 2015; Fujii, Ito, Aihara, Ichinose, & Tsukada, 1996; Nikolic, Fries, & Singer, 2013). Consistent with this view, there is growing evidence from neurophysiological studies supporting the importance of spike-timing dynamics in the brain (Lindsey, Morris, Shannon, & Gerstein, 1997; Mao, Hamzei-Sichani, Aronov, Froemke, & Yuste, 2001; Prut et al., 1998; Softky, 1995). In the current study, we investigate the behavior of a biologically realistic hierarchical neural network model of the primate ventral visual system.

In particular, we explore how the network model develops during training stimulus representations in the form of fixed and regularly repeating temporal chains of spikes emitted by subpopulations of neurons even when the input images are represented by randomized Poisson spike trains. To elaborate further, visual stimuli are represented by specific subpopulations of input neurons with set firing rates, but where the spikes emitted by each neuron have randomized spike times set according to a Poisson probability distribution. However, even though the spike times of the input neurons are randomized, as activity is propagated upward through the layers of the network, we see the gradual emergence of regularly repeating spatiotemporal spike patterns in the output layer. This phenomenon is known as *polychronization*. The emergence of spatiotemporal spike chains should be contrasted with the notion of spike *synchronization*, in which a subpopulation of neurons emits their spikes at the same time. In the simulations reported in this article, we show that polychronization emerges naturally when the model incorporates distributions of nonzero axonal conduction delays of the order of a few milliseconds, which forces neurons to emit their spikes in spatiotemporal chains. After training the network on a set of visual stimuli, we show that different stimuli are represented by distinct spatiotemporal spike patterns in the output layer, which maintain their temporal structure across different presentations of the same stimulus with different input spike times. Such a subpopulation of neurons, which displays regularly repeating spatiotemporal spike patterns, is known as a polychronous group (PG). In order to facilitate this learning process in the simulations, we also explore a mechanism of synaptic delay selection with a biologically plausible learning mechanism: spike-timing-dependent plasticity (STDP).

Building on the above work demonstrating the emergence of polychronization, we also investigate a potential approach to solving the classic feature *binding problem* in visual neuroscience, which concerns how the brain represents the relationships between visual features within a scene. In particular, we are interested in

Akihiro Eguchi, James B. Isbister, Nasir Ahmad, and Simon Stringer, Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, Department of Experimental Psychology, Oxford University.

This research was supported financially by the Oxford Foundation for Theoretical Neuroscience and Artificial Intelligence. The foundation had no other role than providing financial support. We thank B. D. Evans and T. Minot for invaluable assistance and discussion related to the research. The ideas appearing in the manuscript were presented by Simon Stringer at the Microsoft Conference held at Microsoft Berlin, November 16, 2017.

Correspondence concerning this article should be addressed to Simon Stringer, Department of Experimental Psychology, Oxford University, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom. E-mail: simon.stringer@psy.ox.ac.uk

how the visual brain can learn to represent the rich hierarchical binding relations between lower and higher level features at every spatial scale across the visual field as discussed by Duncan and Humphreys (1989). It has been proposed that this type of representation cannot be achieved within a traditional rate-coded model, in which the times of spikes are not explicitly represented because of the phenomenon known as the "superposition catastrophe" (von der Malsburg, 1999). However, in our spiking neural network simulations, we show that the emergence of polychronization leads to a way in which the network may learn to represent these hierarchical binding relations. Specifically, we show the emergence of what we have called *binding neurons* embedded within the spatiotemporal spike chains. Binding neurons represent the binding relationships between low-level and high-level visual features. Such neurons were originally proposed by von der Malsburg (1999). However, it has not previously been shown how these neurons may develop naturally through a biologically plausible process of visually guided learning and self-organization of polychronous groups. In our simulations, we show the emergence of binding neurons that encode binding relationships between visual features across the entire visual field and at every spatial scale. These binding neurons, which develop automatically within the regularly repeating spatiotemporal spike chains during visual training, thus begin to provide a way forward to solving feature binding in primate vision.

Lastly, we show how our proposed mechanism for solving the feature binding problem automatically leads to the bottom-up (feedforward) projection of visual information about lower level visual features, and indeed visual features at every level, through successive neuronal layers to the highest (output) layer of the network. We refer to this as the *holographic principle*. This may be important if subsequent brain areas that guide behavior are only able to read out visual information from the highest stages of the visual system.

Temporal Coding and Polychronization

In the brain, neurons represent information and communicate with each other by pulses in their somatic membrane potential, called *action potentials* or *spikes*. The activity of a somatic spike propagates down the axon of the neuron, causing neurotransmitters to be released from multiple presynaptic axon terminals into their corresponding synaptic clefts. Binding of the neurotransmitters to the receptors of the postsynaptic dendrites causes a change in the electrical activity of the postsynaptic neurons, constituting a communication of information from the presynaptic neuron to the postsynaptic neuron. This neuron also spikes if the excitation of this postsynaptic neuron from its afferent synapses increases the membrane potential above its firing threshold potential. Raising the membrane potential of the postsynaptic neuron above the firing threshold generally requires the activation of afferent synapses within a brief temporal window, as the membrane potential naturally decays quickly back to a resting potential without further afferent excitatory activation.

The relative timings of the spikes emitted by a pair of pre- and postsynaptic neurons has also been shown to affect learning through spike-timing-dependent changes in synaptic efficacy (Bi & Poo, 1998; Markram, Lubke, Frotscher, & Sakmann, 1997), and hence how information and representations are stored and propagated in the network. If a presynaptic neuron fires in a short time period (up to tens of milliseconds) prior to the postsynaptic neuron firing, the synaptic efficacy increases. An increase in synaptic efficacy is known as long-term potentiation (LTP). If the presynaptic neuron instead fires in a short period of time following the firing of a postsynaptic neuron, the efficacy of the synapse is reduced. This reduction in synaptic efficacy is known as long-term depression (LTD). These forms of LTP and LTD, which depend on the relative timings of the pre- and postsynaptic neurons, are known as STDP. Compared with firing rate based synaptic learning rules employed in rate-coded models, an STDP learning rule can result in very different self-organization of the synaptic connectivity in the network when trained on visual scenes containing multiple objects (Evans & Stringer, 2012, 2013).

In a spiking neural network, individual neurons may operate as "coincidence detectors" (Abeles, 1991; Jeanson, 2011). That is, a postsynaptic neuron will fire if spikes from a number of presynaptic neurons arrive within a relatively brief time window of the order of a few milliseconds. This will be the case if the neuronal and synaptic time constants of the postsynaptic neuron are relatively brief, allowing for a fast decay in the cell membrane potential between incoming presynaptic spikes. In this situation, the presynaptic spikes must arrive close together in time in order to combine together to drive up the postsynaptic cell membrane potential to reach its firing threshold. A simple example of a coincidence-detecting neuron is shown in Figure 1a. In the figure, Neurons 1 and 2 represent low-level features such as horizontal and vertical bars, respectively, whereas Neuron 3 is a coincidence detecting neuron that represents a high-level feature or object such as the alphabetic letter "T." Neuron 3 only fires if the spikes emitted by Neurons 1 and 2 arrive at Neuron 3 close together in time. This means that the response of Neuron 3 is sensitive not only to which presynaptic neurons are firing but also to the precise timings of their spikes. As can be seen from this example, such a coincidence detecting neuron can provide a way of constructing higher level symbols through combination of elementary features, and do this in a way that utilizes temporal coding that depends on the timings of spikes.

Simulation studies have shown that if the synaptic connections within a large population of neurons have axonal transmission delays that are drawn from a random distribution of variable magnitudes, from say a few milliseconds to several tens of milliseconds, then groups of coincidence detecting cells emerge through STDP (Izhikevich, Gally, & Edelman, 2004). Furthermore, the network develops repeating temporal chains of spiking activity distributed across subgroups of coincidence detecting neurons, that is, neurons firing in a well-defined temporal sequence. This is referred to as *polychronization* (Izhikevich, 2006). Each subgroup of coincidence detecting neurons that comes together to form a regularly repeating temporal chain of activity is known as a *polychronous group* (PG). It has been hypothesized that each PG could represent a particular sensory (e.g., visual) stimulus such as the letter T or perhaps episodic memory (Izhikevich, 2006). Figure 1b illustrates an example in which a horizontal bar, a vertical bar, and a character T are represented by different PGs. In theory, polychronization in spiking networks can offer a dramatic increase in representational capacity compared with rate-coded models that do not exploit the timings of spikes (Izhikevich, 2006).

Paugam-Moisy, Martinez, and Bengio (2008) have recently examined how PGs selectively respond to artificial input patterns after training with STDP and shed light on the potential of utilizing PGs for real-life machine learning tasks such as handwritten digit recognition.

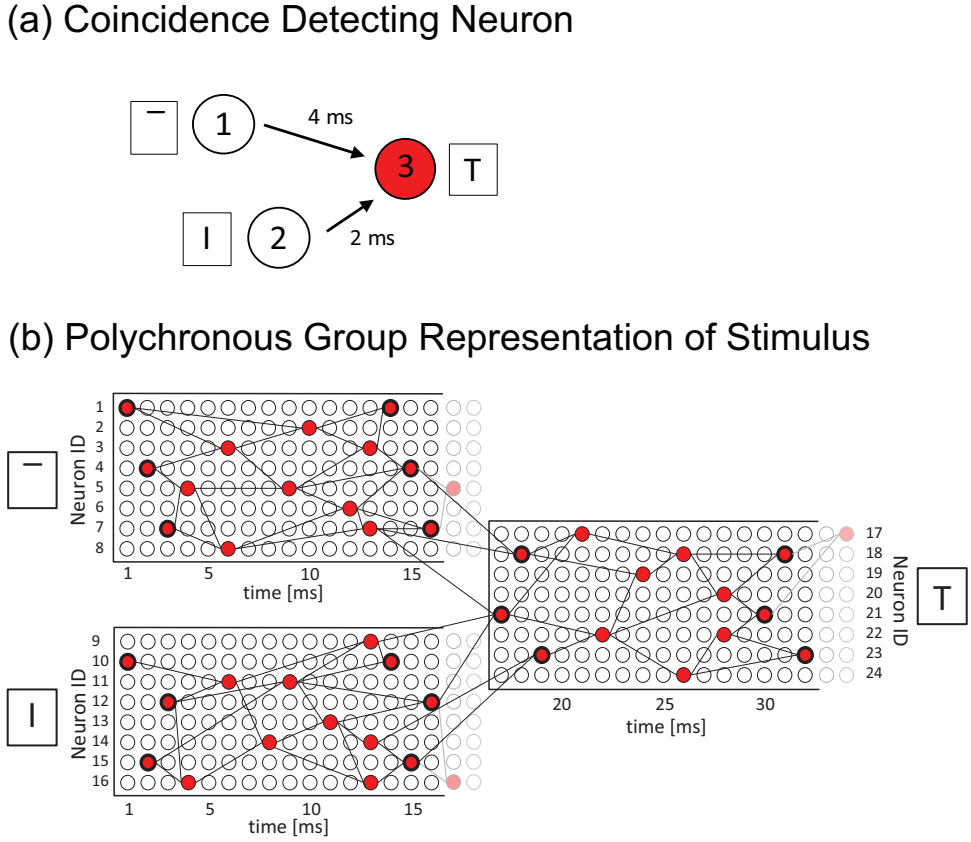

Figure 1. (a) Example of coincidence detecting neuron arrangement. Neurons 1 and 2 represent two different low-level features: a horizontal bar and a vertical bar, respectively. Neuron 3 is a coincidence detecting neuron that represents a high-level feature, namely, the alphabetic letter T. Neuron 3 only fires if the spikes emitted by Neurons 1 and 2 arrive at Neuron 3 close together in time. The action potentials of Neurons 1 and 2 propagate activity to Neuron 3 with delays of 4 ms and 2 ms, respectively. If the action potential of Neuron 1 occurs approximately 2 ms before the action potential of Neuron 2, their propagating activity will arrive simultaneously at Neuron 3 and cause it to spike. Neurons 1 and 2 represent the component vertical and horizontal bars comprising a letter T. In reality, the horizontal and vertical bars, as well as the letter T, would each be represented by a unique polychronous group (PG) of neurons. (b) Example of PG representation of stimulus. A horizontal bar is represented with a PG consisting of Neurons 1–8, a vertical bar is represented with a PG consisting of Neurons 9–16, and a character T is represented with a PG consisting of Neurons 17–24. The red circles represent neurons that are active at different times and form polychronous chains. The red circles toward the beginning and end of the time sequences that have thicker black boundaries represent the trigger neurons for chains of spiking neurons that represent a particular visual input (see Polychronous Group Counting for a mathematical description of such trigger neurons). See the online article for the color version of this figure.

However, the study carried out by these authors did not address three key issues as follows. First, the study carried out by Paugam-Moisy et al. used carefully ordered spike trains to represent input images, which is not biologically plausible. What would happen if the input spike trains contained much more random variation, as would be expected in the brain? Second, their model did not incorporate multiple synaptic connections with different randomized axonal transmission delays between each pair of pre- and postsynaptic neurons. This meant that the axonal transmission delay between any pair of neurons was fixed to a single value and could not be effectively selected from a number of alternatives by STDP learning. Consequently, the set of possible PGs that a neuron could participate in was limited before learning. Third, and perhaps most importantly, Paugam-Moisy et al.'s study did not investigate how feature binding representations, which explicitly encode the binding relations between low and high-level features, might develop through polychronization within a hierarchical model of visual processing. In the simulations presented in this article, we investigate each of these three issues in a hierarchical spiking neural network model of the primate ventral visual pathway, which is tasked with learning representations of the shapes of two-dimensional visual objects.

The Binding Problem and a Limitation of Rate Coding

Descriptions of the binding problem vary but generally address the same question: How does the visual system represent which

elementary features are bound together to form an object? For example, if the two letters T and L are seen together, how does the visual system represent which horizontal and vertical bars are part of which letter? In traditional hierarchical rate-coded visual processing models (e.g., Fukushima, 1980; Riesenhuber & Poggio, 1999; Wallis & Rolls, 1997), simple features (such as horizontal and vertical bars) are represented in the lower visual layers, whereas more complex features (such as letters) are represented in the higher visual layers. However, without a solution to the feature binding problem, there is no way of reading off which bars are part of which letters, and hence where the object's constituent components are in space.

The underlying weakness of rate coding is well illustrated in the classical example of Rosenblatt (1961), which was further explained by von der Malsburg (1999). As Figure 2 illustrates, the example supposes we have a neural network with four output neurons. Output Neurons A and B represent the triangle and square, respectively, invariant to retinal position (top or bottom). Output Neurons C and D are instead location specific, responding to both objects in either the top or bottom location, respectively. When a single object is presented to the network, the responses of the four neurons provide sufficient information to decode both the shape and position of the object. On the other hand, when both objects are presented together, each at a different location, all of the output neurons become highly active; it is no longer clear whether the triangle or the square is in the top retinal location. Thus, the coactivation results in a merging of representations and a loss of information that could have been used to divide the scene into its components. This breakdown is referred to as the *superposition catastrophe* (von der Malsburg, 1999). Similar problems were recently reported in a study modeling the development of border ownership representations in the early visual cortex, driven by top-down modulation from higher layers (Eguchi & Stringer, 2016). This rate-coded model produced neuron responses characterizing border ownership cells in V1. However, these representations catastrophically failed upon the presentation of multiple visual stimuli because of the inability of the rate-coded model to provide spatially selective top-down modulation.

In short, the crucial problem with rate coding is the lack of means to represent information regarding which specific low-level/elementary features have been combined to construct higher level features or objects. Moreover, binding of visual features must operate across the entire visual field and at all spatial scales within a visual scene. How features are bound together underpins how we segment a visual scene into objects and parts of objects, and thus how we make sense of the visual world.

Thus, solving the binding problem is essential to understanding the ability of the primate visual brain to make sense of complex visual scenes, and to developing a next generation of far more powerful computer vision systems with the ability to understand what they are looking at in the same way as the brain. Our simulation results suggest that binding is a much richer phenomenon than traditionally described by visual psychologists. Indeed, the binding mechanism proposed here is potentially so rich that it would be difficult to describe the process at a high psychological level; it requires a description at the neuronal level as presented in this article.

Background Theory, Research Questions, and Hypotheses

We investigate the behavior of a biologically realistic hierarchical neural network model of the primate ventral visual system that incorporates the following key aspects of cortical dynamics and architecture:

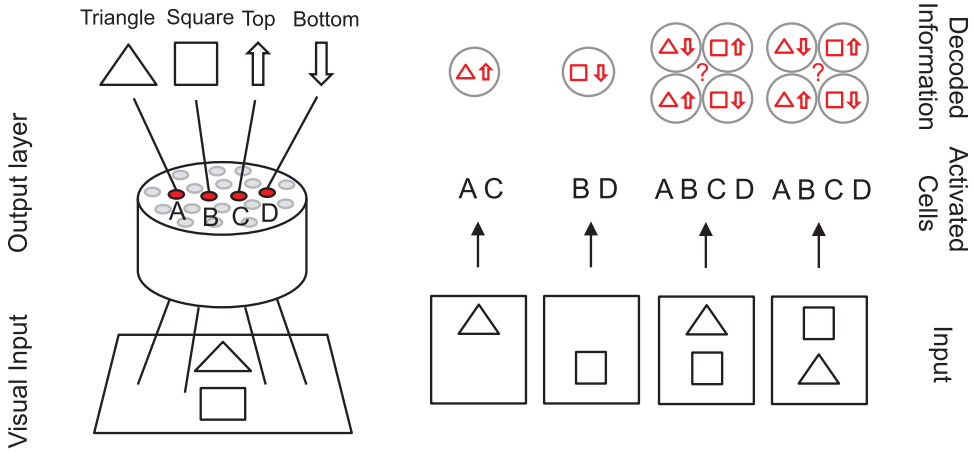

Figure 2. Rosenblatt's (1961) example of a binding problem in a rate-coded network. Left: The input from a visual scene is presented to a neuronal population including a set of four output Neurons A, B, C, and D. The firing rate responses from Neurons A to D, respectively, represent the presence of the following: a triangle, a square, an object in the "top" position, and an object in the "bottom" position. Right: The responses of output Neurons A to D when four different scenes are presented to the network. It can be seen that when only a single object is presented, the network can represent both the object and its position. However, when both objects are presented together, although the network is able to represent that both objects are present, it fails to represent the actual position of each object. See the online article for the color version of this figure.

- The model implements spiking neural dynamics in which the timings of action potentials or "spikes" are simulated explicitly.
- STDP is used to modify the synaptic connections during visually guided learning. If a spike from a presynaptic neuron arrives at a postsynaptic neuron just before the postsynaptic neuron emits a spike, then the synapse is strengthened (LTP). Otherwise, if the spike from the presynaptic neuron arrives at the postsynaptic neuron just after the postsynaptic neuron emits a spike, then the synapse is weakened (LTD).
- The network architecture incorporates bottom-up, top-down, and lateral synaptic connections reflecting the known architecture of the visual cortex.
- The synaptic connectivity between neurons incorporates distributions of axonal conduction delays of varying durations, from a few milliseconds to tens of milliseconds.
- In some simulations, network performance is enhanced by incorporating multiple synaptic connections between each pair of pre- and postsynaptic neurons, where these connections have different axonal transmission delays. This permits STDP to strengthen just one (or a subset) of these connections in order to effectively select the functional transmission delay between the two neurons (Deger, Helias, Rotter, & Diesmann, 2012; Fares & Stepanyants, 2009).

Using this underlying model architecture, the current study investigates the following hypotheses: emergence of polychronization, emergence of binding neurons, and "holographic principle" in the brain.

Emergence of Polychronization

During the initial period of visually guided learning, the network is trained on a set of visual stimuli that are encoded in the input layer by spiking neurons with *randomized* Poisson distributions of spikes. That is, the spike patterns representing the stimuli in the input layer have no special temporal structure, except that the average firing rates of the input neurons are set in accordance with the outputs of Gabor filters that simulate the responses of simple cells in visual area V1. Nevertheless, it was hypothesized that the initial period of visually guided learning with STDP would lead to the development of large numbers of regularly repeating PGs in the higher layers of the network, where individual PGs respond selectively to particular stimuli. Moreover, it was hypothesized that the emergence of large numbers of stimulus-selective PGs would increase the representational capacity of the network beyond that offered by a localist rate-coded representation in that, after training, the number of stimulus-specific PGs would be significantly greater than the number of single cells that responded selectively to a particular stimulus. The representational capacity is thus increased if the network encodes visual stimuli using temporal spike trains distributed over PGs of neurons rather than relying on the average firing rate responses of individual neurons.

Emergence of Binding Neurons

It was hypothesized that the emergence of PGs in the higher layers of the network during visually guided training with STDP could begin to provide a way forward to solving the classic feature *binding problem* in visual neuroscience. That is, how may the network learn to represent the hierarchical binding relations between low-level features such as horizontal or vertical bars and high-level features or objects such as the alphabetic letters T and L? Specifically, we hypothesized that some cells within PGs, which we will call *binding neurons*, will become tuned through STDP learning to respond if a neuron or subset of neurons representing a specific low-level feature is participating in driving neurons representing a particular high-level feature or object, which may be represented in a higher layer. In this case, the binding neuron carries measurable information that the low-level feature (such as a vertical bar at a particular retinal location) is part of the higher level feature or object (such as the letter T). Such binding neurons were originally proposed by von der Malsburg (1999), but without an explanation of how they might emerge naturally during visual development. We now propose, and demonstrate in the simulations presented later, that such binding neurons may develop automatically within the PGs that emerge during visually guided learning with STDP.

Here, we present a simple explanation for how such binding neurons may develop. An actual example is given in Figure 3a. Consider a linked set of three neurons at different stages of the ventral visual pathway: (1) Neuron 1 (in a lower visual layer) represents a low-level visual feature, (2) Neuron 2 (in a higher visual layer) represents a high-level visual feature, and (3) Neuron 3 is a hidden neuron within a local layer, say, the same layer as either Neuron 1 or 2, which may learn to become a binding neuron. Assume that there are the following three synaptic connections between these three neurons: (1) a connection from Neuron 1 to Neuron 2, (2) a connection from Neuron 1 to Neuron 3 (this could be either a lateral or bottom-up connection depending on which layer Neuron 3 is in), and (3) a connection from Neuron 2 to Neuron 3 (this could be either a lateral or top-down connection depending on which layer Neuron 3 is in).

Let us denote the axonal delay from neuron j to neuron i as $\Delta_{(i,j)}$. Then Neuron 1 is participating in driving Neuron 2 if, and only if, a spike emitted by Neuron 2 occurs approximately $\Delta_{(2,1)}$ after a spike emitted by Neuron 1.

If we have a set of three axonal delays such that

$$\Delta_{(3,1)} = \Delta_{(2,1)} + \Delta_{(3,2)}, \quad (1)$$

then the spikes from Neurons 1 and 2 will converge on Neuron 3 (near) simultaneously if, and only if, Neuron 1 is participating in driving Neuron 2.

It is assumed that the hidden Neuron 3 operates as a "coincidence detector," and fires only when the volley of spikes from Neurons 1 and 2 arrive (near) simultaneously. In this case, Neuron 3 will behave as a binding neuron. That is, Neuron 3 will fire if, and only if, Neuron 1 is participating in driving Neuron 2. In this case, STDP will further strengthen the connections from Neurons 1 and 2 onto the Binding Neuron 3.

It is important that an ideal binding neuron responds if, and only if, the neurons representing the low-level feature are actually participating in driving the neurons representing the high-level feature. The binding neuron should not respond if the neurons representing the low-level feature and the neurons representing the high-level feature just happen to be coactive, in which the former are not actually driving the latter. Such unrelated coactivation of low- and high-level features might occur, for example, because of

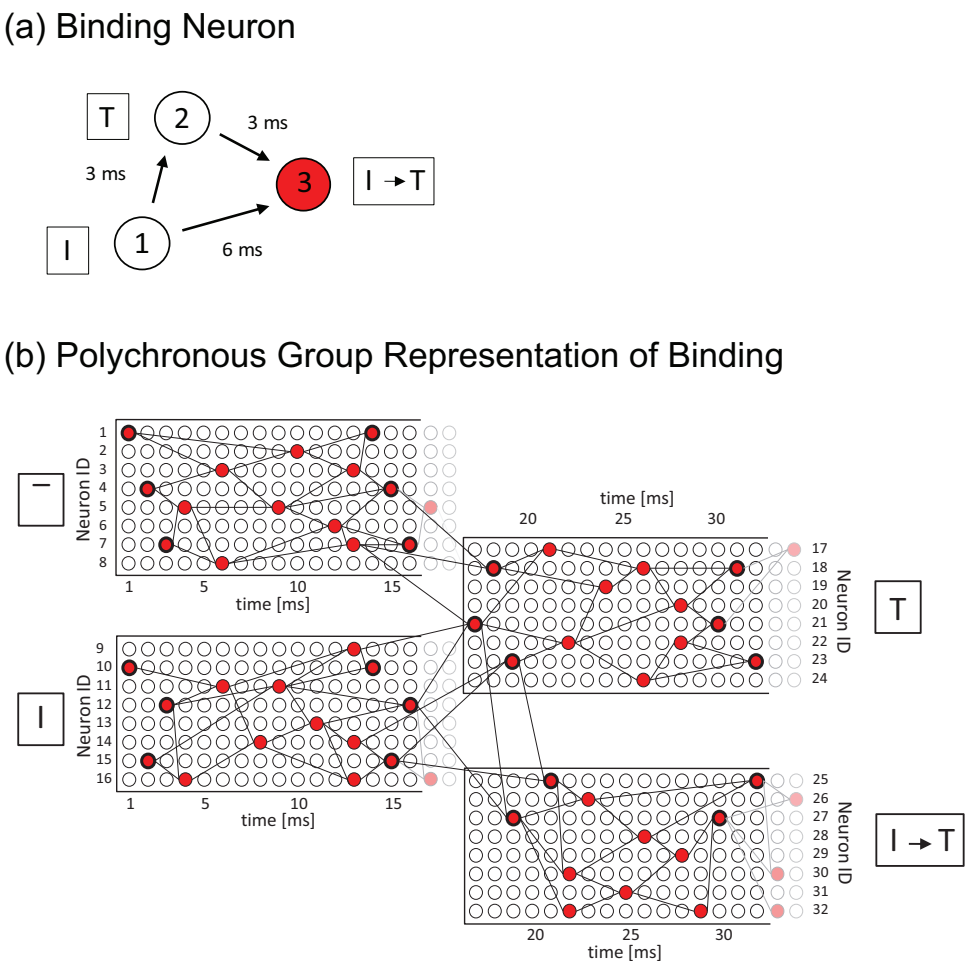

Figure 3. (a) Example of a hypothetical binding neuron. Consider a linked set of three neurons at different stages of the ventral visual pathway: Neuron 1 (in a lower visual layer) represents a low-level visual feature such as a vertical bar, Neuron 2 (in a higher visual layer) represents a higher level visual feature such as the letter T, and Neuron 3 is a binding neuron within a local layer, say, the same layer as either Neuron 1 or 2. Importantly, there are nonzero axonal transmission delays in the connections between these three neurons. In this example, the delays are as follows: The delay from Neuron 1 to Neuron 2 is 3 ms, the delay from Neuron 2 to Neuron 3 is 3 ms, and the delay from Neuron 1 to Neuron 3 is 6 ms. With the particular set of axonal delays given in this example, Neuron 3 will fire if, and only if, Neuron 1 is participating in driving Neuron 2. If Neuron 3 fires, this will encode the fact that the low-level feature (vertical bar) represented by Neuron 1 is part of the higher level feature (the letter T) represented by Neuron 2. (b) Polychronous group (PG) representation of binding. Here, we illustrate how a low-level feature such as a vertical bar may in fact be represented by its own temporal pattern of spikes distributed across a PG of neurons (shown bottom left), the high-level feature or object such as a letter T may also be represented by its own PG (shown top right), and these two PGs may drive a third PG representing the binding relationship between the vertical bar and the letter T (shown bottom right). This more complex scenario, in which the visual features and the binding relations between these features are represented by patterns of spiking activity across their own PGs, is likely to be what actually happens in the brain. The simple three-neuron circuit shown in 3a would then be a small part of the three corresponding PGs (representing a vertical bar, letter T, and binding relation between these two features) shown in 3b. See the online article for the color version of this figure.

the presence of multiple similar objects within a complex natural scene as explained earlier with Rosenblatt's (1961) example (see Figure 2). Suppose a T and L are presented together, then the neurons representing the horizontal bar of the T are coactive with the neurons representing the letter L, but the former are not driving the latter. Thus, the corresponding binding neuron, which would represent that the given horizontal bar was part of the L, should not fire. This kind of temporally specific response is characteristic of a PG, which the three neurons—Neurons 1, 2, and 3—described above comprise.

We hypothesize that with the inclusion of bottom-up, top-down, and lateral connections, there are a variety of possible local net-

work architectures that could self-organize through competitive learning to implement this, with the binding neurons being in any of the nearby lower or higher layers. Wherever the binding neuron is, its activation would still represent that a particular low-level feature is driving the representation of a specific high-level feature or object, and is therefore part of the object. A population of such binding neurons would specify which low-level features within a scene were part of which high-level features or objects, and this information could be read out directly by higher level neurons in the network.

This process could operate across the entire visual field and at every spatial scale within the visual field. Indeed, binding neurons would be expected to emerge throughout successive levels of the feature hierarchy within the network. A rich tapestry of binding neurons through the layers could help to provide a hierarchical structural description of a scene. This proposal may explain why the visual system needs extensive top-down connections between layers and lateral connections within layers in addition to bottom-up connections.

However, in the brain, it is in fact likely that a low-level feature such as a vertical bar and a high-level feature such as the letter T, as well as the binding relationship between these features, would each be represented by their own temporal pattern of spikes distributed across PGs of neurons. This is illustrated in Figure 3b. The binding relations are then represented by PGs (rather than individual neurons), which are replayed if, and only if, the low-level feature is part of the high-level feature. In this scenario, the simple three-neuron circuit shown in Figure 3a would be a small part of the three corresponding PGs (representing a vertical bar, letter T, and binding relation between these two features) shown in Figure 3b. However, in the simulations reported later we focus on identifying individual binding neurons that are part of three-neuron circuits of the general form shown in Figure 3a.

In this investigation, we specifically look at the emergence of such binding neurons among the learned neuronal representations of three simple visual shapes shown in Figure 4, which are presented to the network during visually guided training. We expected to find evidence for the kind of three-neuron binding relationships described earlier and illustrated in Figure 3a. These three-neuron PGs provide the simplest examples of how the network may learn to represent binding relationships in which specific low-level features are part of particular high-level features or objects.

Feedforward Projection of Information About Low-Level Visual Features to Higher Neuronal Layers

The earlier discussion of binding neurons leads directly to a new hypothesis concerning how information about visual features may be projected in a bottom-up (feedforward) manner through successive layers of the network. This might be a useful operation if the behavioral systems of the brain are limited to reading out visual information from the highest layers of the visual system. For example, it is generally conceived that simple visual features such as oriented edges and bars are represented in early cortical visual areas such as V1 and V2, whereas whole objects and faces are represented in higher visual areas. However, when we look at a visual scene, we are perceptually aware of visual features of varying levels of complexity and scale. Does this imply that information about low-level visual features is being projected directly upward through the visual system in some way that preserves the identity of these features, and at the same time also represents the image context of these features (i.e., binding relationships with higher level features)?

Figure 5a shows one simple way in which our network architecture might achieve this. The illustration is very similar to that shown in Figure 3a, except that the Binding Neuron 3 is now in the upper layer, that is, the same layer as Neuron 2, which represents the high-level feature T. Neuron 3 represents the fact that there is a vertical bar in some local region of the retina, which is part of the letter T. In this way, information about the low-level feature (vertical bar at a particular retinal position) along with its image context (the vertical bar is part of the letter T) has been projected up to the same layer as the representation of the high-level feature (the T). This is essentially the same binding mechanism discussed earlier, but where the binding neuron is situated in the same higher layer as the neuron representing the high-level feature. This mechanism for the bottom-up projection of information about low-level features to higher layers, where this information is modulated by local image context (i.e., the low-level feature is part of a particular high-level feature), may again operate up through successive neuronal layers, and hence across the entire visual field and at every spatial scale.

It is possible that the mechanism shown in Figure 5a could be repeated iteratively up through the layers. For example, Figure 5b shows an example in which information about the vertical bar is first projected up from the first layer to the second layer, where it is represented by Binding Neuron 3. Neuron 3 represents the fact

Figure 4. A set of three visual stimuli presented to the network: a circle, a heart, and a star.

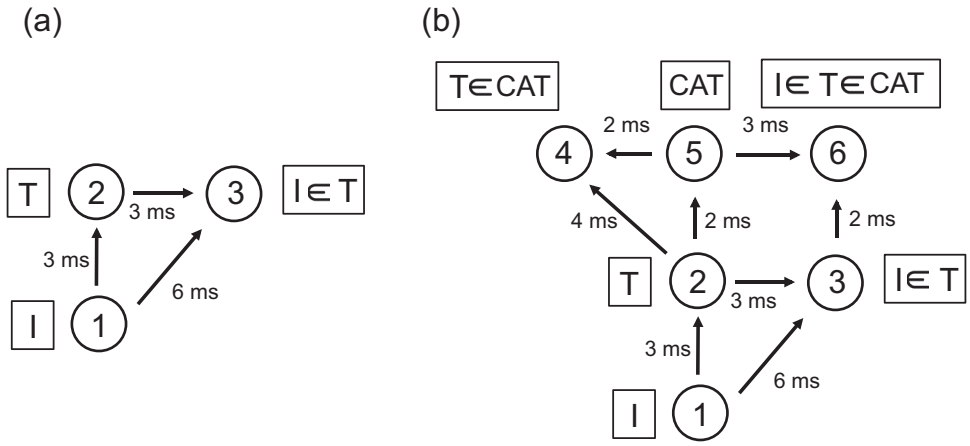

Figure 5. Illustrations of how the proposed binding mechanism may project information about low-level visual features such as a vertical bar up through successive layers of the network. The illustration shown in (a) is very similar to that shown in Figure 3a, except that the Binding Neuron 3 is now located in the upper layer, that is, the same layer as Neuron 2 that represents the high-level feature T. Neuron 3 represents the fact that there is a vertical bar in some local region of the retina, which is part of the letter T. In this way, information about the low-level feature (vertical bar at a particular retinal position) along with its image context (the vertical bar is part of the letter T) has been projected up to the same layer as the representation of the high-level feature (the T). (b) Shows how the mechanism illustrated in (a) could be repeated iteratively up through the layers. Now, a similar binding mechanism combines the output from Binding Neuron 3 with the output of Neuron 5 representing a cat, where these combined outputs drive Binding Neuron 6. Binding Neuron 6 then represents the fact that there is a vertical bar in a local region of the retina, which is part of the letter T, which, in turn, is part of the word CAT. In this case, the information about the lowest level feature (a vertical bar) is preserved in the highest layer of the network.

that there is a vertical bar in a local region of the retina, which is part of the letter T. Then, a similar binding mechanism combines the output from Binding Neuron 3 with the output of Neuron 5 representing a cat, where these combined outputs drive Binding Neuron 6. Binding Neuron 6 then represents the fact that there is a vertical bar in a local region of the retina, which is part of the letter T, which, in turn, is part, of the word CAT. In this case, the information about the lowest level feature is preserved in the highest layer of the network. Indeed, it is theoretically possible that a very large amount of information could be projected upward in this manner and preserved in the highest layers for readout by subsequent behavioral systems. We refer to this as a *holographic principle* for spiking network models of biological vision, because information about visual features at every level of complexity and scale may be preserved in the highest layers.

It is important to note that the Binding Neurons 3 and 6 in the highest layers of the two network architectures shown in Figures 5a and 5b represent the presence of a vertical bar in some local region of the retina that is explicitly part of a higher level feature (e.g., the letter T) or hierarchy of features (e.g., the letter T, which is part of the word CAT). Thus, these binding neurons do not simply respond to the presence of a vertical bar at some retinal location regardless of local image context (i.e., the higher level features/objects that the vertical bar is part of). So the high-level feature/object still needs to be presented to the network in order to elicit a response from these kinds of binding neuron in the upper layers. Thus, the holographic principle described here is consistent with neurophysiological observations that neurons in the later stages of the ventral visual pathway tend to respond to more complex visual forms than the simple oriented bars represented in early cortical stages such as V1 and V2.

Effect of Varying Key Model Parameters

The study also investigates the effects of the following architectural, neuronal, and synaptic parameters on the number of PGs and binding neurons that develop in the network during visually guided training:

- **Synaptic connectivity.** The investigation will explore the performance of the network with the following synaptic connectivities: (a) purely bottom-up, (b) bottom-up and top-down, (c) bottom-up and lateral, and (d) bottom-up, top-down, and lateral. We test which of these architectures best promotes the emergence of polychronization including the representation of visual stimuli by stimulus-specific PGs.
- **STDP time constant.** We hypothesize that a longer STDP time constant will lead to less temporal sensitivity to spike times in the network, which will lead to the model operating in a more rate-coded manner. This, in turn, may reduce the emergence of polychronization including the number of stimulus-specific PGs that develop.
- **Multiple synaptic connections between each pair of pre- and postsynaptic neurons.** Within a network with multiple synaptic contacts (each with a different axonal delay) between each pair of pre- and postsynaptic neurons, we hypothesize that STDP will effectively select which delays to strengthen. If STDP is able to selectively strengthen just one

(or a subset) of the connections, this should help to promote the emergence of polychronization. For example, if each pair of pre- and postsynaptic neurons has two synaptic contacts with quite different axonal delays, then it is expected that STDP will increase one connection but weaken the other. We hypothesize that this will, in turn, increase the number of PGs in the network with maximum information for a particular stimulus.

Model and Performance Measures

Model

Network architecture. The neural network model investigated is shown in Figure 6 and simulates successive neuronal stages of processing along the primate ventral visual pathway. This model was simulated using the SPIKE simulator (see Appendix for link). Specifically, the model is comprised of four layers of excitatory pyramidal neurons, which may be thought of as representing cortical visual areas V2, V4, posterior inferior temporal cortex, and anterior inferior temporal cortex. There are modifiable bottom-up (feedforward) and top-down (feedback) synaptic connections between excitatory pyramidal neurons in successive layers as well as modifiable lateral synapses between excitatory pyramidal neurons within each layer. Some simulations explore the importance of the top-down and lateral connections for polychronization and feature binding by comparing model performance with and without them. Within each layer, there are also inhibitory interneurons with nonplastic lateral synaptic connections to and from the excitatory neurons to produce competition between the excitatory neurons. For all presented simulations, we used $64 \times 64 = 4{,}096$ excitatory neurons and $32 \times 32 = 1{,}024$ inhibitory neurons in each layer, with a fixed number of sparsely distributed topologically organized connections. Table 1a shows the different numbers of afferent connections onto each postsynaptic neuron, as well as the fan-in radius of these connections, for the different types of excitatory-excitatory, excitatory-inhibitory, and inhibitory-excitatory connections between and within the four neuronal layers. The models are developed using the GPGPU based Spiking Neural Network Spike! (See Appendix link).

Differential equations. As originally described in Evans and Stringer (2012), each neuron is based upon the standard conductance-based leaky integrate and fire (LIF) model, whereas the equations for STDP at the Excitatory-Excitatory ($E \rightarrow E$) synapses are adapted from Perrinet, Delorme, Samuelides, and Thorpe (2001). Neuron and synapse constants were chosen to be as biologically realistic as possible based upon the available neurophysiological literature (see Table 1 for a full list).

Cell equations. The neuron's membrane potential is updated according to Equation 2:

$$\tau_m^{\gamma} \frac{dV_i(t)}{dt} = V_0^{\gamma} - V_i(t) + R^{\gamma} I_i(t) \tag{2}$$

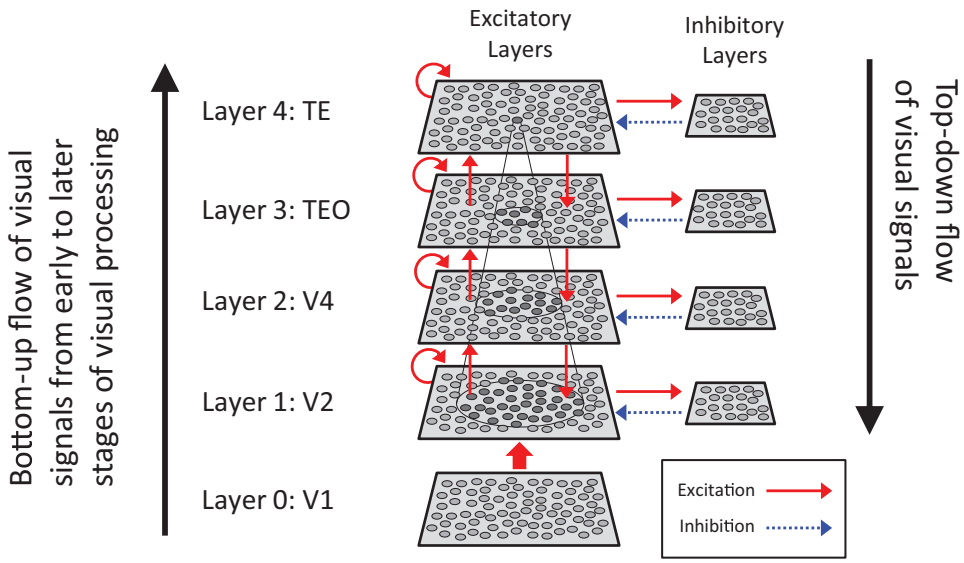

Figure 6. Schematic of the four layer neural network architecture investigated. The model represents successive neuronal stages of processing along the primate ventral visual pathway. The model is comprised of five layers of excitatory pyramidal neurons, which may be thought of as representing cortical visual areas V1, V2, V4, posterior inferior temporal cortex (TEO) and anterior inferior temporal cortex (TE). The layer 0 reflects the output of the Gabor filters of the visual input presented to the network, with an imposed Poisson spike rate of neurons in the layer. These neurons establish only feed-forward connection to the layer 1. Each of the following layers of the model (layer 1–4) consists of $64 \times 64 = 4096$ excitatory neurons and $32 \times 32 = 1024$ inhibitory neurons. Excitatory modifiable connections (red) include bottom-up (feedforward) and top-down (feedback) connections between excitatory pyramidal neurons in successive layers, and lateral connections between excitatory pyramidal neurons within the same layer (shown by the curved red arrows). Each layer of excitatory pyramidal neurons is connected to a population of inhibitory neurons which implement competition between the excitatory neurons in that layer. See the online article for the color version of this figure.

Table 1
Parameters and those Values Used in the Models

Parameter names	Layer 1	Layer 2	Layer 3	Layer 4	
Network parameters					
Number of excit. neurons within each layer	64 × 64	64 × 64	64 × 64	64 × 64	
Number of inhib. neurons within each layer	32 × 32	32 × 32	32 × 32	32 × 32	
Number of FF afferent excit. connections per excit. neuron (*EfE*)	30	100	100	100	
Fan-in radius for FF afferent excit. connections to each excit. neuron (*EfE*)	1.0	8.0	12.0	16.0	
Number of FB afferent excit. connections per excit. neuron (*EbE*)	{0, 10}	{0, 10}	{0, 10}	—	
Fan-in radius for FB afferent excit. connections to each excit. neuron (*EbE*)	8.0	8.0	8.0	—	
Number of LAT afferent excit. connections per excit. neuron (*ElE*)	{0, 10}	{0, 10}	{0, 10}	{0, 10}	
Fan-in radius for LAT afferent excit. connections to each excit. neuron (*ElE*)	4.0	4.0	4.0	4.0	
Number of LAT afferent excit. connections per inhib. neuron (*ElI*)	30	30	30	30	
Fan-in radius for LAT afferent excit. connections to each inhib. neuron (*ElI*)	1.0	1.0	1.0	1.0	
Number of LAT afferent inhib. connections per excit. neuron (*IlE*)	30	30	30	30	
Fan-in radius for LAT afferent inhib. connections to each excit. neuron (*IlE*)	8.0	8.0	8.0	8.0	
Parameters for Gabor filtering of visual images					
Phase shift (Ψ)	0, π				
Wavelength (λ)	2				
Orientation (θ)	0, $\pi/4$, $\pi/2$, $3\pi/4$				
Spatial bandwidth (b)	1.5 octaves				
Aspect ratio (γ)	.5				
Cellular parameters					
Excit. cell somatic capacitance (C_m^E) and inhib. cell somatic capacitance (C_m^I)	500 pF, 214 pF				a
Excit. cell somatic leakage conductance (g_0^E) and inhib. cell somatic leakage conductance (g_0^I)	25 nS, 18 nS				a
Excit. cell membrane time constant (τ_m^E) and inhib. cell membrane time constant (τ_m^I)	20 ms, 12 ms				a
Excit. cell resting potential (V_0^E) and inhib. cell resting potential (V_0^I)	−74 mV, −82 mV				a
Excit. firing threshold potential (Θ^E) and inhib. firing threshold potential (Θ^I)	−53 mv, −53 mV				a
Excit. after-spike hyperpolarization potential (V_H^E) and inhib. after-spike hyperpolarization potential (τ_R)	−57 mV, −58 mV				a
Absolute refractory period (τ_R)	2 ms				a
Synaptic parameters					
Synaptic neurotransmitter concentration (α_C) and Proportion of unblocked N-Methyl-D-aspartic acid (NMDA) receptors (α_D)	.5				b
Presynaptic STDP time constant (τ_C) and Postsynaptic STDP time constant (τ_D)	{5, 25, 125} ms				b
Synaptic learning rate (ρ)	.1				b
Range of synaptic conductance delay	[.1, 10.0] ms				b
Synaptic conductance scaling factor for FF excitatory connections from Gabor filters to Layer 1 excit. cells ($\lambda^{GfE} \cdot \Delta g^{GfE}$)	[0, .4] nS				c
Synaptic conductance scaling factor for FF excit. connections to excit. cells in layers 2, 3 or 4 ($\lambda^{EfE} \cdot \Delta g^{EfE}$)	[0, 1.6] nS				c
Synaptic conductance scaling factor for FB excit. connections to excit. cells in layers 1, 2 or 3 ($\lambda^{EbE} \cdot \Delta g^{EbE}$)	[0, 1.6] nS				c
Synaptic conductance scaling factor for LAT excit. connections to excit. cells in layers 1, 2, 3 or 4 ($\lambda^{ElE} \cdot \Delta g^{ElE}$)	[0, 1.6] nS				c
Synaptic conductance scaling factor for LAT connections from excit. cells to inhib. cells in layers 1, 2, 3 or 4 ($\lambda^{EII} \cdot \Delta g^{EII}$)	40 nS				c
Synaptic conductance scaling factor for LAT connections from inhib. cells to excit. cells in layers 1, 2, 3 or 4 ($\lambda^{IIE} \cdot \Delta g^{IIE}$)	80 nS				c
Excitatory reversal potential ($\hat{V}^E$)	0 mV				a
Inhibitory reversal potential ($\hat{V}^I$)	−70 mV				a
Synaptic time constant for all FF, FB, and LAT connections from Gabor filters and excit. cells to excit. cells (τ_{GfE}, τ_{EfE}, τ_{EbE}, τ_{ElE})	150 ms				b
Synaptic time constant for LAT connections from excit. cells to inhib. cells (τ_{EII})	2 ms				a
Synaptic time constant for LAT connections from inhib. cells to excit. cells (τ_{IIE})	5 ms				a
Parameters for numerical simulation by forward Euler time stepping scheme					
Numerical step size (Δt)	.02 ms				

Note. FF: feedforward; FB: feedback; LAT: lateral; STDP = spiketiming-dependent plasticity. Most integrate and fire parameters were taken from Troyer et al. (1998; derived originally from McCormick et al., 1985), as indicated by the "a" symbol. Plasticity parameters (denoted by the "b" symbols) are taken from Perrinet et al. (2001). Parameters marked with the asterisk (c) were tuned for the reported simulations.

The cell membrane potential for a given Neuron $V_i(t)$, indexed by i, is driven up by current from excitatory conductance-based synapses, and down toward the inhibitory reversal potential by current from inhibitory conductance-based synapses. Neurons decay back to their resting state over a time course determined by the properties of its membrane. Here, τ_m represents the membrane time constant, defined as $\tau_m = C_m/g_0$, where C_m is the membrane capacitance, g_0 is the membrane leakage conductance, and R is the membrane resistance ($R = 1/g_0$). V_0 denotes the resting potential of the cell. Class-specific values (excitatory and inhibitory) are indexed by γ for the neuron parameters. $I_i(t)$ represents the total current input from the afferent synapses (described in Equation 3).

The total synaptic current injected into a neuron is given by the sum of the conductances of all afferent synapses (excitatory and inhibitory), multiplied by the difference between the synapse class reversal potential, $\hat{V}^{\gamma}$, and neuron membrane potential, $V_i(t)$. Excitatory and inhibitory synapses have positive and negative conductances, respectively. The conductance of a given synapse is given by g_{ij} where j and i are the indices of the pre- and postsynaptic neurons, respectively:

$$I_i(t) = \sum_{\gamma} \sum_{j} g_{ij}(t)(\hat{V}^{\gamma} - V_i(t)) \tag{3}$$

Synaptic conductance equations. The synaptic conductance of a particular synapse, $g_{ij}(t)$, is governed by a decay term τ_g and a Dirac delta function (Equation 5) when spikes arrive from the presynaptic neuron j as follows:

$$\frac{dg_{ij}(t)}{dt} = -\frac{g_{ij}(t)}{\tau_g} + \lambda \Delta g_{ij}(t) \sum_{l} \delta(t - \Delta t_{ij} - t_j^l) \tag{4}$$

The conduction delay for a particular synapse is denoted by Δt_{ij}, which ranges from 0.1 to 10.0 ms, and each presynaptic neuron spike is indexed by l. A biological scaling constant, λ, has been introduced to scale the synaptic efficacy, Δg_{ij}, which lies between unity and zero. The Dirac delta function is defined as follows:

$$\delta(x) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where,} \quad \int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{5}$$

Synaptic learning equations. The following differential equations describe the STDP occurring at each modifiable *Excitatory - Excitatory* ($E \rightarrow E$) synapse. That is, these kinds of modifiable synapses occur at all of the bottom-up, top-down, and lateral connections from excitatory cells to excitatory cells throughout Layers 1 to 4.

Here, i labels the postsynaptic neuron. The recent presynaptic activity, $C_{ij}(t)$, is modeled by Equation 6, which may be interpreted as the concentration of neurotransmitter (glutamate) released into the synaptic cleft (Perrinet et al., 2001) and is bounded by [0, 1] for $0 \leq \alpha_C \leq 1$:

$$\frac{dC_{ij}(t)}{dt} = -\frac{C_{ij}(t)}{\tau_C} + \alpha_C(1 - C_{ij}(t)) \sum_{l} \delta(t - \Delta t_{ij} - t_j^l) \tag{6}$$

$C_{ij}(t)$ is governed by a decay term τ_C and is driven up by presynaptic spikes according to the model parameter α_C. The inclusion of the axonal transmission delay Δt_{ij} from presynaptic neuron j to postsynaptic neuron i in Equation 6 means that the variable $C_{ij}(t)$ is driven up at the time the spike from presynaptic neuron j arrives at the postsynaptic neuron i, rather than the time of emission of the spike from the presynaptic cell.

The recent postsynaptic activity, $D_i(t)$, is modeled by Equation 7 and may be interpreted as the proportion of NMDA receptors unblocked by recent depolarization from back-propagated action potentials (Perrinet et al., 2001):

$$\frac{dD_i(t)}{dt} = -\frac{D_i(t)}{\tau_D} + \alpha_D(1 - D_i(t)) \sum_{k} \delta(t - t_i^k) \tag{7}$$

$D_i(t)$ is governed by decay term τ_D and is driven up by postsynaptic spikes according to the model parameter α_D. Postsynaptic neuron spikes are indexed by k. Unlike with the conduction of action potentials toward the synapse, there is no conduction delay associated with D_i, because the cell body is assumed to be arbitrarily close to the receiving synapses and the effects of a postsynaptic spike are assumed to have an equal impact on the neuron's own afferent synapses.

The strength of the synaptic weight, $\Delta g_{ij}(t)$, is modified according to Equation 8, which is governed by the time course variable τ_{Δ_g}:

$$\begin{aligned} \tau_{\Delta_g} \frac{d\Delta g_{ij}(t)}{dt} = \; & \rho[(1 - \Delta g_{ij}(t))C_{ij}(t) \sum_{k} \delta(t - t_i^k) \\ & - \Delta g_{ij}(t) D_i(t) \sum_{l} \delta(t - \Delta t_{ij} - t_j^l)] \end{aligned} \tag{8}$$

Note that the postsynaptic spike train (indexed by k) is now associated with the presynaptic state variable (C), and vice versa. If C is high (because of recent presynaptic spikes having arrived at the postsynaptic neuron) at the time of a postsynaptic spike, then the synaptic weight is increased (LTP), whereas if D is high (from recent postsynaptic spikes) at the time of a presynaptic spike arriving at the postsynaptic neuron, then the weight is decreased (LTD). As noted, the inclusion of the axonal transmission delay Δt_{ij} in Equation 6 means that the variable $C_{ij}(t)$ is driven up at the time the spike from presynaptic neuron j actually arrives at the postsynaptic neuron i. Consequently, this form of STDP learning depends directly on the times that spikes from a presynaptic neuron arrive at a postsynaptic neuron rather than the times that the spikes were originally emitted by the presynaptic neuron.

The weight updates are also multiplicative, meaning that the amount of potentiation decreases as the synapse strengthens, as has been found experimentally (Bi & Poo, 1998). Theoretically, this weight-dependent potentiation yields a normal distribution of synaptic efficacies rather than pushing each weight to one extreme or the other (van Rossum, Bi, & Turrigiano, 2000), as would be the case with an additive form of STDP.

Numerical scheme. The differential equations described earlier are converted to finite difference equations and simulated using the forward Euler numerical scheme with a time step $\Delta t = 0.02$ ms. In the finite difference equations, the Dirac delta function has been replaced by the discrete approximation $S(x)$, as defined in (Amit & Brunel, 1997). Finally, in the original description, the change in synaptic weight (Equation 8) was instantaneous, and so $\Delta t/\gamma_{\Delta_g}$ is defined to be a learning rate constant, ρ, in the corresponding finite difference equation.

Training and stimuli. Before the visual images are presented to the first excitatory layer (Layer 1), they are preprocessed by a set of Gabor filters, which accord with the general tuning profiles of simple cells in V1 (Cumming & Parker, 1999; Jones & Palmer,

1987; Lades et al., 1993). The filters provide a unique pattern of filter outputs for each transform of each visual object, which is passed through to the first layer of the network. These filters are known to provide a good fit to the firing properties of V1 simple cells, which respond to local oriented bars and edges within the visual field (Cumming & Parker, 1999; Jones & Palmer, 1987). The input filters used are computed by the following equations:

$$g(x, y, \lambda, \theta, \psi, b, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \tag{9}$$

with the following definitions:

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \\ \sigma &= \frac{\lambda(2^b + 1)}{\pi(2^b - 1)}\sqrt{\frac{\ln 2}{2}} \end{aligned} \tag{10}$$

where x and y specify the position of a light impulse in the visual field (Petkov & Kruizinga, 1997). The parameter λ is the wavelength ($1/\lambda$ is the spatial frequency), σ controls number of such periods inside the Gaussian window based on λ and spatial bandwidth b, θ defines the orientation of the feature, ψ defines the phase, and γ sets the aspect ratio that determines the shape of the receptive field. In the experiments in this article, an array of Gabor filters is generated at each of 128 × 128 retinal locations with the parameters given in Table 1.

The outputs of the Gabor filters are used as the basis to generate Poisson spike trains as follows:

$$\begin{aligned} &P\{\text{input } cell(x, y, f) \text{ spikes at } t\} \\ &= g(x, y, f) \cdot max\_rate\_scaling\_factor \cdot \Delta t \end{aligned} \tag{11}$$

where f is the index of a filter used for the simulation and *max_rate_scaling_factor* is the maximum input neuron firing rate (set to 100 in the current simulation studies). The outputs of the Gabor filters coded in Poisson spike trains are enacted by the Layer 0 (Gabor Filter) cells which propagate activity to the Layer 1 excitatory neurons of the network according to the synaptic connectivity given in Table 1. That is, each Layer 1 neuron receives connections from 30 randomly chosen Gabor filters localized within a topologically corresponding region of the retina. These distributions are defined by a radius shown in Table 1.

Performance Measures

Information analysis of average firing rate responses of single cells. Information theory is used to quantify how selective the average firing rate responses of individual neurons are for members of a particular stimulus category. If a neuron responds invariantly to the members of a particular stimulus category but not to members of other stimulus categories, then the neuron carries a maximum amount of information about the presence of its preferred stimulus category.

We apply information theory to the average firing rate responses of individual neurons in the network in order to be able to compare the information conveyed by the firing rates of neurons with the information conveyed by the temporal spike patterns emitted by PGs (described later in Information Analysis of Temporal Spike Patterns Emitted by Polychronous Groups). In this way, we are able to demonstrate the very large increase in representational capacity that is possible using the temporal spike time coding available with the emergence of polychronization.

We have previously used information theory to quantify the performance of single neurons tasked with learning a translation invariant response (across multiple retinal locations) to specific visual stimuli (Eguchi, Mender, Evans, Humphreys, & Stringer, 2015). If the responses r of a neuron carry a high level of information about the presence of a particular stimulus s across different retinal locations, then this implies that the neuron will respond selectively to the presence of that stimulus regardless of where the stimulus is presented on the retina.

In this study, we do not explicitly introduce transforms of the visual inputs such as translation or rotation. However, because the input neural spike trains are generated based on Poisson distributions, there is a significant degree of stochasticity involved. This means that the exact timings of the input neuron spikes are different at each run. Therefore, in the current simulation study, different presentations of the same visual input to the network are considered as the "transforms" of the same stimulus category.

The amount of stimulus specific information that a specific cell carries is calculated using the following formula, with details given by Rolls and Milward (2000):

$$I(s, \vec{R}) = \sum_{r \in \vec{R}} P(r|s) log_2 \frac{P(r|s)}{P(r)} \tag{12}$$

Here, s is a particular stimulus, r is the response of a cell to a single stimulus, and $\vec{R}$ is the set of responses of a cell to the set of stimuli.

The maximum information that an ideally developed cell could carry is given by the following formula:

$$\text{Maximum cell information} = log_2(n) \text{ bits}, \tag{13}$$

where n is a number of different stimulus categories.

Information analysis of temporal spike patterns emitted by polychronous groups. We also apply information theory to quantify the amount of information conveyed by the temporal patterns of spikes emitted by PGs. Spike train data consists of time-ordered sequences of spikes. It has been proposed that, in the brain, the temporal spike patterns emitted by PGs may be utilized to encode larger amounts of information than codes relying solely on the average firing rates of neurons.

However, to simplify the analysis, in the simulations, we applied information theory to the analysis of PGs containing only two spikes emitted by a pair of neurons. In the simplest scenario involving only two neurons, A and B, with interspike delay k, the PG episode can be represented using the notation $A[k]B$ (Diekman, Dasgupta, Nair, & Unnikrishnan, 2014). By applying the analytical technique described in this section to the simulations reported later, we are able to demonstrate the emergence of frequently repeating PG episodes of the form $A[k]B$ that are specific to a specific stimulus category. A number of these two-neuron interactions could in principle chain together to form longer, more complex multineuron PGs.

The same information analysis technique described earlier is applied to frequently occurring spike-pair PGs of the form $A[k]B$ to investigate whether the network is able to represent different visual stimulus categories using this form of temporal coding. Based on the spike trains recorded during many stimulus presentations to the

network, we compute the probabilities that a given spike-pair $A[k]B$ will occur in response to the presentation of each of the stimulus categories s. These probabilities are based on the frequency of occurrence of the spike-pair $A[k]B$ across multiple transforms of each stimulus s, that is, across multiple presentations (transforms) of the each stimulus with different stochastic (Poisson) input representations. From these frequency distributions, we construct the following probability table for each stimulus category s:

$$ProbTable\ (i, j, d) = P\{(\text{Presynaptic cell } j \text{ spikes at time } t - d) \mid (\text{Postsynaptic cell } i \text{ spikes at time } t)\} \quad (14)$$

where i and j are the indices of two neurons under consideration, t is the time at which the cell i emits a spike, and d is the time interval that neuron i emits a spike after neuron j. We consider values of d within the range of [0, 10 ms], in which this time interval is divided into 10 equal bins of 1 ms. It is important to note that the probability table is constructed purely based on the actual spike trains emitted by neurons and does not take into account the actual synaptic connectivity between the neurons. This means that this technique highlights the correlations in spike times emitted by the cells involved but does not necessarily reflect actual synaptic connections. Given this method of analysis, there are potentially 167,772,160 ($nCells * nCells * maxDelay$) distinct spike-pair PG representations the output neuronal layer can hold. This is 40,960 ($nCells * maxDelay$) times more than the case of a localist rate-coded neuronal representation.

In applying the information analysis methodology to analyzing the information carried by spike-pair PGs, we regard the probability table given by Equation 14 as $\tilde{R}$, the set of responses to the set of stimuli, used in Equation 12. Thus, Equation 12 may now be used to compute the information carried by spike-pair PGs about the presence of particular stimulus categories s. With this technique, we can quantify how selective such temporal spike-pair PGs are for members of a particular stimulus category. In other words, if a particular spike-pair PG responds invariantly to the members of a particular stimulus category s but not to the members of other stimulus categories, then the spike-pair PG would carry maximum information about the presence of its preferred stimulus category.

Polychronous group counting. A key diagnostic in the simulations reported later is to identify and count the PGs that have emerged in the network after visually guided training.

As discussed in the introduction of this article, a PG is defined as the set of neurons that support the associated time-locked spike pattern. More formally, Martinez and Paugam-Moisy (2009) defined that

> an s-triggered polychronous group refers to the set of neurons that can be activated by a chain reaction whenever the triggers N_k ($1 \leq k \leq s$) fire according to the timing pattern t_k ($1 \leq k \leq s$). The PG is denoted by $N_1 - N_2 - \ldots - N_s$ ($t_1, \ldots, t_s$), where the firing times t_k are listed in the same order as the corresponding triggers N_k. (Martinez & Paugam-Moisy, 2009, p. 26)

We adopted the algorithm used by Izhikevich (2006) and modified it to be applicable for our conductance based LIF neural network model. The basic algorithm is as follows: (a) identify a set of potential triggers consisting of s neurons with specific spike timings (e.g., $N_1 - N_2 - \ldots - N_s$ [$t_1, \ldots, t_s$]), and (b) find PGs by simulating the propagation of activity from activation of this set of triggers.

More specifically, the algorithm first finds all combinations of a given number of s neurons (in our case, $s = 3$) that have at least one postsynaptic cell in common. For each such tuple of neurons, it then looks for the relative spike timings, based on synaptic delay, that can excite the common postsynaptic neuron maximally and enough for it fire. If such neurons exist, then the tuple becomes a trigger set. The algorithm then simulates the firing of the triggers with the identified spike timings and records the propagation of neural activity through the network until it decays to zero. In order to truncate the possible cyclic PGs, an upper limit is set for the time span of a PG and the number of neurons recorded.

Simulations

In the current simulation study, the network was trained and tested on the abstract visual objects shown in Figure 4. The shapes are a circle, a heart, and a star, which are colored black and presented against a 128 × 128 light gray background. Each simulation begins with an initial period of visual training. During each training epoch, each of the three object shapes was presented for 2 s to the network. As explained in the model description, the images are convolved with Gabor filters (Equation 9) that mimic the responses of edge detecting V1 simple cells. The stochastically generated Poisson spikes (Equation 11) are then imposed upon Layer 0 and are then propagated to the first layer (Layer 1) of the network, and thence up through successive Layers 2 to 4. During this, the synaptic connections from the Gabor filters to Layer 1 excitatory neurons, as well as the bottom-up, top-down, and lateral connections between excitatory neurons across all four layers of the network, were modified using the STDP rule described in Equation 8. In order to test the behavior of the network before and after 10 epochs of training, the same set of visual stimuli were also presented to the input layer with STDP turned off before and after training, and the resulting spike trains of neurons in the output layer were recorded for analysis.

Effect of Varying Synaptic Connectivity Within Network

In this section, we explore the performance of the model with different kinds of synaptic connectivity present within the network architecture. Specifically, we simulate the model with the following four different network connectivities: (1) feedforward (FF) connections only, (2) FF + Feedback (FB) connections, (3) FF + Lateral (LAT) connections, and (4) FF + FB + LAT connections. Our aim is to assess the contributions that each of these different types of synaptic connection make toward the operation of the model, including especially the relative amounts of stimulus information carried either in the firing rates of individual neurons or by the spike-pair PGs that emerge after training.

Single-cell information analysis was first conducted on the average firing rate responses of individual neurons in the output layer to the three visual stimuli shown in Figure 4 before and after training. The aim was to measure how much stimulus information was carried by the output neurons under the assumption of traditional rate coding. In this analysis, there are three different stimulus categories ($n = 3$). The maximum amount of information for a single neuron is $log_2(n)$, where n is the number of stimulus

categories = 3. Therefore, the maximum amount of information that a neuron can carry about a particular stimulus is $log_2(3) \approx 1.58$ bits. Each visual stimulus was presented twice during testing, each time for a duration of 2 s. Because the precise spike timings of the input vary for the same visual stimulus between trials because of the stochastic nature of Poisson spike generation, this is conceptually equivalent to presenting two transforms of each stimulus category. Individual Layer 4 neurons would have to respond invariantly over the two transforms of a single stimulus category, and not to transforms of the other stimulus categories, in order to carry maximum information about a single stimulus category.

Figure 7a shows the information analysis results for the Layer 4 neuron responses, based on the average firing rates over 2 s of presentation of each visual stimulus. Results before training are shown for the full network architecture with FF + FB + LAT synaptic connections (gray line). Very few output neurons carry the maximal information before training. Results after training are presented for the following four different network connectivities: (1) FF connections only (black dotted line), (2) FF + FB connections (black dash-dot line), (3) FF + LAT connections (black dashed line), and (4) FF + FB + LAT connections (black solid line). For all four different types of network connectivity, around 50 to 100 cells learned to carry the maximum single cell information (FF = 59, FF + FB = 69, FF + LAT = 85, and FF + FB + LAT = 51). Given that the output layer contains a total of 4,096 neurons, in each simulation, only a relatively small fraction of these neurons learned to carry maximal information about stimulus identity in their average firing rates. Moreover, it is noticeable that the network incorporating all three kinds of connections gave the lowest performance.

We next applied the new technique introduced in this article, spike-pair PG information analysis, which is instead based on frequently occurring temporal spike-pairs as described in Information Analysis of Temporal Spike Patterns Emitted by Polychronous Groups. Figure 7b shows the information analysis results for spike-pair PG responses. Results before training are shown for the full network architecture with FF + FB + LAT synaptic connections (gray line). Before training, very few spike-pair PGs carry the maximal information of 1.58 bits. The four different black lines show the results after training for the four different network connectivities: (1) FF connections, (2) FF + FB connections, (3) FF + LAT connections, and (4) FF + FB + LAT connections. All four network architectures produced large numbers of spike-pair PGs that carried the maximal amount of information about stimulus identity (FF = 66, FF + FB = 244, FF + LAT = 469, and FF + FB + LAT: 973).

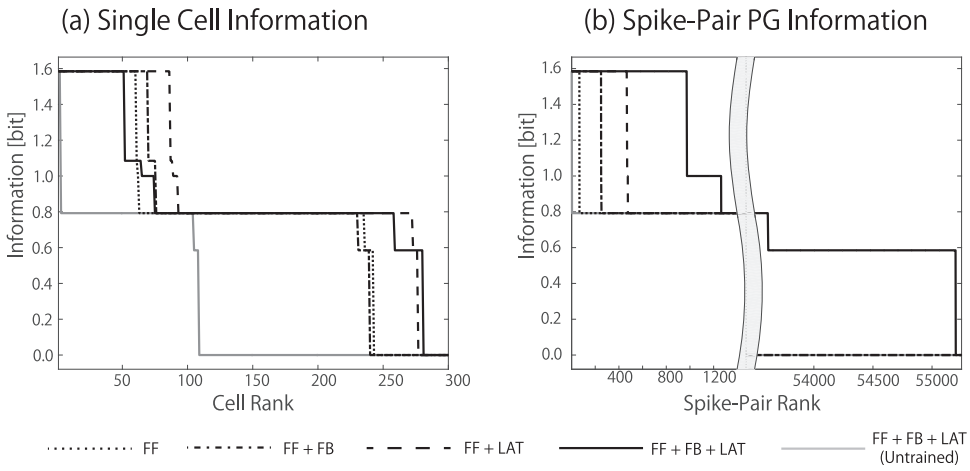

Figure 7. (a) Single-cell average firing rate-based information analysis: We computed the information carried by the output (fourth layer) neurons about a specific object shape. The plot shows the maximum single-cell information carried by 300 cells in Layer 4, where the cells are plotted along the abscissa in rank order. The results before training for the full network architecture with feedforward (FF) + feedback (FB) + lateral (LAT) synaptic connections are plotted in gray. It can be seen that before training, very few output neurons carry the maximal information of 1.58 bits. The four different black lines show the results after training for four different network connectivities: (1) FF connections only, (2) FF + FB connections, (3) FF + LAT connections, and (4) FF + FB + LAT connections. It is evident that all four types of network architecture have produced around 50 to 100 output neurons with maximal single-cell information. (b) Spike-pair polychronous group (PG) information analysis: We computed the information carried by frequently occurring temporal spike-pair PGs in the output (fourth layer) neurons about visual object shape. The plot shows the maximum information carried by spike-pair PGs in Layer 4, where the spike-pair PGs are plotted along the abscissa in rank order. The results before training for the full network architecture with FF + FB + LAT synaptic connections are plotted in gray. It can be seen that before training, very few spike-pair PGs carry the maximal information of 1.58 bits. The four different black lines show the results after training for four different network connectivities: (1) FF connections, (2) FF + FB connections, (3) FF + LAT connections, and (4) FF + FB + LAT connections. It is evident that the full network architecture with FF + FB + LAT connections has produced the most spike-pair PGs with maximal information. Indeed, with the full network architecture, almost 1,000 spike-pair PGs have reached the maximum information of 1.58 bits.

Importantly, it can be seen that the full network architecture with FF + FB + LAT connections produced the most spike-pair PGs with maximal information. Indeed, with the full network architecture almost 1,000 spike-pair PGs reached the maximum information of 1.58 bits. In particular, the number of spike-pair PGs with maximal information in the full network architecture is far greater (about 10 times) than the number of single output neurons achieving maximal information using a firing rate coding shown in Figure 7a.

Thus, the full network developed many spike-pair PGs during visually guided learning that were tuned to specific stimuli. In particular, a major novel result of the current work is that this self-organization of stimulus-specific spike-pair PGs occurred even when the stimulus input representations were *randomized* Poisson spike trains, in which the temporal ordering of spikes varied stochastically across different presentations of the same visual stimulus. The development of (spike-pair) PGs using STDP during visual training in such a spiking network is thus a highly robust process that operates perfectly well with randomized stimulus spike patterns in the lower stages of processing. Furthermore, the information results shown in Figure 7 clearly illustrate the greater potential of temporal coding over traditional rate coding in terms of representational capacity within a biologically realistic spiking neural network with bottom-up, top-down, and lateral connections.

Effects of Varying Key Model Parameters

We next investigate the effects of varying key model parameters in order to identify which factors are important to the emergence of temporal coding by PGs. In particular, we explore the effect of varying the STDP time constant and the number of synaptic contacts between each pair of pre- and postsynaptic neurons on the information carried by spike-pair PGs in the output layer. This part of the investigation uses a full model with all three kinds of synaptic connectivity (FF + FB + LAT).

Figure 8a shows the spike-pair PG information carried by frequently occurring temporal spike-pairs in the output layer with the STDP time constants τ_C and τ_D both set to either 5 ms (solid line), 25 ms (dashed line), or 125 ms (dotted line). The results show that shortening the STDP time constants promotes the emergence of spike-pair PGs with maximal information about which stimulus is presented to the network. In particular, the network develops the largest number of such stimulus specific spike-pair PGs when the STDP time constants are shortest (i.e., 5 ms). However, as the STDP time constant increases, the number of object specific spike-pair PGs decreases. Increasing the STDP time constant makes the precise timing of the spikes less important for learning, making the effect of learning more similar to that expected from traditional Hebbian learning in a rate-coded model. This result implies an important role of temporally precise STDP for the development of temporal coding.

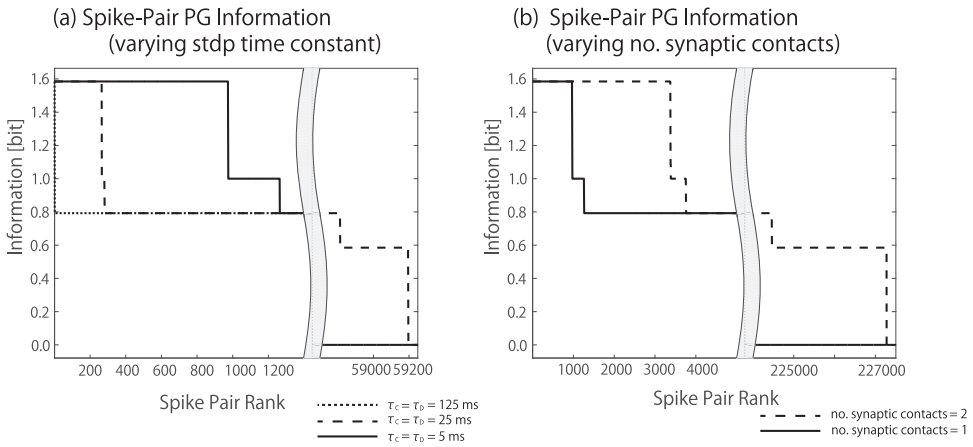

Figure 8. Spike-pair polychronous group (PG) information analysis: We computed the information carried by frequently occurring temporal spike-pair PGs in the output layer of a trained network about visual object shape. The two subplots show the maximum information carried by spike-pair PGs in Layer 4, where the spike-pair PGs are plotted along the abscissa in rank order. (a) Spike-pair PG information scores for three values of the spike-timing-dependent plasticity (STDP) time constants $\tau_C = \tau_D = 125$ ms, 25 ms, or 5 ms. It is evident that shortening the STDP time constants promotes the emergence of spike-pair PGs with maximal information about which stimulus is presented to the network. In particular, the network develops the largest number of such stimulus-specific spike-pair PGs when the STDP time constants are shortest (i.e., 5 ms). As the STDP time constants are lengthened, this reduces the temporal precision of the STDP and so degrades the emergence of PGs. (b) Spike-pair PG information scores for the cases in which the number of plastic synaptic contacts between each pair of pre- and postsynaptic excitatory neurons is either one or two. It can be seen that there is a large increase in the number of spike-pair PGs with maximal stimulus information when there are two synaptic contacts with different transmission delays, rather than just one contact, between each pair of pre- and postsynaptic excitatory neurons. The presence of two synaptic contacts between each pair of pre- and postsynaptic excitatory neurons enables the STDP to select which of the transmission delays to strengthen in order to promote the development of PGs.

Figure 8b compares the information carried by frequently occurring temporal spike-pair PGs in the output layer when the number of plastic synaptic contacts between each pair of pre- and postsynaptic excitatory neurons is either one or two. In the latter case, the two synaptic contacts between each pair of pre- and postsynaptic excitatory neurons had different durations that were randomly assigned at the beginning of the simulation and remained fixed throughout. Only the strengths of these connections could be modified by STDP during visually guided learning. The results show that having multiple synaptic contacts, and hence multiple axonal transmission delays, between each pair of excitatory neurons, increases the number of spike-pair PGs with maximum information. The presence of two synaptic contacts between each pair of pre- and postsynaptic excitatory neurons enables the STDP to select which of the transmission delays to strengthen in order to promote the development of (stimulus specific) PGs as hypothesized.

The results shown in Figure 8b demonstrate how such a spiking model may exploit the ability to effectively select (self-organize) the durations of the axonal delays in the plastic connections between excitatory neurons. If there is no self-organization during visual learning over synaptic delay lengths, then it is effectively prespecified by the initial random distribution of axonal transmission delays within the network whether a given neuron can be a part of a particular PG. By allowing multiple plastic synaptic contacts, with different delays, between each pair of pre- and postsynaptic excitatory neurons, we expected that STDP would effectively select which of these axonal delays to strengthen.

We next took a deeper look into the selective strengthening and weakening of synaptic contacts with different axonal transmission delays between each pair of pre- and postsynaptic excitatory neurons. This analysis was carried out on the same simulation with two synaptic contacts with different fixed delays between each pair of neurons. In this case, STDP could select one of the delays to be strengthened while weakening the other during visual training. For each pair of connected pre- and postsynaptic neurons, we calculated the absolute difference between the synaptic weights of the two synaptic contacts both before and after training. The absolute difference in the values of the two synaptic weights after training should reflect how effectively the STDP has selectively strengthened one connection with a particular delay but weakened the other connection with a different delay, which is necessary to promote the emergence of many stimulus specific spike-pair PGs. Specifically, before and after training, we computed frequency histograms in which pairs of pre- and postsynaptic excitatory neurons were binned according to the absolute difference between the weights of their two synaptic contacts.

Figure 9a shows the result of dividing the frequency histogram after training by the frequency histogram before training on a bin by bin basis. Thus, the subplot marked "a" shows the factor by which the number of pairs of neurons with a particular absolute difference between the weights of their two synaptic contacts changes after training. It can be seen that after training, there was a large increase in the number of pairs of pre- and postsynaptic excitatory neurons with the maximum possible synaptic weight difference. This implies that during visual training, one of the synaptic weights went to its maximum value of 1.0, whereas the other synaptic weight went to the minimum value of 0.0. This represents successful synaptic delay selection by STDP.

Figure 9b shows examples of synaptic modifications for four pairs of pre- and postsynaptic excitatory neurons, where each such pair of neurons has two synaptic contacts with different transmission delays. For each pair of neurons, these plots show selective strengthening of one synaptic contact with a particular delay but weakening of the other synaptic contact with a different delay. These results clearly demonstrate that STDP is able to selectively strengthen and weaken synaptic connections during visual learning according to their respective transmission delays. The model thus selects and self-organizes its effective synaptic delays, which can greatly facilitate the emergence of stimulus specific (spike-pair) PGs.

The Emergence of Larger Scale Polychronous Groups

In this section, we explored the development of larger scale PGs (i.e., containing more than just two neurons). In particular, we investigated how the development of these PGs is influenced by changing the kind of synaptic connectivity implemented within the network. For each simulation with a different connectivity structure (FF only, FF + FB, FF + LAT, and FF + FB + LAT), we identified all the potential PGs triggered from cells in the third layer of the network based on the synaptic connectivity, conduction delays, and synaptic weights as explained in Polychronous Group Counting. Furthermore, based on the actual spike trains recorded during testing, we investigated whether any of the activated PGs had learned to be stimulus specific.

Table 2 shows the statistics of the PGs that emerged in network models with different kinds of synaptic connectivity after training. The top row shows the total number of PGs that were identified. The general trend is that as the synaptic connectivity becomes more complex, that is, with more types of connection, the number of PGs increases. In particular, by far the largest number of PGs was found in the full network architecture with FF + FB + LAT connections. The middle row of Table 2 shows the mean number of spikes in each PG. Lastly, the bottom row presents the mean longest path length of each PG, where the longest path is defined as the number of neurons involved in the longest chain of spikes emitted by the PG because of the activation of the trigger neurons (Izhikevich, 2006). It can be seen that we get an increase in both of these statistics as the network architecture includes more types of synaptic connection. Indeed, the full network architecture (FF + FB + LAT) also gives rise to the largest mean number of spikes in each PG and the mean longest path length of each PG. The full network architecture is clearly the most efficacious for promoting the emergence of polychronization.

The detailed statistical distributions underlying the mean values shown in the second and third rows of Table 2 are shown as box plots in Figure 10. Figure 10a shows the distribution of the average number of spikes in a PG, whereas Figure 10b presents the distribution of the longest path of spikes in a PG. For both subplots, the red horizontal lines indicate the medians, and the red circles indicate the means. As already shown in Table 2, the full trained network architecture (FF + FB + LAT) gives rise to the largest mean number of spikes in each PG and mean longest path length of each PG compared with the three other reduced network connectivities. The results from the four trained networks are compared with those from the untrained full network architecture (FF + FB + LAT) shown on the right of each subplot. By comparing the results for the full network architecture before and after training, it can be seen that training has led to a

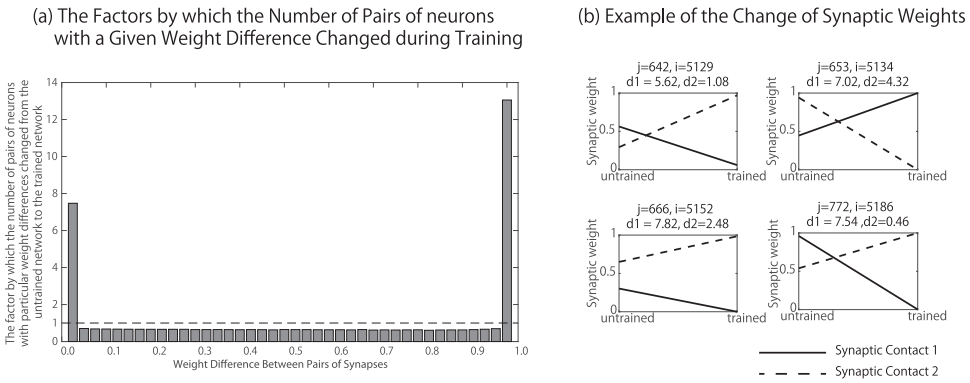

Figure 9. Simulation results demonstrating selection of effective synaptic delays by spike-timing-dependent plasticity (STDP) during visual training. In this simulation, the model has two synaptic contacts with different fixed delays between each pair of pre- and postsynaptic excitatory neurons. STDP can then select one of the delays to be strengthened while weakening the other during visual training. For each pair of connected pre- and postsynaptic excitatory neurons we calculated, both before training and after training, the absolute difference between the synaptic weights of their two synaptic contacts. We then computed two frequency histograms, corresponding to before and after training, in which all such pairs of pre- and postsynaptic excitatory neurons were binned according to the absolute difference between the weights of their two synaptic contacts. These two frequency histograms were then used to compute the plot shown in 9a, which shows the result of dividing the frequency histogram after training by the frequency histogram before training on a bin by bin basis. Thus, the subplot marked "a" shows the factor by which the number of pairs of neurons with a particular absolute difference between the weights of their two synaptic contacts changes after training. It can be seen that after training there was a large increase in the number of pairs of pre- and postsynaptic excitatory neurons with the maximum possible synaptic weight difference. This implies that during visual training, one of the synaptic weights went to its maximum value of 1.0, whereas the other synaptic weight went to the minimum value of zero. This represents successful synaptic delay selection by STDP. (b) Shows four examples of the changes in the weights of the two synaptic contacts between a pair of pre- and postsynaptic excitatory neurons that occurred during training. The solid and dashed lines in each subplot represent the strengths of the two synaptic contacts between a pre- and postsynaptic neuron, each with a different synaptic delay. For each pair of neurons, these plots show selective strengthening of one synaptic contact with a particular delay but weakening of the other synaptic contact with a different delay. Again, it is clearly evident that STDP is able to selectively strengthen and weaken synaptic connections during visual learning according to their respective transmission delays.

significant increase in the mean number of spikes in each PG and the mean longest path length of each PG.

We next investigated whether the PGs that developed after training in the full network architecture (with all three connectivity types FF + FB + LAT) had learned to respond to a particular stimulus category. This was done by analyzing the responses of the actual trigger events for these PGs in Layer 3 to the three visual stimuli: the circle, heart, and star. The stimulus-selective PG trigger events were identified as being selective for one of the stimuli by using information analysis. This was done using a similar information analysis to that used for single cells, but instead using the occurrences of the PG trigger events in the spike trains of Layer 3 neurons.

Figure 11 plots the occurrences of the stimulus-selective PG trigger events (identified by the information analysis) when the network was tested on the three visual stimuli: the circle, heart, and star. The figure shows the occurrences of these PG trigger events when each of the stimuli is presented twice to the network, each time for 2 s. Specifically, the circle is presented twice, followed by two presentations of the heart and then two presentations of the star. It can be seen that PG Trigger Events 1 to 70 respond

Table 2
Statistics of the PGs That Emerged in Network Models With Different Kinds of Synaptic Connectivity After Training

Parameters	Synaptic connectivity							
	FF		FF + FB		FF + LAT		FF + FB + LAT	
Number of PGs	562		1,689		827		32,317	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total number of spikes in PG	8.91	6.16	8.94	7.30	8.70	5.98	11.41	6.56
Longest path of spikes in PG	1.00	.00	1.12	.44	1.31	.53	2.55	.97

Note. FF: feed-forward; FB: feedback; LAT: lateral; PG = polychronous group.

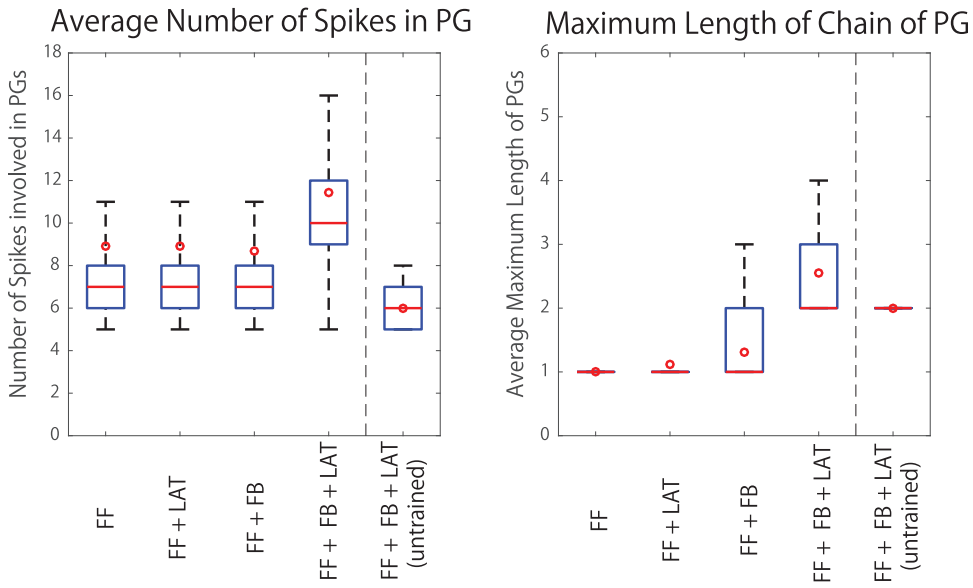

Figure 10. Box plots showing key performance statistics of the polychronous groups (PGs) that emerged in network models with different kinds of synaptic connectivity (FF only, FF + FB, FF + LAT, and FF + FB + LAT) after training. The subplot marked "a" shows the distribution of the average number of spikes in a PG, whereas "b" presents the distribution of the longest path of spikes in a PG. For both subplots, the red horizontal lines indicate the median, and the red circles indicate the means. It is evident that the full trained network architecture (feedforward [FF] + feedback [FB] + lateral [LAT]) gives rise to the largest mean number of spikes in each PG and mean longest path length of each PG compared with the three other reduced network connectivities. The results from the four trained networks are compared with those from the untrained full network architecture (FF + FB + LAT) shown on the right of each subplot. By comparing the results for the full network architecture before and after training, it can be seen that training has led to a significant increase in the mean number of spikes in each PG and the mean longest path length of each PG. See the online article for the color version of this figure.

selectively to the circle, PG Trigger Events 71 to 102 respond to the heart, and PG Trigger Events 103 to 123 respond to the star.

These results confirm that in the trained full network architecture (FF + FB + LAT), large numbers (i.e., greater than 100) of PGs respond selectively to just one of the stimuli, and do so across different presentations of that stimulus.

The Emergence of Binding Neurons

Finally, we analyzed the PGs from the full network model (FF + FB + LAT) that were found to respond to specific stimuli in The Emergence of Larger Scale Polychronous Groups for the presence of the hypothesized binding neurons as illustrated in Figure 3a. Figure 12 shows examples of activated PGs, binding neurons, and visual features represented by input Gabor filters that drive the cells in the PGs. Simulation results are presented from the trained full network architecture when tested on the three visual stimuli: the circle, heart, and star. Each row (a-c) represents a PG of neurons that responds selectively to one of the stimuli (subplot in left pane) and the visual features represented by the input Gabor filters with strong connections to particular neurons explicitly identified in the PGs (two subplots in right pane). The PGs shown in the rows marked "a," "b," and "c" respond selectively to the circle, heart, and star, respectively.

In the PG plots shown on the left of Figure 12, the neurons are identified by small circles and the strengthened connections between the neurons are represented by lines. In particular, the rows marked "a" and "c" present clear examples of the hypothesized binding neurons. In these PG plots, the three neurons that make up the three-neuron binding circuit (as illustrated in Figure 3a) have bold connections between them. For example, in row "a," there are three neurons in the binding circuit as follows: Neuron 12,686 (a PG trigger neuron in Layer 3) represents the low-level feature, Neuron 18,657 (a Layer 4 output neuron) represents the high-level feature, and Neuron 18,396 is the related binding neuron between these two features. It can be seen that the low-level Feature Neuron 12,686 sends a connection to the high-level Feature Neuron 18,657, and both Feature Neurons 12,686 and 18,657 send connections to the Binding Neuron 18,396. In particular, it can be seen from the axonal transmission delays shown in the plot that if the low-level Feature Neuron 12,686 is driving the high-level Feature Neuron 18,657, then the spikes emitted by both of these feature neurons will arrive at the Binding Neuron 18,396 at about the same time and so reinforce each other. Thus, the Binding Neuron 18,396 will fire if the low-level Feature Neuron 12,686 is actually driving the high-level Feature Neuron 18,657. A similar binding relationship between three neurons is shown in row "c." Row "b" shows that mixtures of polychronous representation types emerge in the same neuronal layers.

The two subplots presented in the right panes of rows (a-c) in Figure 12 show the visual features represented by the input Gabor

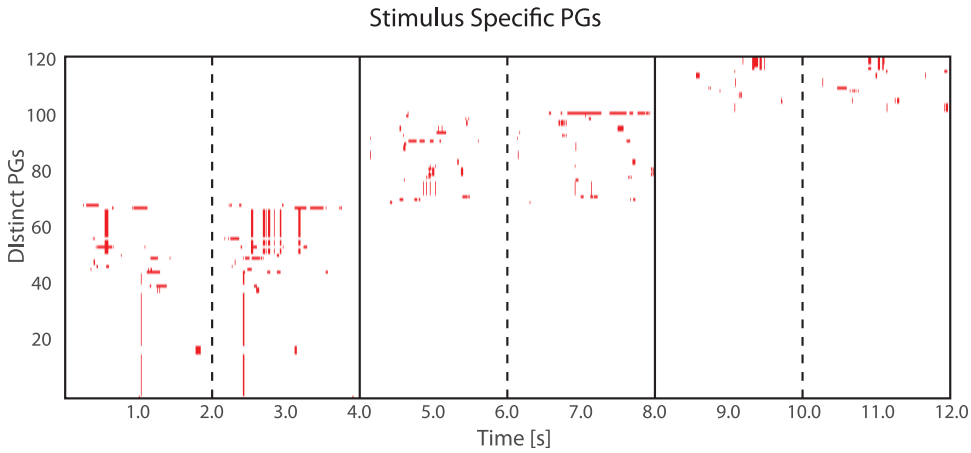

Figure 11. Graphical representation of the occurrences of stimulus-selective polychronous group (PG) trigger events in Layer 3 of the trained full network architecture (feedforward [FF] + feedback [FB] + lateral [LAT]) when tested on the three visual stimuli: the circle, heart, and star. The stimulus-selective PG trigger events were first identified as being selective for one of the stimuli by using information analysis. In the figure, we show the occurrences of these PG trigger events when each of the stimuli is presented twice to the network, each time for 2 s. Specifically, the circle is presented during 0 to 2 s, and then again during 2 to 4 s. Next, the heart is presented during 4 to 6 s, and then again during 6 to 8 s. Finally, the star is presented during 8 to 10 s, and then again during 10 to 12 s. The distinct stimulus-selective PG trigger events identified by the information analysis are numbered along the ordinate. It can be seen that PG Trigger Events 1 to 70 responded selectively to the circle, PG Trigger Events 71 to 102 responded to the heart, and PG Trigger Events 103 to 123 responded to the star. See the online article for the color version of this figure.

filters that have strong feedforward connections to two neurons from the PG shown in the left pane. Specifically, the right pane shows a Layer 3 trigger neuron for the PG (left), and a Layer 4 neuron from within the same PG (right). To produce these subplots, the feedforward synaptic connections between successive layers are traced back to the input Gabor filters in order to determine the specific visual features that drive the responses of the higher layer neurons. Starting from a particular neuron in Layer 3 or 4, we select the connections from the previous layer that have the highest weights, repeating this process through successive layers until the connections reach the Gabor filters in the input layer. This then allows us to plot the pattern of Gabor input filters that the neuron in Layer 3 or 4 has become tuned to. Looking at the right panes of Figure 12 (a and c), it is clear that the Layer 3 neurons (left) represent simpler low-level features, whereas the Layer 4 neurons (right) represent more global features of the entire object. For example, in row "a" of Figure 12, it can be seen that the low-level Feature Neuron 12,686 in Layer 3 (left) receives strong connections from a simpler set of input Gabor filters than the high-level Feature Neuron 18,657 in Layer 4 (right). Moreover, comparison with the corresponding PG plots in the left panes of Figure 12 (a and c) shows that the layer four neurons are being driven by the simpler Layer 3 neurons, with the outputs of both Layer 3 and 4 neurons driving an associated binding neuron. These results are consistent with our underlying theoretical framework about binding taking place between low-level and high-level visual features.

Feedforward Projection of Information About Low-Level Visual Features to Higher Neuronal Layers

Simulations of the full spiking network architecture (FF + FB + LAT) provided examples of the kind of feedforward propagation of visual information hypothesized in Feedforward projection of information about low-level visual features to higher neuronal layers and illustrated in Figure 5a. The binding neurons presented in the rows marked "a" and "c" of Figure 12 were in fact in Layer 4. Thus, in each of these examples, the low-level feature neuron was in Layer 3, the high-level feature neuron was in Layer 4, and the binding neuron was also in Layer 4. In these cases, information about the low-level feature represented in Layer 3, including its local image context (i.e., that the low-level feature represented in Layer 3 is part of the high-level feature represented in Layer 4), is projected onto the binding neuron in Layer 4. These simulation results confirm the feasibility of the hypothesis that low-level visual information is propagated forward (i.e., bottom-up) to higher layers in the manner proposed earlier. This could allow information about low-level features to be represented in the highest layers of the network, where, in principle, this information could be read out by subsequent behavioral systems.

Discussion

In this article, we explored the operation of a biologically detailed neural network model of the primate ventral visual system. The model incorporates the following key aspects of cortical dynamics and architecture: (a) the model implements spiking neural dynamics in which the timings of action potentials or "spikes" are simulated explicitly; (b) STDP is used to modify the synaptic connections during visually guided learning; (c) the network architecture incorporates bottom-up, top-down, and lateral synaptic connections; (d) the synaptic connectivity between neurons incorporates distributions of axonal conduction delays of varying durations; and (e) in some simulations multiple synaptic connections with different axonal transmission delays are incor-

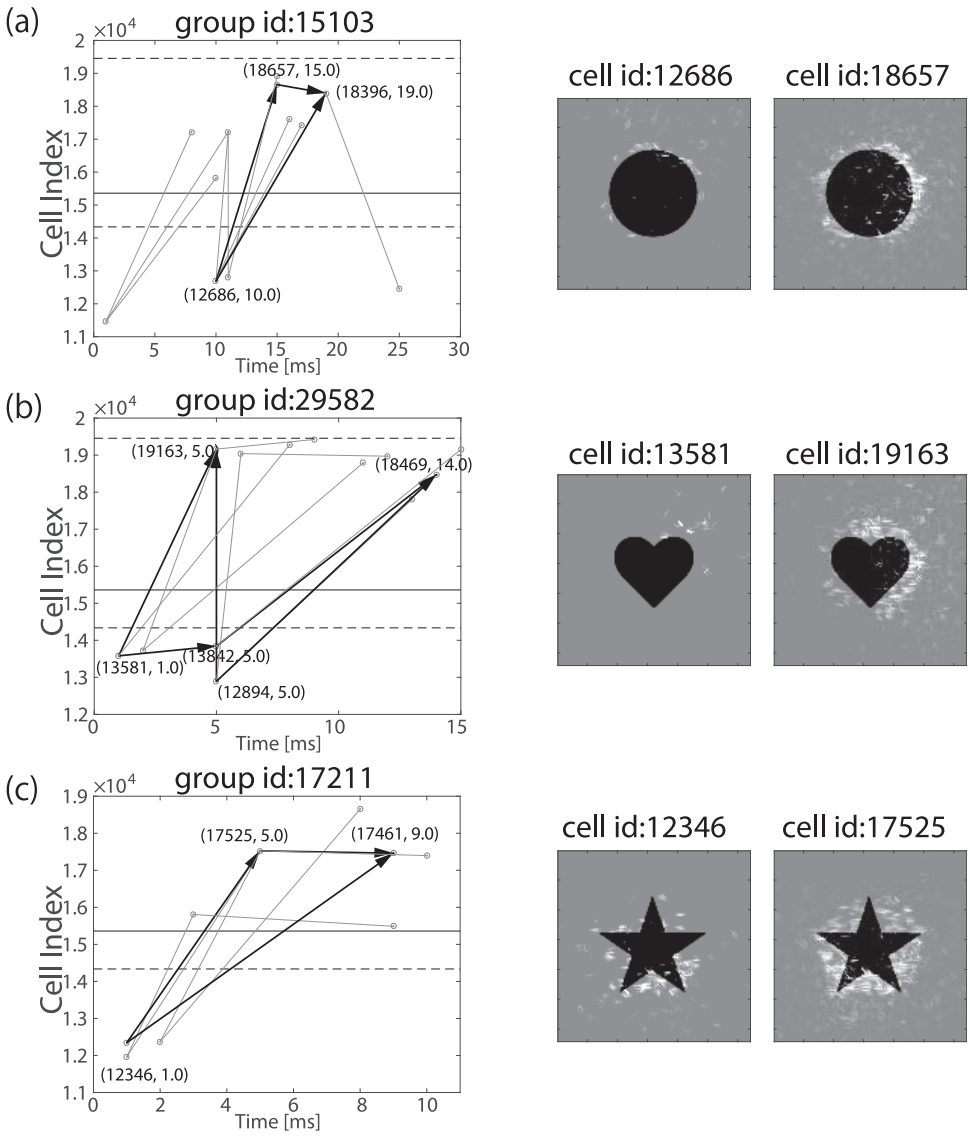

Figure 12. Examples of activated polychronous groups (PGs), binding neurons, and visual features represented by input Gabor filters that drive the cells in the PGs. Simulation results are presented from the trained full network architecture (feedforward [FF] + feedback [FB] + lateral [LAT]) when tested on the three visual stimuli: the circle, heart, and star. The left side of the figure presents three examples of the PGs identified using the technique explained in Polychronous Group Counting. In our model, the cell index is assigned as follows: Cells with index 1 to 4,096 (64 × 64) are the excitatory cells in the first layer; 4,096 + 1 to 4,096 + 1,024 (32 × 32) are the inhibitory cells in the first layer; 5,120 + 1 to 5,120 + 4,096 are the excitatory cells in the second layer; 9,216 + 1 to 9,216 + 1024 are the inhibitory cells in the second layer; and so on. To help identify the different kinds of cells in the plots, the dotted horizontal lines mark the boundary between the excitatory and inhibitory neurons within a layer, whereas the solid horizontal lines indicate the separation between the inhibitory neurons in one layer and the excitatory neurons in the next higher layer. Each row (a-c) represents a PG of neurons that responds selectively to one of the stimuli (subplot in left pane) and the visual features represented by the input Gabor filters with strong connections to particular neurons explicitly identified in the PGs (two subplots in right pane). In the PG plots (shown on the left), the neurons are identified by small circles and the strengthened connections between the neurons are represented by lines. The neurons are plotted along the abscissa according to the relative timings of their spikes within the PGs, which was determined by the axonal transmission delays of the strengthened connections between the neurons. In particular, rows "a" and "c" present clear examples of the hypothesized binding neurons.

porated between each pair of pre- and postsynaptic neurons. These are basic, known aspects of the architecture and function of the visual cortex. Using this model architecture, we explored a number of major computational hypotheses as follows.

Emergence of Polychronization

Previous authors have considered the potential role of synchronization in solving the feature binding problem, whereby the neurons representing the visual features of a particular object emit their spikes clustered closely together in time (Evans & Stringer, 2012, 2013; Kreiter & Singer, 1996). However, in this paper we have investigated the potential role of polychronization in solving feature binding, in which subpopulations of neurons (PGs) emit their spikes in fixed spatiotemporal patterns that repeat across different presentations of the same stimulus.

In the simulations reported above, we demonstrated that polychronization emerges naturally in the network when distributions of randomised axonal transmission delays of the order of several milliseconds are incorporated. The incorporation of such axonal delays has the effect of flipping the model behaviour from synchronization to polychronization. In particular, we showed that even if the visual stimuli (circle, heart, and star) presented to the network were encoded in the input layer by randomized Poisson spike trains, the synaptic connectivity in the later layers of the network could still self-organize using STDP during visually guided learning such that polychronous groups (PGs) emerged naturally. Moreover, we found that individual PGs learned to respond to particular stimuli that the network was trained on, that is, the PGs responded in a stimulus-specific manner.

During each training epoch, each object shape was presented for 2 s to the network while the synaptic weights were adapted using an STDP learning rule. After 10 epochs of training, the network had learned stimulus specific representations of each object. The change in the distribution of the synaptic weights is shown in Figure 9.

The output (fourth) layer was found to carry more stimulus information if we assumed a temporal coding based on patterns of spike times within PGs instead of assuming traditional rate coding by individual neurons. Our results found that the inclusion of feedback and lateral connections in the network structure led to an increase in the number and length of PGs (especially spike-pairs). In particular, the full network architecture with FF, FB, and LAT synaptic connections produced the most spike-pair PGs with maximal stimulus information. These spike-pair PGs were tuned to specific stimuli.

A major novel result of the current work is that this self-organization of stimulus-specific spike-pair PGs occurred even when the stimulus input representations were *randomized* Poisson spike trains, in which the temporal ordering of spikes varied stochastically across different presentations of the same visual stimulus. The development of (spike-pair) PGs using STDP during visual training in such a spiking network is thus a highly robust process that operates perfectly well with randomized stimulus spike patterns in the lower stages of processing.

The development of temporal PG codes was shown to be dependent on the temporal specificity of the STDP learning rule used to modify the synaptic connections. It was found that the network develops the largest number of spike-pair PGs with maximal information about which stimulus is presented to the network when the STDP time constants are shortest (i.e., 5 ms). However, increasing the STDP time constants in the simulations had the effect of decreasing the number of object specific spike-pair PGs that emerged. The explanation for these observations is that increasing the STDP time constants makes the precise timing of the spikes less important for learning, which, in turn, makes the synaptic weight modification more similar to traditional Hebbian learning in a rate-coded model. Consequently, these simulation results suggest an important role for temporally precise STDP in the development of temporal coding.

Another novel feature of some of the simulations reported in this article was the incorporation of multiple synaptic contacts with different axonal transmission delays between each pair of pre- and postsynaptic neurons. This corresponds to a presynaptic neuron making multiple synaptic connections on different parts of the dendritic branching of a postsynaptic neuron, as is seen among real neurons in the brain. In such a network architecture, STDP was able to select which synapses to strengthen and which synapses to weaken, which promoted the visually guided development of PGs of spiking neurons. Thus, during self-organization, the network is able to effectively select for synaptic transmission delays between pre- and postsynaptic neurons, which results in a greater representational capacity.

Importantly, the stimulus-specific representations that developed in the output layer were robust with respect to random jitter in the input spike patterns, as the visual objects presented to the network were represented by randomized Poisson spike trains as described in Training and Stimuli.

Emergence of Binding Neurons

The feature *binding problem* in visual neuroscience is expressed by different authors in rather different ways. However, it generally boils down to the question of how the visual system represents which features are bound together as part of the same object. For example, if the two letters T and L are seen together, how does the visual system represent which horizontal and vertical bars are part of which letter? Over the last 20 years, our laboratory has developed a hierarchical, *rate-coded* neural network model, VisNet, of the primate ventral visual pathway (Eguchi, Humphreys, & Stringer, 2016; Galeazzi, Minini, & Stringer, 2015; Wallis & Rolls, 1997). This network model represents low-level visual features in the lower layers and higher level features or objects in the higher layers, but there is no way to identify which features are part of which objects from the activity of these neurons. How visual features are bound together must underpin how we segment a visual scene into objects and parts of objects, and thus how we make sense of the visual world. Duncan and Humphreys (1989) provide a good description of this hierarchical process as follows:

> A fully hierarchical representation is created by repeating segmentation at different levels of scale. Each structural unit, contained by its own boundary, is further subdivided into parts by the major boundaries within it. Thus, a human body may be subdivided into head, torso, and limbs, and a hand into palm and fingers. Such subdivision serves two purposes. The description of a structural unit at one level of scale (animal, letter, etc.) must depend heavily on the relations between the parts defined within it (as well as on properties such as

> color or movement that may be common to the parts). Then, at the next level down, each part becomes a new structural unit to be further described with its own properties, defined among other things by the relations between its own sub-parts. At the top of the hierarchy may be a structural unit corresponding to the whole input scene, described with a rough set of properties (e.g., division into light sky above and dark ground below). (p. 445)

The new generation of spiking neural network simulations reported in this article, in which the timings of action potentials or spikes are explicitly simulated, aim to begin to solve this hierarchical feature binding problem. Our basic conception is that within the PGs that emerge during visually guided learning are embedded *binding neurons* that represent the binding relationships between low-level and high-level visual features. It is assumed that neurons in the network behave as "coincidence detectors," in that they require a volley of spikes from presynaptic cells to arrive simultaneously at the postsynaptic cell in order for the postsynaptic cell to fire itself. The basic three-neuron binding circuit is illustrated in Figure 3a.

Simulations of the full spiking network architecture (FF + FB + LAT) presented demonstrated the emergence of binding neurons, which were part of the same kind of three-neuron binding circuit as shown in Figure 3a. These simulation results were shown in Figure 12. In these simulations, the binding neurons represented the binding relationships between lower level feature neurons in Layer 3 and higher level feature neurons in Layer 4. Moreover, the individual PGs, in which these binding neurons were embedded, responded to specific visual stimuli (the circle, heart or star). Such binding neurons were originally proposed by von der Malsburg (1999), but without an explanation of how they might emerge naturally during visual development. In the simulations reported, we demonstrated that such binding neurons may develop automatically within the PGs that emerge during visually guided learning with STDP. In particular, these binding representations were shown to emerge even when the visual stimuli are encoded by randomized (Poisson) spike trains in the input layer. The binding neurons that develop carry measurable information about which low-level features are driving (and hence part of) which high-level features. Our theory predicts that such binding neurons should develop across the visual field, at every layer of the feature hierarchy, and at every spatial scale within a natural visual image.

The utilisation of polychronisation within a spiking neural network allows the model to develop binding neurons with the crucial property that they respond *if and only if* a low-level feature neuron is actually participating in driving a high-level feature neuron. Only in this case will the binding neuron be fully informative that the low-level feature is part of the high-level feature. It is important that the binding neuron does not fire whenever the low-level feature neuron and the high-level feature neuron happen to be simultaneously active. For example, if the letters T and L are presented together, then the network dynamics should not activate the binding neuron linking the low-level feature neuron encoding the vertical bar of the T (represented in a lower layer) to the high-level feature neuron encoding the letter L (represented in a higher layer). For this reason, we propose that binding may not be soluble within a traditional rate-coded network, but will instead require the full spiking neuronal dynamics of the brain.

Our model of the primate ventral visual pathway contains bottom-up, top-down, and lateral synaptic connections in order to reflect the known architecture of this part of the brain. Given this kind of connectivity, there are a variety of ways of realizing the three-neuron binding circuit shown in Figure 3a. For example, the binding neuron might be in the same layer as the low-level feature neuron, or in the same layer as the high-level feature neuron, or in a different area completely. We discuss the feedforwarded projection of information about low-level features that may occur if the binding neuron is in the same layer as the high-level feature neuron (holographic principle). However, wherever the binding neurons are located, they will carry measurable information about the binding relations within a visual scene. Moreover, the theory presented in this article implies that binding neurons will develop throughout all visual processing areas of the visual cortex, thus representing the binding relations across the visual field and at every spatial scale. A rich tapestry of binding neurons through the layers could help to provide a rich hierarchical structural description of a scene, rather analogous to that described earlier by Duncan and Humphreys (1989).

The example given in Figure 3a shows how binding neurons may learn to represent the fact that a particular low-level visual feature such as a horizontal or vertical bar is driving, and therefore part of, a given high-level feature such as the letter T. However, binding neurons may learn to represent many other kinds of relationship between features within an image. For example, a binding neuron might learn to respond when a low-level feature (such as a vertical bar) is part of an intermediate-level feature (such as the letter T), which is, in turn, part of a high-level feature (such as the word CAT). In this case, the binding neuron receives simultaneous inputs from the low-level, intermediate, and high-level feature neurons, as shown in Figure 13a. Alternatively, a binding neuron could represent that a low-level feature (such as a vertical bar) is simultaneously part of two different higher level features (such as the letter T and the word CAT), as shown in Figure 13b. Or a binding neuron could represent that two low-level features (such as a vertical bar and a horizontal bar) are both part of a higher level feature (such as the letter T), as shown in Figure 13c. There are a vast number of such relationships that could be represented by binding neurons. What kinds of relationships actually get represented will depend on the visual images used to train the network model. In future research, we will explore what kinds of binding relationship become represented in the model as it is trained on lots of natural images. Such binding information is essential to the rich semantic analysis and interpretation of visual images performed by the visual brain.

The binding hypothesis proposed in this paper also provides a way in which the visual system might localise (parts of) objects in space. The ventral visual pathway of the primate brain is thought to extract visual features of increasing complexity as one moves up along the pathway. For example, simple cells in the primary visual cortex represent oriented bars and edges in localised regions of retinal space, while neurons in higher stages of visual processing may represent whole objects or faces in a (location, view and scale) invariant manner (Booth and Rolls, 1998; Perry et al., 2010; Wallis and Rolls, 1997). When we look at a visual scene, we are aware of visual features at all such spatial scales. In particular, we are aware of the (e.g.\ retinal) locations of low-level features such as the edges of objects. This kind of information may be represented by edge detecting (e.g.\ simple) cells in lower visual areas,

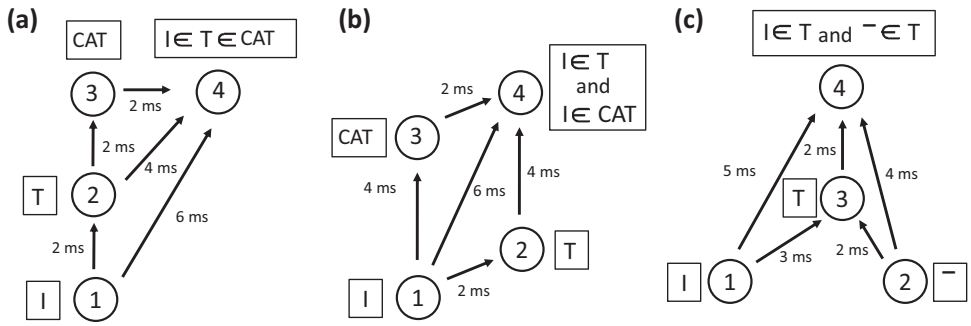

Figure 13. Examples of three different kinds of more complex binding relationships. (a) A binding neuron might learn to respond when a low-level feature (such as a vertical bar) is part of an intermediate-level feature (such as the letter T), which is, in turn, part of a high-level feature (such as the word CAT). In this case, the binding neuron receives simultaneous inputs from the low-level, intermediate, and high-level feature neurons. (b) Alternatively, a binding neuron could represent that a low-level feature (such as a vertical bar) is simultaneously part of two different higher level features (such as the letter T and the word CAT). (c) A binding neuron could represent that two low-level features (such as a vertical bar and a horizontal bar) are both part of a higher level feature (such as the letter T).

which have small receptive fields of about 1 or 2 degrees in size. Neurons with such small receptive fields can effectively localise the edge of an object in space. However, through a process of feature binding, we also see these edges as parts of the boundaries of their respective objects. Thus, the binding of a localised edge represented in an early visual area to an object representation at a higher stage of processing provides a way in which (the parts of) objects may be localised in space. And, indeed, neurophysiological studies have revealed the existence of border ownership cells (Zhou et al., 2000) in the lower cortical visual areas V1 and V2, which are thought to play a key role in binding border edges to objects. Such border ownership cells may be examples of the kind of binding neurons proposed here. Hence, the development of binding neurons within polychronous groups as proposed in this paper provides a plausible explanation for how such binding might operate and play a key role in the localisation of (parts of) objects in space.

In the brain, the low-level and high-level visual features may in fact be represented by their own temporal patterns of spikes distributed across polychronous group of neurons, and these two polychronous groups may then drive a third polychronous group representing the binding relationship between these visual features. This more complex scenario, in which the visual features and the binding relations between these features are represented by their own polychronous groups, is likely to be what actually happens in the brain. The simple three neuron binding circuit shown in Figure 3a would then be a small part of the three corresponding polychronous groups shown in Figure 3b.

Our new approach to solving the feature binding problem in biological spiking neural networks relies on polychrony instead of synchrony. Some previous authors have proposed that a visual scene could be partitioned into separate object regions by synchronisation of neuronal firing (Kreiter and Singer, 1996; Evans and Stringer, 2012, 2013). In this scenario, the spikes emitted by the neurons representing each individual object become synchronised in time, while the spikes emitted by neurons encoding different objects become desynchronised. This mechanism of synchronisation allows a spiking network model to, say, segment and individually bind several different object regions of an image. However, we have previously found that such neuronal synchronisation may be destroyed if we include natural distributions of axonal transmission delays, as we have done in this paper. But more fundamentally, how can simply segmenting a visual scene into several distinct object regions accord with the semantically rich, hierarchical visual experience of primate vision as described by Duncan and Humphreys (1989)? For these reasons, in this paper we have proposed that feature binding may depend on polychronisation rather than synchronisation. The use of polychronisation with binding neurons seems to offer far greater richness in terms of the structural and semantic representation of visual scenes.

This proposal sharply contrasts with the feature integration theory of Treisman and Gelade (1980), which posits that that there is only a single locus of attention within the visual field where visual features are bound together. Some researchers have tried to relate feature binding to the speed of visual search for target objects among nontarget distractors. Given that feature integration theory assumes there is only a single locus of attention where feature binding takes place, this implies a serial search for a visual search task that requires feature binding but allows faster parallel search for other search tasks that do not require feature binding. In contrast, we have proposed that feature binding is carried out by binding neurons that operate simultaneously across the whole visual field, including at every spatial scale. In this case, there is no need for feature binding to be limited to a single spatial locus of attention although spatial attention may still facilitate binding at particular retinal locations. If feature binding does occur across the visual field, then the time taken for visual search would not be governed by the need to perform a serial search with a single locus of attention. Instead, binding may operate in parallel across the visual field, and the search time would be related to other factors determining the intrinsic difficulty of the task. This is supported by the study of Duncan and Humphreys (1989). These authors found no clear dichotomy between serial and parallel modes of search. Instead, they reported that search efficiency was found to decrease as the targets became more similar to nontargets, or if the nontargets became more dissimilar to each other. This finding contradicts the assumption of feature integration theory that there is a single locus of feature bind-

ing, which leads to serial search for those tasks that require feature binding and parallel search for tasks that do not.

However, although our theory permits feature binding to operate in parallel across the entire visual field, it would still be expected that visual processing, which includes feature binding, would be somewhat degraded away from the spatial locus of attention. This could occur because the neural representation of the part of the visual scene at the site of spatial attention, which might be highlighted because of high-acuity foveal fixation or top-down attentional facilitation, would compete strongly with visual processing of the rest of the scene through inhibitory interneurons. This strong inhibition from the attended part of the scene would likely degrade visual processing elsewhere, including feature binding operations. This would explain various psychophysical findings about binding in human vision (Wolfe & Cave, 1999). However, this is quite different from the underlying assumption of feature integration theory, which actually requires only a single spatial locus of attention to perform feature binding, and so cannot permit any binding elsewhere.

We do not mean to suggest that this article provides a complete solution to the feature binding problem. Rather, we believe that the mechanisms illustrated here point the way toward understanding feature binding in the visual cortex. The kinds of three-neuron binding circuits that emerge during visually guided learning, as shown in Figure 3a, are merely the simplest expression of how poychronization could encode the hierarchical binding relations between features within an image. We are at the beginning of exploring this. For example, as one ascends through the network layers, there appears to be a rapid increase in the representational capacity, that is, the number of binding neurons, needed to encode all of the possible binding relations between features at every spatial scale. This article does not provide a detailed analysis of how a network such as this one might resolve the capacity issue. However, we propose a number of hypotheses that follow from this work. For example, as mentioned in the article, each feature binding representation may take the form of a polychronous neuronal group. In this case, individual binding neurons could occur in many binding representations, thus dramatically increasing the representational capacity. One way of investigating this could be to simulate very-large-scale networks trained on millions of natural images and analyze the nature of the features represented as well as the binding relations between them. Moreover, although, in principle, a spiking neural network model can represent visual features and their binding relations across the entire visual field, in the visual brain, processing is usually focused on the high-acuity foveal region of the retinal space. Because of the larger number of neurons devoted to covering this region, we would expect a more fine-grained representation of visual features and the binding relations between them here. Additionally, attention may sometimes be covertly directed to some parafoveal region where visual processing, including feature binding, may then also be enhanced. This might be achieved through greater structured neuronal activity representing the spatial locus of visual attention. In this way, spatial attention may contribute to feature binding at those specific locations. Spatial visual attention, whether at the fovea or some parafoveal location, may thus play a role in reducing the capacity needed to process natural images. Nevertheless, unlike feature integration theory, our model does allow for a degree of parallel feature binding some distance from the spatial locus of attention.

Feedforward Projection of Information About Low-Level Visual Features to Higher Neuronal Layers

In the simulations presented in this article, we showed that visual information about low-level features was, in fact, being propagated up through the neuronal layers of the network in a similar fashion to that illustrated in Figure 5a. This kind of feedforward propagation of low-level visual information may be important if the behavior-related areas of the brain are restricted to reading out visual information from only the higher processing stages of the visual system. As discussed above, low-level visual features such as oriented bars and edges are represented in the earliest cortical stages (e.g., V1 and V2) of visual processing. However, when we perceive an object, we are aware of visual features at every spatial scale and complexity of visual form. The simulations reported in this article show how all of this visual information could, in principle, be projected upward through successive stages of visual processing. In particular, the neural representation of a low-level feature in the higher layers of the network encodes both the identity of the low-level feature as well as its local image context in terms of hierarchical binding relationships to higher level features. For example, in Figure 5a, Binding Neuron 3 represents the fact that there is a vertical bar in some localized region of the retina and that this vertical bar is part of the alphabetic letter T.

The bottom-up projection of low-level visual information through successive layers of visual processing is an automatic consequence of our hypothesized solution to the feature binding problem using polychronization and the emergence of binding neurons. The most simple mechanism for achieving the bottom-up projection of low-level visual information is illustrated in Figure 5a. The mechanism is essentially the same as the three-neuron binding circuit shown in Figure 3a but with the Binding Neuron 3 situated in the same higher layer as Neuron 2 that represents the high-level feature T. As described, Binding Neuron 3 represents that there is a vertical bar in some local region of the retina and that this vertical bar is part of the letter T. Thus, Figure 5a shows how information about the presence of a low-level feature (i.e., there is a vertical bar in some localized region of the retina) in the lower layer has been projected up to the higher layer along with its local image context (i.e., the vertical bar is part of the letter T). This proposed mechanism for the bottom-up projection of information about low-level features could operate through successive cortical stages of visual processing, including across the visual field and at every spatial scale.

Simulations of the full network architecture (FF + FB + LAT) provided examples of this kind of feedforward propagation of visual information. The binding neurons presented in rows "a" and "c" of Figure 12 were, in fact, in Layer 4. Thus, in each of these examples, the low-level feature neuron was in Layer 3, the high-level feature neuron was in Layer 4, and the binding neuron was also in Layer 4. In these cases, information about the low-level feature represented in Layer 3, including its local image context (i.e., that the low-level feature represented in Layer 3 is part of the high-level feature represented in Layer 4), is projected onto the binding neuron in Layer 4. These simulation results confirm the feasibility of the hypothesis that low-level visual information is propagated forward (i.e.,

bottom-up) to higher layers in the manner proposed in Feedforward projection of information about low-level visual features to higher neuronal layers.

Figure 5b shows how the basic mechanism illustrated in Figure 5a could be repeated iteratively up through successive layers in order to project information about low-level features into the highest (output) layer of the network. In Figure 5b, visual information about the presence of a vertical bar is first projected up from the first neuronal layer to the second layer, where it is represented by Binding Neuron 3. Neuron 3 represents the fact that there is a vertical bar in some localized region of the retina and that this vertical bar is part of the alphabetic letter T. Then, a similar binding mechanism combines the output from Binding Neuron 3 with the output of Neuron 5 representing a cat, where these combined outputs drive Binding Neuron 6. Binding Neuron 6 then represents the fact that there is a vertical bar in a local region of the retina, which is part of the letter T, which, in turn, is part of the word CAT. In this way, the information about the lowest level feature is projected upward and preserved in the highest layer of the network. Indeed, it is possible that a large amount of information about low-level features could be projected upward in this manner and preserved in the highest layers for readout by subsequent behavioral systems. We refer to this as a *holographic principle* because information about visual features at every level of complexity and scale may be preserved in the highest (output) layer(s) of the network. In using the term *holographic principle*, we are conscious of a somewhat similar usage to describe the preservation of information at the event horizon surface of a black hole (Susskind, 1995).

It is important to note that the Binding Neurons 3 and 6 in the highest layers of the two network architectures shown in Figures 5a and 5b represent the presence of a vertical bar in some local region of the retina that is explicitly part of a higher level feature/object (e.g., the letter T) or hierarchy of features (e.g., the letter T, which is part of the word CAT). Consequently, such binding neurons do not simply respond to the presence of a vertical bar at some retinal location regardless of local image context (i.e., the higher level features/objects that the vertical bar is part of). The high-level feature/object still needs to be presented to the network in order to elicit a response from these kinds of binding neuron in the upper layers. The holographic principle described here is thus consistent with neurophysiological observations that neurons in the later stages of the ventral visual pathway tend to respond to more complex visual forms than the simple oriented bars represented in early cortical stages such as V1 and V2.

Lastly, the bottom-up projection of information about low-level visual features, as illustrated in Figure 5, would seem to negate the need for top-down synaptic connections in the network architecture. This presents something of a conundrum. If the holographic principle holds true in some way, then we will need to develop a theory of how top-down signal transmission fits into this framework. This might lead to much greater complexity in visual processing than so far considered here. However, we believe that the observed architecture and neurodynamics of the visual cortex provide the necessary signposts for eventually understanding and simulating the singular semantic richness of biological vision.

Future Work

The kind of spiking network architecture discussed in this article seems to be needed to solve the feature binding problem. In rate-coded models, such as our laboratory's own VisNet model, individual postsynaptic neurons do not record which subset of presynaptic neurons are actually driving them, and, consequently, the network as a whole does not maintain an explicit representation of which presynaptic neurons are driving particular postsynaptic neurons throughout the network. Thus, in a sense, the rate-coded network "leaks" this essential binding information, which is necessary for representing, and making sense of, how the visual features within a scene are related to each other. This is particularly problematic when postsynaptic neurons represent high-level visual features (such as a complex visual form, object, or face) with some degree of transform (e.g., location, view or scale) invariance, as is typical in the higher layers of the primate ventral visual pathway (Booth & Rolls, 1998; Perry, Rolls, & Stringer, 2010; Wallis & Rolls, 1997). It is particularly in this situation that the network needs to maintain a representation of exactly which presynaptic neurons are driving each postsynaptic neuron in order to represent the relationships between the lower level and higher level features throughout the visual field and at every spatial scale. In the current article, we have not trained the network to develop transform (e.g., location) invariant responses to the visual stimuli. However, we plan to do this in future work when modeling the development and operation of binding neurons.

Further theoretical evidence that rate coding may be insufficient to solve feature binding has been provided by Eguchi and Stringer (2016), who demonstrated the failure of binding within a rate-coded model of *border ownership cells*. This class of visual cells, which have been found in cortical areas V1 and V2, respond to oriented edges like simple cells but are also modulated by which side of an object the edge occurs on (Zhou, Friedman, & von der Heydt, 2000). Such border ownership cells are clearly modulated by top-down visual signals about local object context from outside their classical receptive field. Importantly, border ownership cells are thought to represent the binding relationship between a localized border edge region of an object and the object itself. Eguchi and Stringer provided a detailed argument for why their rate-coded model of border ownership cells failed on binding when more than one object was presented to the network at a time and also proposed that spiking dynamics would be needed to solve this problem. In future work, we will explore how border ownership cells may develop in the kind of spiking network model investigated in this article, where the border ownership cells are perhaps examples of our hypothesized binding neurons.

A particularly interesting feature of the proposed theories in this article is that it potentially reveals a sharp contrast between processing in the visual brain and the operation of biologically implausible rate-coded neural network algorithms such as backpropagation of error. The architecture of the visual cortex, which is simulated in the spiking neural network models presented in this article, could potentially enable the development of binding neurons that represent the binding relationships between low-level and high-level features at all spatial scales throughout a visual scene. However, a biologically implausible neural network algorithm such as rate-coded backpropagation of error (Hertz, Krogh, &

Palmer, 1991) would not develop binding neurons and so could not represent such binding information. That is, although rate-coded networks (trained by backpropagation of error or otherwise) might be efficient at learning arbitrary mappings, they would not be able to represent the essential binding information needed to semantically analyze natural visuospatial scenes in the same way as the primate brain.

As a first step toward this, in future work, we propose developing *hybrid* neural networks that combine the kind of biologically inspired spiking (unsupervised learning) network presented in this article with a more traditional engineering (supervised learning) network such as backpropagation of error. In such a hybrid network, the biological network may operate as a preprocessing stage that extracts not only the visual features but also the binding relationships between those features across the visual field and at every spatial scale. All of this visual information may then be propagated from the biological network to the engineering network for, say, image classification.

References

Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. New York, NY: Cambridge University Press.

Akolkar, H., Meyer, C., Clady, X., Marre, O., Bartolozzi, C., Panzeri, S., & Benosman, R. (2015). What can neuromorphic event-driven precise timing add to spike-based pattern recognition? *Neural Computation, 27,* 561–593.

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex, 7,* 237–252.

Bi, G.-Q. and Poo, M.-M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience, 18,* 10464–10472.

Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex, 8,* 510–523.

Cumming, B. G., & Parker, A. J. (1999). Binocular neurons in v1 of awake monkeys are selective for absolute, not relative, disparity. *The Journal of Neuroscience, 19,* 5602–5618.

Deger, M., Helias, M., Rotter, S., & Diesmann, M. (2012). Spike-timing dependence of structural plasticity explains cooperative synapse formation in the neocortex. *PLoS Computational Biology, 8,* e1002689.

Diekman, C., Dasgupta, K., Nair, V., & Unnikrishnan, K. P. (2014). Discovering functional neuronal connectivity from serial patterns in spike train data. *Neural Computation, 26,* 1263–1297.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Eguchi, A., Humphreys, G. W., & Stringer, S. M. (2016). The visually-guided development of facial representations in the primate ventral visual pathway: A computer modelling study. *Psychological Review, 123,* 696–739.

Eguchi, A., Mender, B. M. W., Evans, B., Humphreys, G., & Stringer, S. (2015). Computational modeling of the neural representation of object shape in the primate ventral visual system. *Frontiers in Computational Neuroscience, 9,* 100.

Eguchi, A., & Stringer, S. M. (2016). Neural network model develops border ownership representation through visually guided learning. *Neurobiology of Learning and Memory, 136,* 147–165.

Evans, B. D., & Stringer, S. M. (2012). Transformation-invariant visual representations in self-organizing spiking neural networks. *Frontiers in Computational Neuroscience, 6,* 46.

Evans, B. D., & Stringer, S. M. (2013). How lateral connections and spiking dynamics may separate multiple objects moving together. *PLoS ONE, 8*(8), e69952.

Fares, T., & Stepanyants, A. (2009). Cooperative synapse formation in the neocortex. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 16463–16468.

Fujii, H., Ito, H., Aihara, K., Ichinose, N., & Tsukada, M. (1996). Dynamical cell assembly hypothesis: Theoretical possibility of spatio-temporal coding in the cortex. *Neural Networks, 9,* 1303–1350.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36,* 193–202.

Galeazzi, J. M., Minini, L., & Stringer, S. (2015). The development of hand-centred visual representations in the primate brain: A computer modelling study using natural visual scenes. *Frontiers in Computational Neuroscience, 9,* 147.

Hertz, J. A., Krogh, A. S., & Palmer, R. G. (1991). Introduction To The Theory Of Neural Computation. *Physics Today, 44,* 12.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation, 18,* 245–282.

Izhikevich, E. M., Gally, J. A., & Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex, 14,* 933–944.

Jeanson, F. (2011). Coincidence detection: Towards an alternative to synaptic plasticity. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 33,* 976.

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58,* 1187–1211.

Kreiter, A. K., & Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *Journal of Neuroscience, 16,* 2381–2396.

Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42,* 300–311.

Lindsey, B. G., Morris, K. F., Shannon, R., & Gerstein, G. L. (1997). Repeated patterns of distributed synchrony in neuronal assemblies. *Journal of Neurophysiology, 78,* 1714–1719.

Mao, B.-Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C., & Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices. *Neuron, 32,* 883–898.

Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science, 275,* 213–215.

Martinez, R., & Paugam-Moisy, H. (2009). Algorithms for structural and dynamical polychronous groups detection. In C. Alippi, M. Polycarpou, C. Panayiotou, & G. Ellinas (Eds.), *Artificial neural networks – ICANN 2009. Lecture Notes in Computer Science* (Vol. 5769, pp. 75–84). Berlin, Germany: Springer.

McCormick, D. A., Connors, B. W., Lighthall, J. W., & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology, 54,* 782–806.

Nikolic, D., Fries, P., & Singer, W. (2013). Gamma oscillations: Precise temporal coordination without a metronome. *Trends in Cognitive Sciences, 17,* 54–55.

Paugam-Moisy, H., Martinez, R., & Bengio, S. (2008). Delay learning and polychronization for reservoir computing. *Neurocomputing, 71,* 1143–1158.

Perrinet, L., Delorme, A., Samuelides, M., & Thorpe, S. J. (2001). Networks of integrate-and-fire neuron using rank order coding A: How to implement spike time dependent Hebbian plasticity. *Neurocomputing, 38–40,* 817–822.

Perry, G., Rolls, E. T., & Stringer, S. M. (2010). Continuous transformation learning of translation invariant representations. *Experimental Brain Research, 204,* 255–270.

Petkov, N., & Kruizinga, P. (1997). Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: Bar and grating cells. *Biological Cybernetics, 76,* 83–96.

Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H., & Abeles, M. (1998). Spatiotemporal structure of cortical activity: Properties and behavioral relevance. *Journal of Neurophysiology, 79,* 2857–2874.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Rolls, E. T., & Milward, T. (2000). A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation, 12,* 2547–2572.

Rosenblatt, F. (1961). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan Books.

Softky, W. R. (1995). Simple codes versus efficient codes. *Current Opinion in Neurobiology, 5,* 239–247.

Susskind, L. (1995). The world as a hologram. *Journal of Mathematical Physics, 36,* 6377–6396.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Troyer, T. W., Krukowski, A. E., Priebe, N. J., & Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: Thalamocortical input tuning and correlation-based intracortical connectivity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 18,* 5908–5927.

van Rossum, M. C., Bi, G. Q., & Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 20,* 8812–8821.

von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron, 24,* 95–104.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51,* 167–194.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24,* 11–17.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience, 20,* 6594–6611.

Appendix

Data Sharing

The SPIKE simulator can be downloaded from http://oftnai.github.io/Spike/.

Received January 31, 2017
Revision received January 9, 2018
Accepted January 10, 2018 ■