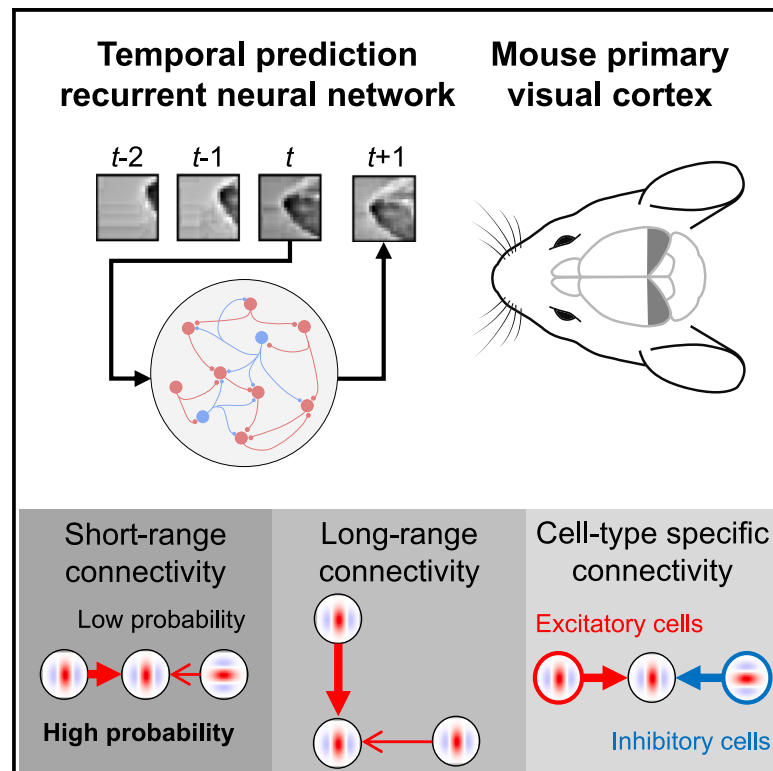


Prediction of future input explains lateral connectivity in primary visual cortex

Graphical abstract



Authors

Sebastian Klavinskis-Whiting,
Emil Fristed, Yosef Singer,
M. Florencia Iacaruso, Andrew J. King,
Nicol S. Harper

Correspondence

sebastian.klavinskis-whiting@chch.ox.ac.uk (S.K.-W.),
andrew.king@dpag.ox.ac.uk (A.J.K.),
nicol.harper@dpag.ox.ac.uk (N.S.H.)

In brief

Klavinskis-Whiting et al. show that feature-specific connectivity patterns in mouse primary visual cortex are captured by simple recurrent neural networks trained to predict their future inputs. This temporal prediction model accounts for a range of observed connectivity motifs, providing a unifying explanation for cortical functional connectivity.

Highlights

- Neurons in primary visual cortex show feature-specific connectivity patterns
- Recurrent networks trained for temporal prediction recapitulate this connectivity
- Temporal prediction outperforms models based on alternative training objectives
- The connectivity of early visual cortex may be optimized for temporal prediction



Article

Prediction of future input explains lateral connectivity in primary visual cortex

Sebastian Klavinskis-Whiting,^{1,2,*} Emil Fristed,^{1,2} Yosef Singer,¹ M. Florencia Iacaruso,^{1,4} Andrew J. King,^{1,5,*} and Nicol S. Harper^{1,3,*}

¹Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford OX1 3PT, UK

²These authors contributed equally

³Senior author

⁴Present address: The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

⁵Lead contact

*Correspondence: sebastian.klavinskis-whiting@chch.ox.ac.uk (S.K.-W.), andrew.king@dpag.ox.ac.uk (A.J.K.), nicol.harper@dpag.ox.ac.uk (N.S.H.)

<https://doi.org/10.1016/j.cub.2024.11.073>

SUMMARY

Neurons in primary visual cortex (V1) show a remarkable functional specificity in their pre- and postsynaptic partners. Recent work has revealed a variety of wiring biases describing how the short- and long-range connections of V1 neurons relate to their tuning properties. However, it is less clear whether these connectivity rules are based on some underlying principle of cortical organization. Here, we show that the functional specificity of V1 connections emerges naturally in a recurrent neural network optimized to predict upcoming sensory inputs for natural visual stimuli. This temporal prediction model reproduces the complex relationships between the connectivity of V1 neurons and their orientation and direction preferences, the tendency of highly connected neurons to respond more similarly to natural movies, and differences in the functional connectivity of excitatory and inhibitory V1 populations. Together, these findings provide a principled explanation for the functional and anatomical properties of early sensory cortex.

INTRODUCTION

An increasing number of studies—mostly focusing on mouse primary visual cortex (V1)—have begun to uncover the underlying rules specifying how cortical neurons connect.^{1–4} Some findings, such as the tendency of V1 neurons to synapse with other neurons that show similar orientation selectivity, follow a simple like-for-like pattern.^{2–4} By contrast, other results, such as the spatial organization of synaptic inputs to orientation- and direction-tuned visual neurons, appear more complex and less amenable to a unifying theoretical explanation.^{1,5} An outstanding question, then, is how to understand these putative connectivity rules and whether they can be explained by a single general principle.

By taking a normative approach, we can ask whether the patterns of structure and function observed at the level of individual neurons are optimized for achieving a particular goal that is important behaviorally or from an evolutionary perspective. One such promising normative framework is that of temporal prediction, which posits that sensory systems are optimized to represent those features in the recent past that are predictive of the immediate future sensory inputs.^{6,7} Why might temporal prediction serve as a useful objective for an organism's sensory systems? First, if sensory systems construct a model of the world, then a good model should predict future inputs well.^{8,9} Second, predictive features are valuable for guiding actions and for compensating for neural conduction and processing delays,¹⁰ enabling, for example, a cat to catch a bird in flight. Third, extracting predictive features reduces the vast amount of

information the brain needs to manage. Finally, temporal prediction requires no explicit teaching signal beyond the sensory input itself, making it inherently more biologically plausible as an unsupervised principle than other supervised counterparts.^{11,12}

When optimized for temporal prediction, feedforward networks have been shown to capture many of the receptive field characteristics and response properties of V1 neurons, as well as motion processing along the visual pathway.^{6,7} Nevertheless, existing feedforward temporal prediction models neglect the role of recurrency, which experimental and theoretical studies have implicated in a range of key brain functions.^{13–15}

Here, we show that a recurrent network optimized for temporal prediction on dynamic natural visual scenes can capture many motifs of local connectivity in visual cortex. Furthermore, when we compared network models optimized for different normative objectives, temporal prediction stood out in its capacity to explain these connectivity motifs. Hence, the relationship between the connectivity patterns of V1 neurons and their response characteristics appears to be optimized to support the predictive processing of dynamic stimuli.

RESULTS

Model response properties

We trained a recurrent network model to predict the upcoming visual input (40 ms ahead) based on the recent stimulus history (Figure 1A). The model was trained on a diverse dataset of movies of natural scenes, including wildlife and panning over



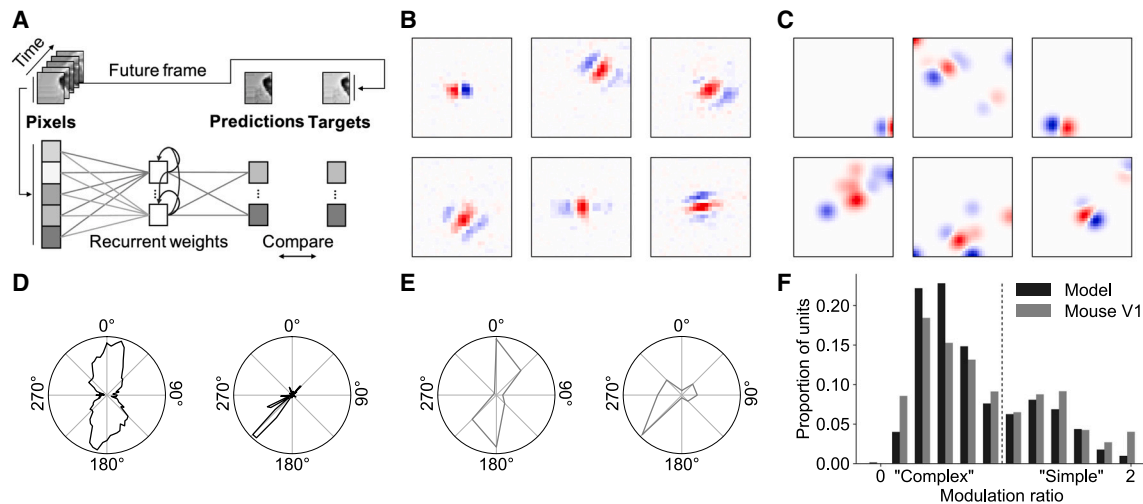


Figure 1. The recurrent temporal prediction model captures basic tuning properties of V1 neurons

(A) Schematic of the recurrent temporal model. 2,332 hidden units (90%) were excitatory with non-negative outgoing recurrent weights, and the remaining 260 hidden units (10%) were inhibitory with non-positive outgoing recurrent weights.

(B) Response-weighted-average receptive field estimates of model units (units in pixels).

(C) Mouse V1 receptive fields from publicly available recordings,¹⁷ pre-processed for visualization by thresholding and smoothing with a Gaussian filter (units in pixels).

(D) Example tuning curves for orientation (left) and direction (right) tuned model units. Response is the normalized hidden unit activity as a function of the stimulus value.

(E) Example tuning curves for orientation- and direction-tuned cells in V1.¹⁷ Response is the normalized firing rate as a function of the stimulus value.

(F) Distribution of modulation values across model and pooled excitatory and inhibitory mouse V1 units¹⁸; typically a modulation ratio <1 is taken as a complex cell and >1 as a simple cell.

See also Figures S1 and S2.

natural environments, which were bandpass filtered to approximate the retina.¹⁶ After training, we first compared model hidden unit response properties with those of neurons in mouse V1.

We estimated the model units' receptive fields by means of the response-weighted average (Figure 1B). Like simple cells in V1 (Figure 1C), model units often had well-defined receptive fields with a Gabor-like structure consisting of oriented excitatory and inhibitory subfields. To probe the tuning properties of the model units, we recorded the model's responses to oriented full-field drifting gratings. Model units were generally orientation tuned (24%) or direction tuned (57%; Figures 1D and 1E), with a similar distribution of orientation selectivity indices (OSI) and direction selectivity indices (DSI) as found in mouse V1 (Figure S1). As in visual cortex, model units varied in their phase responsiveness, characterized by their modulation ratio (Figure 1F). Highly modulated units were classed as simple-cell like, while units that displayed little or no phase modulation were classed as complex-cell like.^{19,20} At the population level, there was a bimodal distribution similar to that found in mouse V1,^{18,21} indicating the existence of distinct populations of simple-cell-like and complex-cell-like model units (Figure 1F).

Some neurons in V1 are thought to encode prediction errors.^{22,23} While the model hidden units do not explicitly represent prediction errors by default, they could emerge spontaneously during the inference process in response to "unexpected" stimuli that violate a recurring pattern (Figure S2A). We found a small number of units that responded selectively to these unexpected stimuli (omitted stimuli: 0.2% of units; deviant stimuli: 5.7% of units), while a slightly larger number of model units showed

mixed sensitivity and responded to both standard and unexpected stimuli (omitted stimuli: 0.5% of units; deviant stimuli: 9% of units). Prediction-error responses have been reported in some neurons in mouse V1^{23–25} and to varying degrees in other primary sensory areas.^{26–30} Moreover, this mixed sensitivity aligns with the biology where neurons rarely exist in a strictly error-coding capacity.^{25,31}

Short- and long-range functional connectivity

Short-range functional connectivity between excitatory units in the model resembled that of mouse V1 neurons (Figure 2A).^{1,3} Model units were more likely to synapse with other units with the same orientation preference (Figure 2B), with the connection probability monotonically decreasing as the difference in orientation preference increased ($p < 0.0001$, Cochran-Armitage). Likewise, connectivity between direction-tuned excitatory model units resembled that of mouse V1 neurons.³ Model units were most likely to connect when tuned to either the same or opposite direction of motion (both $p < 0.0001$, Cochran-Armitage), with connection probability decreasing as the presynaptic unit's tuning became more orthogonal to the postsynaptic unit's preferred direction (Figure 2C).

For inhibitory model units, co-tuning with the postsynaptic unit was much weaker (Figure 2D), as has been previously reported in V1.^{5,32,33} Neither inhibitory-to-excitatory ($p = 0.102$, Cochran-Armitage) nor inhibitory-to-inhibitory ($p = 0.606$, Cochran-Armitage) model unit connections showed a significant linear dependence of connection probability on the difference in orientation preference. For direction-tuned inhibitory-to-excitatory model

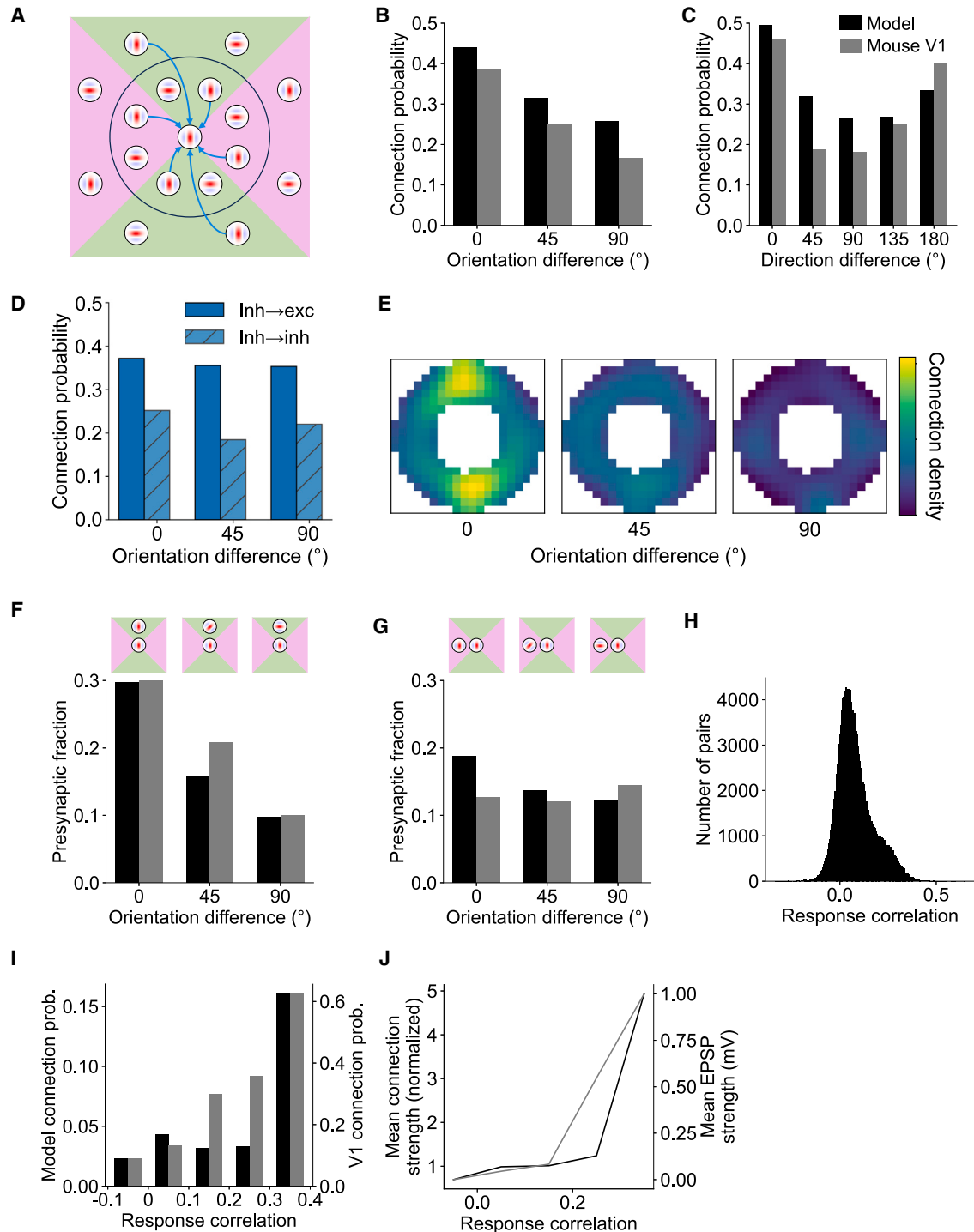


Figure 2. The model captures short- and long-range functional connectivity in V1

(A) Schematic of short- and long-range connectivity in V1. Short-range connections are more prevalent for similarly tuned V1 neurons, whereas long-range connection probability is greater for V1 neurons with similar orientation-tuning when their receptive fields are located in co-axial space.

(B and C) Short-range connections are more prevalent when excitatory model units have similar orientation tuning (B) and for direction-tuned units that have similar or opposite preferred directions of motion (C), as is also the case in V1.³

(D) As in (B), but for inhibitory-to-inhibitory and inhibitory-to-excitatory connections in the model.

(legend continued on next page)

units, the model predicts a weak but significant monotonic trend of increasing connection probability as the difference in preferred direction increases (Figure S3; $p = 0.032$, Cochran-Armitage), distinct from the u-shaped trend observed for excitatory model units and excitatory V1 neurons. No significant trend was found for inhibitory-to-inhibitory direction-tuned model units ($p = 0.484$, Cochran-Armitage). Finally, for orientation- and direction-tuned excitatory-to-inhibitory model units, the model predicts a similar trend as for excitatory-to-excitatory model units and V1 neurons ($p < 0.0001$, Cochran-Armitage), albeit with the minimum connection probability shifted to the 135° bin for direction-tuned model units (Figure S3).

We further investigated how connectivity differs among units with simple-cell-like versus complex-cell-like responses (Figures S3E and S3F). Overall, we found few differences between these two populations. For the short-range connectivity motifs investigated (Figures S3E and S3F), the connectivity pattern was qualitatively similar to the population distributions described above (Figures 2B and 2C), though the aggregate connection probability was higher for simple-cell-like units compared with complex-cell-like units.

Across long-range connections, the model also replicated the dependence of connectivity on neuronal orientation preference and receptive field location found in visual cortex.¹ We measured the connection probability between pre- and postsynaptic model units as a function of their difference in preferred orientation and the presynaptic unit's receptive field location in visual space relative to that of the postsynaptic unit (Figure 2E). As for mouse V1, model units were more likely to project to the postsynaptic unit if their receptive field aligned along the axis of the postsynaptic unit's receptive field. To quantify this effect, we divided visual space relative to the postsynaptic unit into four quadrants. Those quadrants that aligned with the postsynaptic unit's orientation tuning were referred to as "co-axial" space (green regions in Figure 2A), while those quadrants orthogonal to the unit's preferred orientation were referred to as "co-orthogonal" space (pink regions in Figure 2A). As with the biology, orientation-tuned model units were more likely to synapse with other units when they had similar orientation selectivity and their receptive fields were located in co-axial visual space (Figure 2F; $p < 0.0001$, permutation test). Although a similar effect was found for co-orthogonal units (Figure 2G; $p < 0.0001$, permutation test), this relationship was much weaker. In particular, there was a significantly higher proportion of model units in the 0° orientation difference bin and a significantly lower proportion in the 90° bin for receptive fields in co-axial versus co-orthogonal space (both $p < 0.001$, permutation test).

As in V1,^{2,4} model units whose responses to natural movies were highly correlated were also more likely to be connected. In both cases, the distribution of correlation values between pairs of connected model units was skewed, with the majority of pairs of units showing a relatively low response correlation and a smaller proportion that were highly correlated (Figure 2H). A qualitatively similar relationship was seen for both model units and V1 neurons between the response correlation and both the connection probability (Figure 2I; $p < 0.001$, Cochran-Armitage) and the average strength of those connections (Figure 2J).

Excitatory-inhibitory functional connectivity of direction-tuned units

As for orientation-tuned cells, synaptic inputs to direction-tuned cells in V1⁵ are not uniformly distributed in visual space (Figure 3A). In particular, direction-tuned excitatory cells preferentially receive inputs from other excitatory cells whose receptive fields are situated in the opposite sector of visual space (i.e., behind) the postsynaptic cell's preferred direction. By contrast, the opposite effect is observed for inhibitory cells, which preferentially synapse with excitatory cells if the location of their receptive fields is ahead of the postsynaptic cell's preferred direction of motion. This connectivity motif provides a plausible circuit basis for direction selectivity, whereby a spatial offset combined with a conductance delay for inhibitory cells facilitates the detection of moving stimuli.⁵

In line with this evidence from V1, excitatory presynaptic ensembles in the model were more numerous in the opposite sector relative to the postsynaptic unit's preferred motion direction (Figures 3B–3E; $t(113) = 2.93$, $p = 0.004$). Conversely, the opposite pattern was found for the inhibitory model units, whose connection probability with direction-tuned excitatory units was higher if their receptive fields were located in the sector of visual space ahead of the preferred direction of motion (Figures 3B–3E; $t(113) = -2.80$, $p = 0.006$). This effect was specific to direction-tuned excitatory model units, with no significant difference in the spatial locations of excitatory and inhibitory presynaptic ensembles synapsing with weakly direction-selective units, as defined by a DSI ≤ 0.8 (see STAR Methods; Figure 3E; $t(113) = 225$, $p = 0.337$).

Comparing V1 response prediction and connectivity across models

To assess how well temporal prediction performed as a normative model of mouse V1, we examined how well it captured the properties of V1 neurons compared with several other commonly used models (see STAR Methods). First, to directly compare

(E–G) In both the model and V1,¹ long-range connection probability is higher for presynaptic model units with similar orientation preferences when their receptive fields are located in co-axial (F) than in co-orthogonal (G) locations relative to the receptive field of the postsynaptic unit. For (B), (C), (F), and (G), data are binned using the same convention as in Ko et al.³ with orientation bins of 0° – 22.5° , 22.5° – 67.5° , and 67.5° – 90° and motion direction bins of 0° – 22.5° , 22.5° – 67.5° , etc. Heatmap (E) shows the normalized connection probability over visual space across differences in orientation tuning for model units. Heatmap is smoothed for display purposes with a Gaussian filter (standard deviation, $\sigma = 2$ pixels).

(H) Histogram of the response correlation distribution across pairs of connected model units for natural stimuli. The distribution is right skewed, indicating that a minority of model units have highly correlated responses.

(I and J) As for mouse V1,² response correlation for model units co-varies with the connection probability (I) as well as the input connection strength (J). These results were abolished after randomly shuffling the recurrent weights between the units when measuring connectivity, resulting in uniform distributions (Figure S4). Accordingly, the model connectivity biases cannot be explained by the underlying distribution of orientation- and direction-tuning preferences among model units.

See also Figures S2–S4.

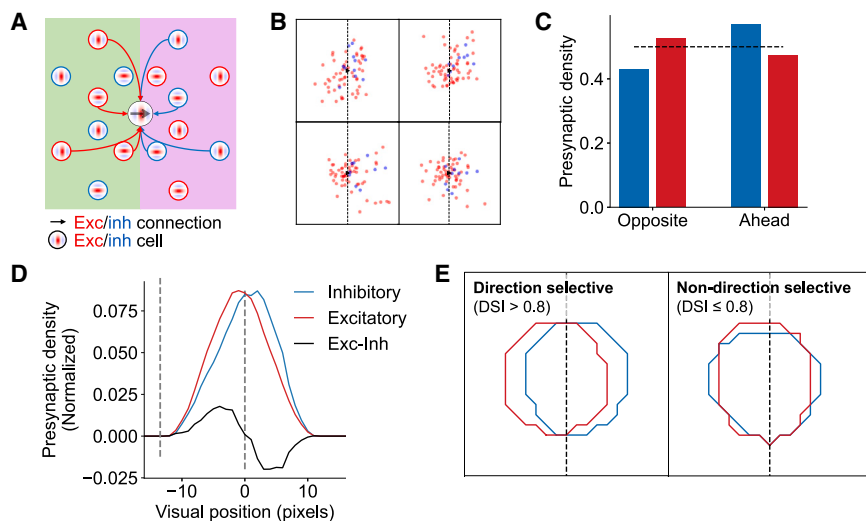


Figure 3. The model captures direction-dependent differences in functional connectivity between excitatory and inhibitory populations in V1

(A) Schematic of connectivity biases in excitatory and inhibitory inputs to direction-tuned cells in V1.⁵

(B) Exemplar excitatory and inhibitory presynaptic ensembles for direction-tuned excitatory model units.

(C) Model unit presynaptic density for excitatory and inhibitory subpopulations in the sectors of visual space opposite to and ahead of the postsynaptic unit's preferred direction of motion. Dashed line represents equal density (0.5).

(D) Profile of model presynaptic unit location density across horizontal visual space for excitatory and inhibitory inputs. Profiles smoothed with a 5-pixel moving average.

(E) Pooled location density contours over visual space across all excitatory (red) and inhibitory

(blue) model units for direction- and non-direction-selective postsynaptic excitatory units. These results were abolished after randomly shuffling the recurrent weights between the units when measuring connectivity, with no difference in density across either sector of visual space for excitatory or inhibitory model units (Figure S4).

See also Figure S4.

each model's learned representations with V1, we used regression to predict the responses of awake mouse V1 neurons from the model's hidden unit activity for two natural movie clips (Figure 4A). Second, we quantitatively compared model unit properties and connectivity biases to those found in V1. Overall, the recurrent temporal prediction model predicted the recorded neural responses well (mean normalized correlation coefficient $[CC_{norm}] = 0.242$) compared with the other models tested (Figures 4B and 4C) and best accounted for V1 connectivity motifs (Figure 4D).

The linear-nonlinear baseline model consisted of the same fitting procedure as the other models surveyed but was applied directly to the input stimuli and performed significantly worse than the temporal prediction model in predicting V1 responses ($CC_{norm} = 0.189$, $t(738) = -6.03$, $p < 0.0001$).

VGG-19 and PredNet are both published models that capture elements of V1 responses. VGG-19 is a deep feedforward network trained for object recognition,³⁴ while PredNet implements a form of predictive coding.³⁵ PredNet was similarly trained for next-frame prediction but, unlike temporal prediction, explicitly uses prediction errors during the inference process.³⁶ The temporal prediction model also exceeded the performance of PredNet (mean $CC_{norm} = 0.113$, $t(738) = -8.97$, $p < 0.0001$) but performed significantly worse than VGG-19 (mean $CC_{norm} = 0.296$; $t(738) = 5.40$, $p < 0.0001$). However, as a supervised model, VGG-19 is fitted using human-annotated image labels, providing it with constraints not available to other models. Thus, the model is in some senses descriptive rather than normative and not directly comparable to the other models, which were unsupervised.

We also considered two variants of the temporal prediction model, comparing an untrained network with random weights and a feedforward temporal prediction model. In both cases, there was a significant benefit from both training (untrained mean $CC_{norm} = 0.175$, $t(738) = -7.03$, $p < 0.0001$) and the addition of recurrency (feedforward mean $CC_{norm} = 0.184$, $t(738) =$

-5.74 , $p < 0.0001$), implying the importance of these features for modeling V1 neural responses.

Finally, we considered three models based on the same recurrent network architecture but trained using alternative learning objectives. The denoising network aimed to recover the original current frame from noise, the inpainting network aimed to predict the complete current frame given inputs with patches blanked out, and the sparse autoencoder aimed to reproduce the current frame while minimizing hidden unit activity. The recurrent temporal prediction model performed significantly better than the inpainting network (mean $CC_{norm} = 0.187$, $t(738) = -4.09$, $p < 0.0001$) and the sparse autoencoder network (mean $CC_{norm} = 0.206$, $t(738) = -2.31$, $p = 0.021$), and non-significantly better than the denoising network (mean $CC_{norm} = 0.222$, $t(743) = -1.63$, $p = 0.110$), highlighting the importance of the training objective over model architecture alone.

The distribution of orientation- and direction-selective model units varied substantially among these models (Figure 4C). The temporal prediction model most closely reproduced the overall distribution of unit types found in mouse V1,¹⁷ with comparable proportions of orientation-selective (V1 = 31%, temporal prediction = 24%), direction-selective (V1 = 39%, temporal prediction = 57%), and non-selective units (V1 = 30%, temporal prediction = 19%), albeit with an overrepresentation of direction-selective units. By contrast, the denoising and inpainting networks had a clear overrepresentation of orientation-selective units (62% and 97%, respectively) relative to mouse V1. Similarly, far fewer units met the criteria for direction selectivity across these alternative models, implying that their learned representations had no or only a weak temporal component.

In tandem, we compared how well each model recapitulated the connectivity biases found in mouse V1 (Figure 4D). For each model, we calculated a model connectivity score as the average correlation between the model and V1 connectivity distributions (Figures 2B, 2C, 2F, 2G, 2I, and 2J). Overall, the temporal prediction model (mean = 0.68) had a much closer

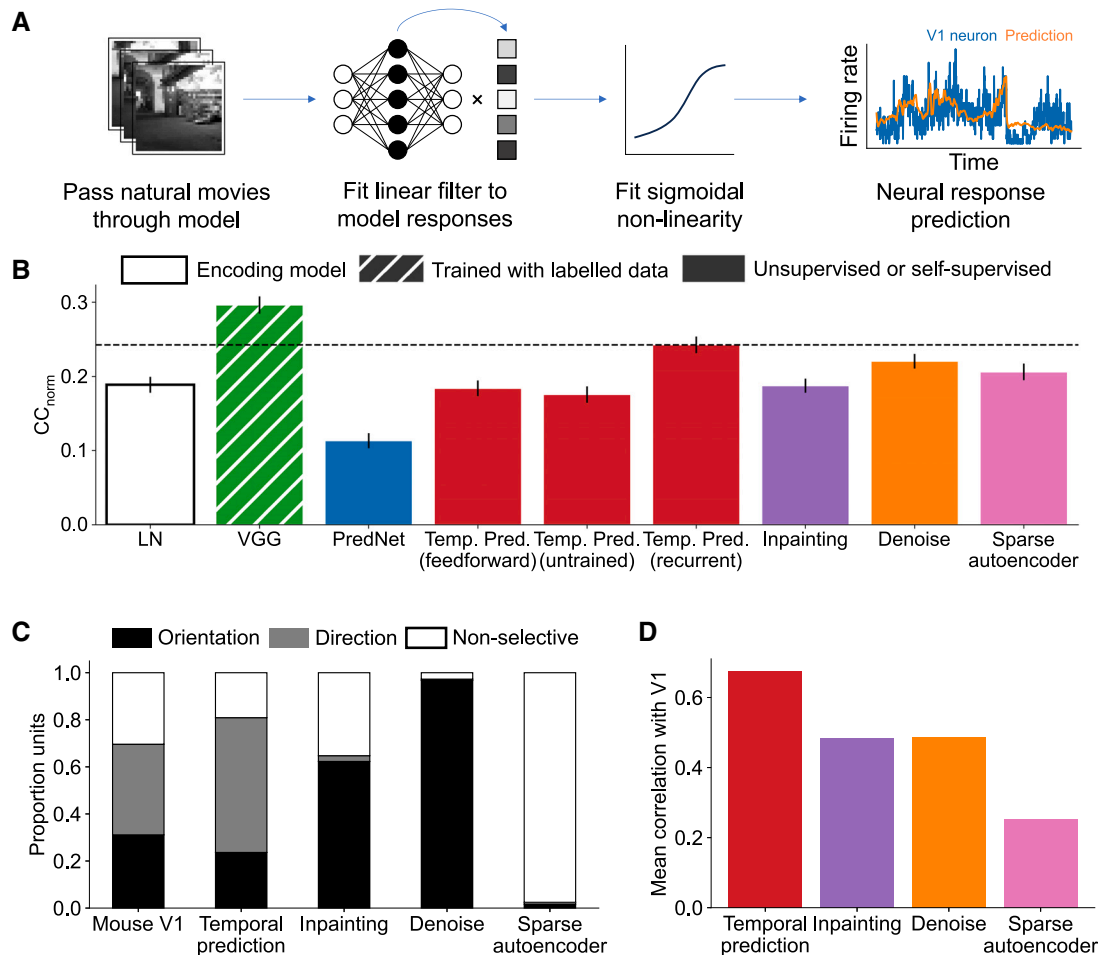


Figure 4. Comparison of neural prediction performance and model connectivity to V1 across alternative models

(A) Schematic of the neural response fitting procedure.

(B) Performance (average CC_{norm}) of the recurrent temporal prediction model (dashed line) relative to other comparison models in predicting neural responses to natural movies. Error bars indicate s.e.m.

(C) Distribution of orientation-, direction-, or non-selective units across mouse V1 and each model.

(D) Comparison for each model of the average correlation of the functional connectivity profiles (Figures 2B, 2C, 2F, 2G, 2I, and 2J) with those found in mouse V1. A higher correlation indicates an overall better fit with the V1 data. Note that VGG-19 and PredNet could not be included in the functional connectivity analysis. See also Figure S5.

correspondence to the connectivity profiles found in mouse V1 compared with the other models (in-painting mean = 0.48, denoising mean = 0.49, sparse autoencoder = 0.25). We found no significant correlation between each model's connectivity score and its neural prediction performance (CC_{norm} ; $r = 0.602$, $p = 0.398$). Thus, the capacity of a model to predict neural responses in V1 does not imply that it can accurately capture the underlying organization of cortical connectivity.

We additionally compared the learned connectivity of the PredNet model where possible (Figure S5). Although the model's structure precluded analysis of connectivity motifs dependent on visual space or excitatory and inhibitory subpopulations, we were able to analyze for the motifs described by Ko et al.³ and found that these were not well captured by PredNet. However, we were not able to look at lateral connectivity in the VGG model as it lacks recurrency. Hence, due to the limited or lacking lateral

connectivity of the VGG and PredNet models, we could not calculate their connectivity scores.

Variants of the temporal prediction model

We further explored how different modeling parameters impacted the capacity of the temporal prediction model to predict V1 connectivity.

We first investigated how the future prediction offset affected the model's connectivity motifs. Using a high frame-rate (120 Hz) dataset,³⁷ we produced a continuum of models trained to predict 0–10 frames (0–83 ms) into the future. As expected, we found that the aggregate connectivity score increased smoothly as a function of the future offset, to a maximum of 0.62 at 33 ms (Figure 5A). Similarly, if we trained the model to predict a span of offsets (i.e., the future frames at 0, 0–25, 0–45 ms, etc.), we found that the connectivity score similarly increased with the future

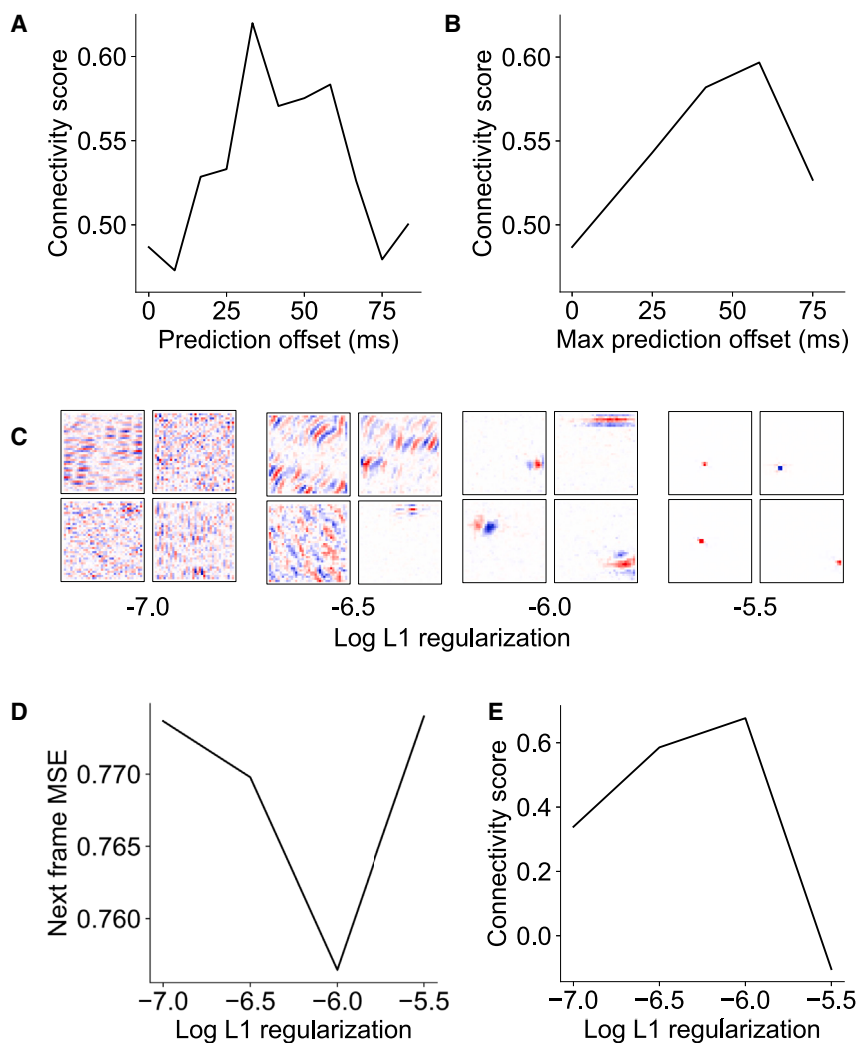


Figure 5. Variants of the temporal prediction model

(A) The connectivity score varies as a function of the future prediction offset, with a maximum value reached when predicting the next frame at 33 ms into the future.

(B) Connectivity score as a function of increasing future prediction span.

(C and D) Example unit receptive fields (C) and next-frame prediction mean squared error (D) as a function of L1 regularization. The optimal V1-like receptive fields coincide with the L1 regularization value that minimizes the mean squared error in the test set.

(E) Connectivity score as a function of the L1 regularization value, with the maximum again occurring at the optimal L1 regularization value that minimizes the mean squared error.

See also Figure S6.

further evidence for temporal prediction as a principle of neural sensory processing.

Model connectivity supports temporal prediction

To determine how the observed motifs are related to temporal prediction, we perturbed the network's orientation- and direction-dependent connectivity while measuring next-frame prediction performance (Figures 6A and 6B). To that end, we ablated (set to the median weight) an increasing number of connections in the model across different functional classes of units while measuring next-frame prediction performance. Given the finding in both V1 and the model that similarly tuned

units are more likely to be connected, we hypothesized that ablating connections between units with similar orientation or direction tuning ("co-tuned" units for orientation-selective units; "co/anti-tuned" units for direction-selective units) would result in a larger impairment in prediction performance than for units with a large orientation difference or more orthogonal difference in direction preference ("orthogonal" units).

offset, with a maximum score of 0.60 when predicting 0–58 ms into the future (Figure 5B). Notably, for the offset-span model, we found that units that integrated information further back into the past tended to project further into the future ($r = -0.15$, $p < 0.0001$; Figure S6). Thus, across these networks, as the prediction target was shifted away from an intermediate temporal offset, the connectivity score declined.

Finally, we investigated the role of the model's wiring constraint (L1 regularization). Varying L1 regularization had a marked effect on receptive field structure (Figures 5C and 5D). At the highest regularization strength, receptive fields were reminiscent of the center-surround organization of retinal receptive fields, while at weaker regularization strengths, receptive fields were only weakly spatially localized. Only at the optimal regularization strength for next-frame prediction did receptive fields resemble the Gabor-like structure of V1 neurons. A similar dependence on regularization strength was found for the aggregate connectivity score (Figure 5E), where the connectivity score peaked at the optimal L1 regularization strength, declining across neighboring strengths. That the most future-predictive hyperparameter setting is also the most brain-like provides

units are more likely to be connected, we hypothesized that ablating connections between units with similar orientation or direction tuning ("co-tuned" units for orientation-selective units; "co/anti-tuned" units for direction-selective units) would result in a larger impairment in prediction performance than for units with a large orientation difference or more orthogonal difference in direction preference ("orthogonal" units).

In support of our hypothesis, as the number of ablated connections increased, there was a relatively greater increase in the error when ablating connections between co-tuned and co/anti-tuned versus orthogonally tuned units (orientation: $t(1,998) = 290$, $p < 0.0001$; direction: $t(1,998) = 382$, $p < 0.0001$). Thus, ablating the connectivity motifs in the model described by Ko et al.⁹ specifically impaired temporal prediction.

From a more mechanistic perspective, we investigated how connectivity motifs may impact the network's internal representations. Considering the finding that units with similar orientation preferences are more likely to be connected, we hypothesized that this connectivity motif may help improve the discriminability of network representations. Within both the visual pathway and deep neural networks, one outcome of processing is to

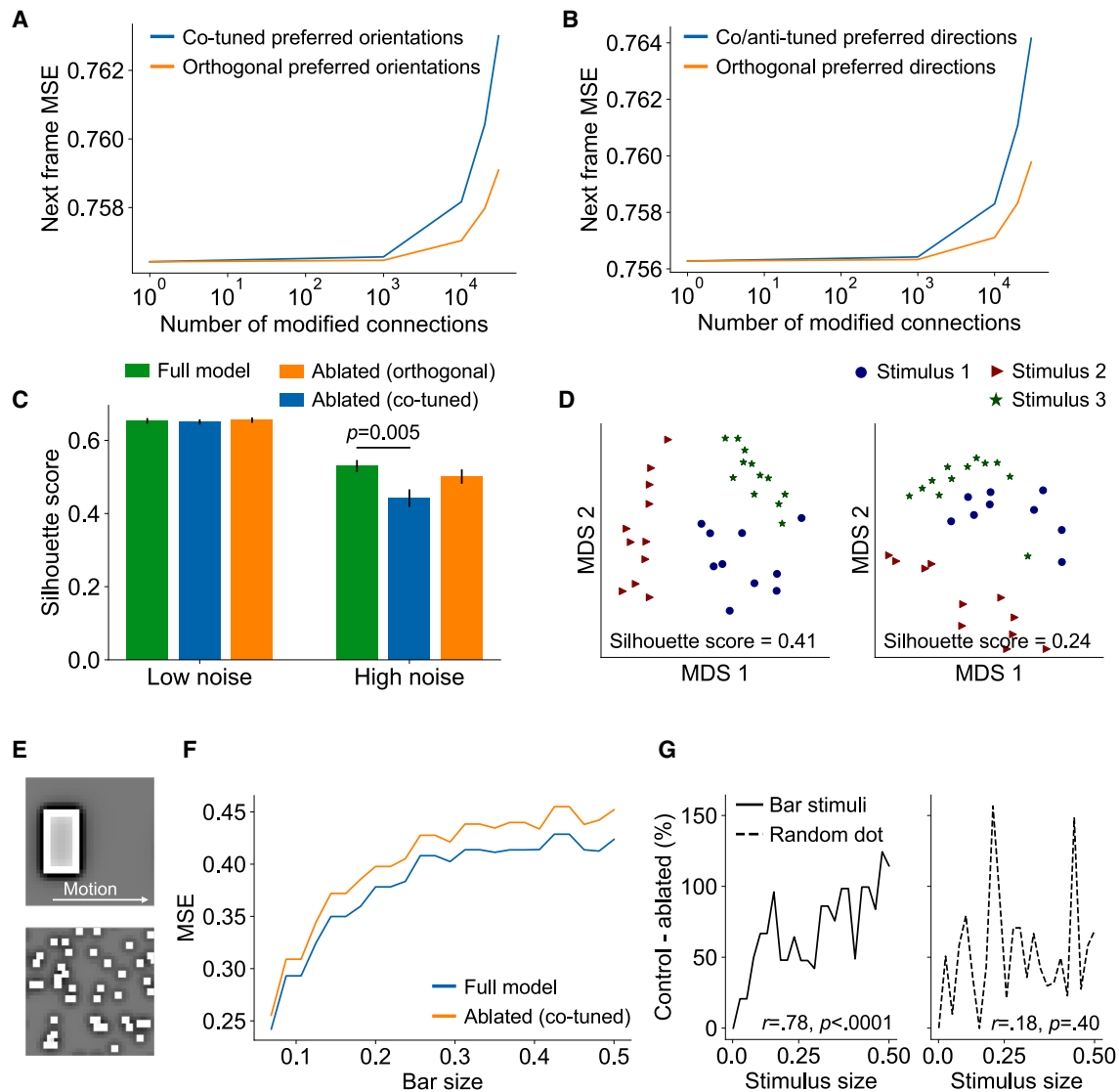


Figure 6. Model connectivity motifs support temporal prediction

(A and B) Ablating connections between connected units with similar preferences for stimulus orientation (co-tuned units) or similar and opposite preferences for direction of motion (co/anti-tuned units) impairs next-frame prediction more than for units with orthogonal orientation or direction tuning (orthogonal units).

(C) Clustering between network representations (measured via the silhouette score) is significantly lower under high noise when connections between pairs of units that are co-tuned (for orientation) are ablated. This suggests that this connectivity motif may improve the robustness of cortical representations of visual stimuli in noise. Error bars indicate s.e.m.

(D) Example plots of unit network activity projected onto two dimensions using multidimensional scaling. The silhouette scores illustrate the clustering across different stimulus types in the default network (left) or the network with ablated connections between co-tuned units (right).

(E) Illustration of the moving bar or moving random dot stimuli used to probe the network's next-frame prediction performance across different stimuli when disrupting specific connectivity motifs.

(F) The mean squared error for next-frame prediction in the ablated network increases relative to the control network as a function of bar length.

(G) Stimulus size, equated across both stimulus types, predicts the ablation deficit (the percentage increase in mean squared error for the ablated versus control networks) for bar but not random dot stimuli.

transform representations to facilitate linear readout to downstream regions.³⁸ In the case of next-frame prediction, transforming these representations—for example, to increase their separability—could facilitate model performance by helping to decompose visual inputs and thereby better capture their underlying causes.

To test this hypothesis, we presented oriented grating stimuli with different levels of noise while measuring how distinctly the network represented these stimulus classes in low-dimensional space. We used the silhouette score to measure clustering in the model's hidden activity³⁹ and compared these scores between the default and ablated networks (Figures 6C and 6D). Under

low-noise conditions (0 dB SNR), we did not find any difference in the silhouette score between the full and ablated networks (co-tuned: $t(46) = 0.240$, $p = 0.814$; orthogonal: $t(46) = -0.186$, $p = 0.853$). However, under high-noise conditions (−9.5 dB SNR), a higher silhouette score (indicating improved clustering) was found for the full network compared with the co-tuned but not orthogonal ablated networks (co-tuned: $t(46) = 2.94$, $p = 0.005$; orthogonal: $t(46) = 1.09$, $p = 0.282$). Thus, under conditions of high noise, connections between similarly orientation-tuned units may help disentangle the internal representations of different stimuli.

Finally, we examined the potential roles of the connectivity motifs described by Iacaruso et al.¹ in temporal prediction. We hypothesized that the tendency of V1 neurons to synapse with neurons of similar orientation tuning in co-axial space—that is, when their receptive fields are aligned in visual space—might facilitate the detection of contiguous moving edges. This should aid the network in next-frame prediction.

Under this hypothesis, we predicted that ablating connections between these model units would impair the next-frame prediction of contiguous-edge-like features (i.e., a moving bar). Furthermore, this impairment should increase with the bar length, under the assumption that this motif detects elongated features. As a control, we compared the bar stimulus with moving random dot stimuli, equating the total area of both (Figure 6E). Since random dot stimuli are non-contiguous, increasing the stimulus area (unlike bar length) should not affect the degree of impairment produced by ablating model connectivity. In line with this hypothesis, we found that stimulus area was correlated with the impact of ablations on mean squared error for the moving bar stimuli (Figure 6F; $r = 0.778$, $p < 0.0001$) but not random dot stimuli (Figure 6G; $r = 0.180$, $p = 0.400$).

Together, these results suggest that the identified connectivity motifs have specific functions in supporting temporal prediction.

DISCUSSION

The recurrent temporal prediction model exhibits response properties and functional connectivity patterns remarkably akin to those found in mouse V1, providing a unifying normative explanation for these wiring biases. In particular, the model captured the relationship between both short- and long-range connectivity patterns and neuronal preferences for stimulus orientation and direction of motion, as well as spatial differences in the inputs from excitatory and inhibitory cells to direction-selective cortical neurons.

The extent to which cortical circuits are fundamentally stereotyped remains an enduring question in systems neuroscience. The concept of a canonical microcircuit proposes that cortical networks follow the same basic organization in which functional differences are defined primarily by their inputs and outputs, rather than by idiosyncratic, local circuits.^{40,41} In support of this hypothesis, recurring cortico-thalamic and cortico-cortical loop motifs, as well as cell-type- and layer-specific patterns of connectivity, have been found to be consistent across many cortical areas.^{42,43} However, the extent to which cortex-wide connectivity motifs might relate to the functional properties of cortical neurons is still emerging.⁴⁴

Within mouse V1, our study demonstrates that a plausible computational principle—temporal prediction—can account for

these functional connectivity patterns. Crucially, the network model was not optimized for specific response properties of visual neurons (e.g., particular receptive field characteristics). Instead, the resulting patterns of connectivity arose naturally as an emergent function of optimizing for the more general objective of predicting the neurons' future inputs. These results suggest that wiring biases found in mouse V1 are not arbitrary but rather that they underpin an important cortical function.

Comparison to other normative models

Despite the clear functional importance of recurrent connectivity in V1, there are comparatively few normative modeling studies addressing this topic. The key contribution of the present work is in uniting different aspects of both short-range and long-range V1 connectivity with neuronal feature preferences under a single unsupervised learning objective.

Sparse coding networks have been widely employed in modeling receptive fields and response properties^{16,45,46} and more recently have been applied to local connectivity in the visual cortex. Sparse coding argues that the brain is optimized to represent stimuli efficiently such that only a small number of neurons are strongly activated at a given time. When trained on static images, sparse coding models have been found to replicate the like-for-like connectivity pattern among units with similar orientation tuning.^{47,48} Where motion has been included in these models, they can capture the asymmetry in excitatory and inhibitory inputs for direction tuning.⁴⁹ However, these sparse coding models have been shown to replicate simple-cell responses only, but not complex-cell responses. Similarly, such models have not been shown to reproduce other distinct connectivity profiles across separate excitatory and inhibitory cell populations or the long-range tuning biases reported in this study.

Finally, local recurrent connectivity in V1 has also been approached from a Bayesian perspective, where the dependence of cortical connectivity on the similarity in orientation tuning is argued to represent an optimal means of integrating contextual information.⁵⁰ However, this Bayesian model depends on hard-coded basis functions derived from V1 simple cells, and unlike our approach, it can neither be said to be truly unsupervised nor learned exclusively from natural stimulus statistics.

In the context of this study, we found that only the temporal prediction model could closely reproduce the observed relationships between V1 neurons and their functional connectivity. Thus, the results cannot be accounted for by the choice of dataset or model architecture but are specific to the temporal prediction model's training objective. The temporal prediction model therefore provides a more complete explanation than the other models for the relationship between the connectivity of visual cortical neurons and the stimulus features to which they are tuned. In turn, these results suggest that the functional specificity of connections in V1 enables the brain to process dynamic stimuli by facilitating the prediction of upcoming sensory information. These predictions are critical for guiding complex actions,^{9,10} such as those required to catch a moving prey or, in the case of a tennis player, to return the ball, which depend on estimating the future state of the world.

Comparison to the biology

While the current temporal prediction model is trained using backpropagation through time, the principle of temporal

prediction itself is largely agnostic to the underlying learning mechanisms. Indeed, novel and more biologically plausible learning algorithms are being developed that could, in principle, be applied to learn temporal prediction for the current network.^{51,52} In this sense, the present work does not preclude either a hard-wired or learned origin for the connectivity patterns found in visual cortex.^{6,7}

From an evolutionary perspective, temporal prediction is likely to confer several advantages. By encoding only those features that are efficiently predictive of future sensory inputs, temporal prediction provides a principled way of extracting underlying variables and discarding non-predictive, and therefore less behaviorally relevant, information.⁹ Furthermore, given the inherent delays due to neural conduction and processing in sensory pathways and in preparing motor outputs, some form of predictive processing may be essential to accurately guide an animal's actions.¹⁰

Given that the model's structure is learned from an initial random state, such a configuration can, at least in theory, emerge from the interplay of some optimization principle and the natural statistics of visual inputs. Following the onset of vision, the connectivity of mouse V1 neurons that respond to similar visual features progressively increases.⁵³ These response-specific connectivity patterns still develop in dark-reared mice, indicating that the emergence of like-for-like wiring biases is not dependent on visual experience.⁵⁴ Nevertheless, the relationship between connection probability and the similarity of V1 responses to natural movies (but not the similarity of their orientation preferences) was found to be weaker in dark-reared mice than in animals reared with normal visual inputs. Thus, it is likely that these biases in functional connectivity result from an interplay of innate developmental programs that, at least to some extent, are later fine-tuned by sensory experience.⁵⁵ Such learning might then require some neurons or neuronal compartments to represent prediction or prediction errors congruent with the temporal prediction model output units and loss function. Notably, the hidden units in the temporal prediction model sometimes developed mixed sensitivity to both visual stimuli and prediction errors, which is consistent with cortical circuits in the brain where neurons rarely exist in a pure error-coding capacity.^{25,31}

In conclusion, we show that many aspects of functional connectivity in mouse V1 can be parsimoniously described by a single framework—temporal prediction. By optimizing a recurrent network for temporal prediction, model units naturally recapitulate both structural and functional properties of mouse visual cortex. Consequently, even seemingly disparate examples of connectivity rules may be united by a simple underlying principle of cortical organization.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for code or data should be directed to Andrew J. King (andrew.king@dpag.ox.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data, accessible at <https://portal.brain-map.org/circuits-behavior/visual-coding-neuro-pixels>.
- Original code used to generate analyses and figures presented in this manuscript has been deposited on GitHub and is publicly available as of the date of publication at https://github.com/sebbkw/temporal_prediction_connectivity.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

This work was funded by a Wellcome Principal Research Fellowship (WT108369/Z/2015/Z) to A.J.K. S.K.-W. was supported by a studentship funded by the Nuffield Department of Clinical Neurosciences at the University of Oxford. Illustrations in the graphical abstract were provided by Luigi Petrucco (<https://doi.org/10.5281/zenodo.3925903>) and Federico Claudi (<https://doi.org/10.5281/zenodo.5642927>) from SciDraw.

AUTHOR CONTRIBUTIONS

Conceptualization, S.K.-W., E.F., and N.S.H.; methodology and investigation, S.K.-W. and E.F.; writing—original draft, S.K.-W. and E.F.; writing—review and editing, S.K.-W., E.F., Y.S., M.F.I., A.J.K., and N.S.H.; funding acquisition, A.J.K.; and supervision, Y.S., M.F.I., A.J.K., and N.S.H.

DECLARATION OF INTERESTS

E.F. is currently the Chief Executive Officer of Novoic Ltd.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Dataset
 - Network model
 - Implementation
 - Comparison models
 - Neural response predictions
- QUANTIFICATION AND STATISTICAL ANALYSES
 - Receptive field mapping
 - Unit inclusion criteria
 - Unit tuning characteristics
 - Natural movie response correlations
 - Prediction errors analyses
 - Unit connectivity
 - Visual space-dependent connectivity
 - Ablation experiments
 - Control analyses
 - Mouse V1 comparisons data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2024.11.073>.

Received: May 17, 2024

Revised: October 21, 2024

Accepted: November 29, 2024

Published: January 10, 2025

REFERENCES

- Iacuruso, M.F., Gasler, I.T., and Hofer, S.B. (2017). Synaptic organization of visual space in primary visual cortex. *Nature* 547, 449–452. <https://doi.org/10.1038/nature23019>.
- Cossell, L., Iacuruso, M.F., Muir, D.R., Houlton, R., Sader, E.N., Ko, H., Hofer, S.B., and Mrsic-Flogel, T.D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* 518, 399–403. <https://doi.org/10.1038/nature14182>.
- Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature* 473, 87–91. <https://doi.org/10.1038/nature09880>.
- Ding, Z., Fahey, P.G., Papadopoulos, S., Wang, E., Celii, B., Papadopoulos, C., Kunin, A.B., Chang, A., Fu, J., Ding, Z., et al. (2024). Functional connectomics reveals general wiring rule in mouse visual cortex. Preprint at bioRxiv. <https://doi.org/10.1101/2023.03.13.531369>.
- Rossi, L.F., Harris, K.D., and Carandini, M. (2020). Spatial connectivity matches direction selectivity in visual cortex. *Nature* 588, 648–652. <https://doi.org/10.1038/s41586-020-2894-4>.
- Singer, Y., Teramoto, Y., Willmore, B.D., Schnupp, J.W., King, A.J., and Harper, N.S. (2018). Sensory cortex is optimized for prediction of future input. *eLife* 7, e31557. <https://doi.org/10.7554/eLife.31557>.
- Singer, Y., Taylor, L., Willmore, B.D.B., King, A.J., and Harper, N.S. (2023). Hierarchical temporal prediction captures motion processing along the visual pathway. *eLife* 12, e25299. <https://doi.org/10.7554/eLife.25299>.
- Lamberti, M., Tripathi, S., van Putten, M.J.A.M., Marzen, S., and le Feber, J. (2023). Prediction in cultured cortical neural networks. *PNAS Nexus* 2, pgad188. <https://doi.org/10.1093/pnasnexus/pgad188>.
- Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity, and learning. *Neural Comput.* 13, 2409–2463. <https://doi.org/10.1162/089976601753195969>.
- Johnson, P.A., Blom, T., van Gaal, S., Feuerriegel, D., Bode, S., and Hogendoorn, H. (2023). Position representations of moving objects align with real-time position in the early visual response. *eLife* 12, e82424. <https://doi.org/10.7554/eLife.82424>.
- Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M.C., DiCarlo, J.J., and Yamins, D.L.K. (2021). Unsupervised neural network models of the ventral visual stream. *Proc. Natl. Acad. Sci. USA* 118, e2014196118. <https://doi.org/10.1073/pnas.2014196118>.
- O'Reilly, R.C., Wyatte, D., and Rohrlich, J. (2014). Learning through time in the thalamocortical loops. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1407.3432>.
- Chariker, L., Shapley, R., and Young, L.-S. (2016). Orientation selectivity from very sparse LGN inputs in a comprehensive model of macaque V1 cortex. *J. Neurosci.* 36, 12368–12384. <https://doi.org/10.1523/jneurosci.2603-16.2016>.
- Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. <https://doi.org/10.1016/j.neuron.2012.10.038>.
- Muckli, L., and Petro, L.S. (2013). Network interactions: non-geniculate input to V1. *Curr. Opin. Neurobiol.* 23, 195–201. <https://doi.org/10.1016/j.conb.2013.01.020>.
- Olshausen, B.A., and Field, D.J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.* 37, 3311–3325. [https://doi.org/10.1016/S0042-6989\(97\)00169-7](https://doi.org/10.1016/S0042-6989(97)00169-7).
- Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* 592, 86–92. <https://doi.org/10.1038/s41586-020-03171-x>.
- Durand, S., Iyer, R., Mizuseki, K., de Vries, S., Mihalas, S., and Reid, R.C. (2016). A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice. *J. Neurosci.* 36, 12144–12156. <https://doi.org/10.1523/JNEUROSCI.1741-16.2016>.
- Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, 284–299. <https://doi.org/10.1364/josaa.2.000284>.
- Rust, N.C., Schwartz, O., Movshon, J.A., and Simoncelli, E.P. (2005). Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46, 945–956. <https://doi.org/10.1016/j.neuron.2005.05.021>.
- Niell, C.M., and Stryker, M.P. (2008). Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* 28, 7520–7536. <https://doi.org/10.1523/jneurosci.0623-08.2008>.
- Gavornik, J.P., and Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* 17, 732–737. <https://doi.org/10.1038/nn.3683>.
- Fiser, A., Mhringer, D., Oyibo, H.K., Petersen, A.V., Leinweber, M., and Keller, G.B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci.* 19, 1658–1664. <https://doi.org/10.1038/nn.4385>.
- Westerberg, J.A., Xiong, Y.S., Nejat, H., Sennesh, E., Durand, S., Cabasco, H., Belski, H., Gillis, R., Loeffler, H., Bawany, A., et al. (2024). Stimulus history, not expectation, drives sensory prediction errors in mammalian cortex. Preprint at bioRxiv. <https://doi.org/10.1101/2024.10.02.616378>.
- Price, B.H., Jensen, C.M., Khoudary, A.A., and Gavornik, J.P. (2023). Expectation violations produce error signals in mouse V1. *Cereb. Cortex* 33, 8803–8820. <https://doi.org/10.1093/cercor/bhad163>.
- Jamali, S., Bagur, S., Brémont, E., Van Kerkoerle, T., Dehaene, S., and Bathellier, B. (2024). Parallel mechanisms signal a hierarchy of sequence structure violations in the auditory cortex. *eLife* 13, RP102702. <https://doi.org/10.7554/elife.102702.1>.
- Li, J., Liao, X., Zhang, J., Wang, M., Yang, N., Zhang, J., Lv, G., Li, H., Lu, J., Ding, R., et al. (2017). Primary auditory cortex is required for anticipatory motor response. *Cereb. Cortex* 27, 3254–3271. <https://doi.org/10.1093/cercor/bhx079>.
- Wang, M., Li, R., Li, J., Zhang, J., Chen, X., Zeng, S., and Liao, X. (2018). Frequency selectivity of echo responses in the mouse primary auditory cortex. *Sci. Rep.* 8, 49. <https://doi.org/10.1038/s41598-017-18465-w>.
- Musall, S., Haiss, F., Weber, B., and von der Behrens, W. (2017). Deviant processing in the primary somatosensory cortex. *Cereb. Cortex* 27, 863–876. <https://doi.org/10.1093/cercor/bhv283>.
- Auksztulewicz, R., Rajendran, V.G., Peng, F., Schnupp, J.W.H., and Harper, N.S. (2023). Omission responses in local field potentials in rat auditory cortex. *BMC Biol.* 21, 130. <https://doi.org/10.1186/s12915-023-01592-4>.
- Tang, M.F., Kheradpezhoh, E., Lee, C.C.Y., Dickinson, J.E., Mattingley, J.B., and Arabzadeh, E. (2023). Expectation violations enhance neuronal encoding of sensory information in mouse primary visual cortex. *Nat. Commun.* 14, 1196. <https://doi.org/10.1038/s41467-023-36608-8>.
- Scholl, B., Wilson, D.E., Jaepel, J., and Fitzpatrick, D. (2019). Functional logic of layer 2/3 inhibitory connectivity in the ferret visual cortex. *Neuron* 104, 451–457.e3. <https://doi.org/10.1016/j.neuron.2019.08.004>.
- Harris, K.D., and Mrsic-Flogel, T.D. (2013). Cortical connectivity and sensory coding. *Nature* 503, 51–58. <https://doi.org/10.1038/nature12654>.
- Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1409.1556>.
- Lotter, W., Kreiman, G., and Cox, D. (2017). Deep predictive coding networks for video prediction and unsupervised learning. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1605.08104>.
- Rao, R.P.N., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. <https://doi.org/10.1038/4580>.

37. Taylor, L., Zenke, F., King, A.J., and Harper, N.S. (2024). Temporal prediction captures retinal spiking responses across animal species. Preprint at bioRxiv. <https://doi.org/10.1101/2024.03.26.586771>.
38. Cohen, U., Chung, S., Lee, D.D., and Sompolinsky, H. (2020). Separability and geometry of object manifolds in deep neural networks. *Nat. Commun.* *11*, 746. <https://doi.org/10.1038/s41467-020-14578-5>.
39. Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* *20*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
40. Di Lazzaro, V., Capone, F., Paolucci, M., Assenza, F., Brunelli, N., Ricci, L., and Florio, L. (2016). Canonical cortical circuits: current evidence and theoretical implications. *Neurosci. Neuroecon.* *5*, 1–8. <https://doi.org/10.2147/NAN.S70816>.
41. Nelson, S.B. (2002). Cortical microcircuits: diverse or canonical? *Neuron* *36*, 19–27. [https://doi.org/10.1016/s0896-6273\(02\)00944-3](https://doi.org/10.1016/s0896-6273(02)00944-3).
42. Shepherd, G.M.G., and Yamawaki, N. (2021). Untangling the cortico-thalamo-cortical loop: cellular pieces of a knotty circuit puzzle. *Nat. Rev. Neurosci.* *22*, 389–406. <https://doi.org/10.1038/s41583-021-00459-3>.
43. Brown, S.P., and Hestrin, S. (2009). Cell-type identity: a key to unlocking the function of neocortical circuits. *Curr. Opin. Neurobiol.* *19*, 415–421. <https://doi.org/10.1016/j.conb.2009.07.011>.
44. Reid, R.C. (2012). From functional architecture to functional connectomics. *Neuron* *75*, 209–217. <https://doi.org/10.1016/j.neuron.2012.06.031>.
45. Chalk, M., Marre, O., and Tkačik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. USA* *115*, 186–191. <https://doi.org/10.1073/pnas.1711114115>.
46. Ecke, G.A., Bruijns, S.A., Hoelscher, J., Mikulasch, F.A., Witschel, T., Arrenberg, A.B., and Mallot, H.A. (2020). Sparse coding predicts optic flow specificities of zebrafish pretectal neurons. *Neural Comput. Appl.* *32*, 6745–6754. <https://doi.org/10.1007/s00521-019-04500-6>.
47. Capparelli, F., Pawelzik, K., and Ernst, U. (2019). Constrained inference in sparse coding reproduces contextual effects and predicts laminar neural dynamics. *PLoS Comput. Biol.* *15*, e1007370. <https://doi.org/10.1371/journal.pcbi.1007370>.
48. Garrigues, P., and Olshausen, B. (2007). Learning horizontal connections in a sparse coding model of natural images. *Advances in Neural Information Processing Systems* *20*. <https://doi.org/10.1167/2.7.130>.
49. Pachitariu, M., and Sahani, M. (2017). Visual motion computation in recurrent neural networks. Preprint at bioRxiv. <https://doi.org/10.1101/099101>.
50. Iyer, R., and Mihalas, S. (2017). Cortical circuits implement optimal context integration. Preprint at bioRxiv. <https://doi.org/10.1101/158360>.
51. Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T., and Lin, Z. (2016). Towards biologically plausible deep learning. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1502.04156>.
52. Millidge, B., Tang, M., Osanlouy, M., Harper, N.S., and Bogacz, R. (2024). Predictive coding networks for temporal prediction. *PLoS Comput. Biol.* *20*, e1011183. <https://doi.org/10.1371/journal.pcbi.1011183>.
53. Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S.B., and Mrsic-Flogel, T.D. (2013). The emergence of functional microcircuits in visual cortex. *Nature* *496*, 96–100. <https://doi.org/10.1038/nature12015>.
54. Ko, H., Mrsic-Flogel, T.D., and Hofer, S.B. (2014). Emergence of feature-specific connectivity in cortical microcircuits in the absence of visual experience. *J. Neurosci.* *34*, 9812–9816. <https://doi.org/10.1523/JNEUROSCI.0875-14.2014>.
55. Kiorpes, L. (2015). Visual development in primates: neural mechanisms and critical periods. *Dev. Neurobiol.* *75*, 1080–1090. <https://doi.org/10.1002/dneu.22276>.
56. Meyer, H.S., Schwarz, D., Wimmer, V.C., Schmitt, A.C., Kerr, J.N.D., Sakmann, B., and Helmstaedter, M. (2011). Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proc. Natl. Acad. Sci. USA* *108*, 16807–16812. <https://doi.org/10.1073/pnas.1113648108>.
57. van Versendaal, D., and Levitt, C.N. (2016). Inhibitory interneurons in visual cortical plasticity. *Cell. Mol. Life Sci.* *73*, 3677–3691. <https://doi.org/10.1007/s00018-016-2264-4>.
58. Sahani, M., and Linden, J. (2002). How linear are auditory cortical responses? *Advances in Neural Information Processing System* *15*. <https://doi.org/10.1523/jneurosci.3377-07.2008>.
59. Rahman, M., Willmore, B.D.B., King, A.J., and Harper, N.S. (2020). Simple transformations capture auditory input to cortex. *Proc. Natl. Acad. Sci. USA* *117*, 28442–28451. <https://doi.org/10.1073/pnas.1922033117>.
60. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* *17*, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
61. Schoppe, O., Harper, N.S., Willmore, B.D.B., King, A.J., and Schnupp, J.W.H. (2016). Measuring the performance of neural models. *Front. Comput. Neurosci.* *10*, 10. <https://doi.org/10.3389/fncom.2016.00010>.
62. Hsu, A., Borst, A., and Theunissen, F.E. (2004). Quantifying variability in neural responses and its application for the validation of model predictions. *Network* *15*, 91–109. <https://doi.org/10.1088/0954-898x/15/2/002>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Allen Institute Visual Coding database	The Allen Institute	https://portal.brain-map.org/circuits-behavior/visual-coding-neuropixels
Software and algorithms		
Analysis scripts	This paper	https://github.com/sebbkw/temporal_prediction_connectivity
Python	https://www.python.org/	N/A
PyTorch	https://pytorch.org/	N/A

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental data used for neural response prediction was taken from the Allen Institute for Brain Science Neuropixels Visual Coding dataset, whose detailed experimental procedures have been published elsewhere.¹⁷ The dataset used for neural fitting was comprised of recordings from 15 male wildtype C57BL6/J mice ($n = 739$ neurons).

METHOD DETAILS

Dataset

Training data consisted of natural wildlife videos using the same dataset as described previously.⁶ Videos were taken from the repository <http://www.arkive.org/species> and contributed by: BBC Natural History Unit, <http://www.gettyimages.co.uk/footage/bbcmotiongallery>; BBC Natural History Unit & Discovery Communications Inc, <http://www.bbcmotiongallery.com>; Granada Wild, <http://www.itnsource.com>; Mark Deeble & Victoria Stone Flat Dog Productions Ltd., <http://www.deeblestone.com>; Getty Images, <http://www.gettyimages.com>; National Geographic Digital Motion, <http://www.ngdigitalmotion.com>. In brief, videos were converted to grayscale, bandpass filtered then downsampled to 180x180 pixels. Finally, each video was cropped into non-overlapping 36x36 pixel patches of 50 frames each, leading to a total of 40,000 clips for the training dataset and 4,000 clips for the validation dataset used for hyperparameter selection. In addition, to mimic the effects of noise present in the nervous system, Gaussian noise was added to each video clip during training with a signal-to-noise ratio of 6 dB.⁶

For the models in which the temporal offset was varied, training data consisted of a set of 120 Hz naturalistic videos³⁷ (available at https://figshare.com/articles/dataset/Natural_movies/24265498). For each temporal offset, for the target to be predicted, but not for the input, we shifted the 120 Hz videos 0-10 frames (0-83 ms) into the future. Both the input and target datasets were then downsampled to 24 Hz to approximately match the 25 Hz frame rate of the original temporal prediction model, and then cut into 50-frame long clips. This setup ensured that the input dataset was the same across temporal offsets, while only the frame targets varied as a function of temporal offset.

Network model

The model was implemented as a single-layer recurrent network with a linear readout layer to project the hidden activity to the network's output predictions. The network's input consisted of a 50-frame video clip, where the model was trained to predict each subsequent frame given the preceding video frames in the clip. More formally, the network receives a 1,296 length vector $\mathbf{u}[t]$ at each time step t , consisting of the flattened 36x36 pixel video frame. The 2,592 length hidden state vector $\mathbf{s}[t]$ at each time step t is then given by:

$$\mathbf{s}[t] = \mathbf{f}(\mathbf{W}_{\text{in}}\mathbf{u}[t] + \mathbf{W}_{\text{rec}}\mathbf{s}[t - 1] + \mathbf{b}_{\text{rec}})$$

where \mathbf{f} is the ReLU function, \mathbf{W}_{in} is the weight matrix which describes the input weights to the network, \mathbf{W}_{rec} is the weight matrix which describes the hidden, recurrent weights mapping the previous state $\mathbf{s}[t - 1]$ to the new hidden state $\mathbf{s}[t]$, and \mathbf{b}_{rec} is the bias term.

The hidden activity vector $\mathbf{s}[t]$ at each time step t is then mapped to the output predictions $\hat{\mathbf{v}}[t]$ by:

$$\hat{\mathbf{v}}[t] = \mathbf{W}_{\text{out}}\mathbf{s}[t] + \mathbf{b}_{\text{out}}$$

where \mathbf{W}_{out} is the weight matrix describing the linear mapping from the hidden state to the output prediction and \mathbf{b}_{out} is the bias term for the output weights.

In addition, to enforce Dale's Law whereby hidden units make exclusively excitatory or inhibitory connections, each recurrent weight w was constrained during the forward pass as:

$$w \leftarrow \begin{cases} -|w| & \text{if inhibitory} \\ +|w| & \text{if excitatory} \end{cases}$$

with a total of 2,332 (90%) units set as excitatory and the remaining 260 (10%) as inhibitory units.^{56,57}

The network was then optimized using backpropagation to minimize the loss function E :

$$E = \sum_{n=1}^N \sum_{t=1}^T \|\hat{\mathbf{v}}_n[t] - \mathbf{v}_n[t+1]\|_2^2 + \lambda(\|\mathbf{W}_{in}\|_1 + \|\mathbf{W}_{rec}\|_1 + \|\mathbf{W}_{out}\|_1)$$

where n is the clip number, N is the total number of clips in a minibatch, T is the total number of time steps, and $\hat{\mathbf{v}}_n[t] - \mathbf{v}_n[t+1]$ is the difference between the predicted pixel values $\hat{\mathbf{v}}_n[t]$ and true future pixel values $\mathbf{v}_n[t+1]$. Finally, L1 regularization is included as the sum of absolute values of all weights in the network, weighted by the λ hyperparameter.

Implementation

The temporal prediction model was implemented in PyTorch, with gradient descent performed using the ADAM optimizer set at a learning rate of 10^{-4} . Unless otherwise noted, the regularization strength hyperparameter λ was set at 10^{-6} after a hyperparameter search across lambda values (λ range = $10^{-5.5} - 10^{-7}$) to minimize the mean squared error on the held-out validation set.

Comparison models

The inpainting, denoising and sparse autoencoder networks consisted of the same network architecture as the recurrent temporal prediction model but with modified datasets and training objectives. For the inpainting network, the input dataset was masked with 8 randomly placed 8x8 pixel patches on each frame. For the denoising network, the input was combined with Gaussian noise with a signal-to-noise ratio of 3 dB. Finally, the sparse autoencoder was trained on the same dataset as the temporal prediction model but was trained to recover the current frame under a sparsity constraint. In all these networks, the models were optimized to produce the unmodified current frame (rather than subsequent frame, as for the temporal prediction model) by minimizing the mean squared error between the predicted and actual current frame:

$$E = \sum_{n=1}^N \sum_{t=1}^T \|\hat{\mathbf{v}}_n[t] - \mathbf{v}_n[t]\|_2^2 + \lambda(\|\mathbf{W}_{in}\|_1 + \|\mathbf{W}_{rec}\|_1 + \|\mathbf{W}_{out}\|_1)$$

For the sparse autoencoder, an additional regularization term λ_{act} was included as the absolute sum of activity across all units, to encourage sparsity in the network's representations:

$$E = \sum_{n=1}^N \sum_{t=1}^T \|\hat{\mathbf{v}}_n[t] - \mathbf{v}_n[t]\|_2^2 + \lambda(\|\mathbf{W}_{in}\|_1 + \|\mathbf{W}_{rec}\|_1 + \|\mathbf{W}_{out}\|_1) + \lambda_{act} \sum_{n=1}^N \sum_{t=1}^T \|\mathbf{s}_n[t]\|_1$$

For the inpainting and denoising networks, the L1 weight regularization hyperparameter was chosen as for the temporal prediction network to minimize the mean squared error on the validation set across a range of values. For the sparse autoencoder, where no such comparable selection criterion exists, the hyperparameter set was qualitatively optimized to produce the most biologically realistic receptive fields.⁷

For VGG-19, we used a publicly available model from the PyTorch 'torchvision.models' package, pre-trained on the ImageNet dataset for object recognition.³⁴ For each layer of the network, we took the hidden activity as the concatenated, flattened feature maps of each filter in the layer.

For PredNet, we reimplemented the model described by Lotter et al.,³⁵ trained for next-frame prediction on the same dataset as the main temporal prediction model. However, because the resolution halves at each layer, we used a slightly larger input size of 40x40 pixels. Again, as for VGG-19, we took the hidden activity across layers as the concatenated, flattened feature maps of each filter in that layer.

The LN model used the same basic fitting procedure as the other models. However, whereas in the other models used to predict neural responses we regressed the neural responses on each pre-trained network's hidden activity, the output of the LN model was directly fitted to the neural responses.

Neural response predictions

Neural data were taken from the Allen Institute's Neuropixels Visual Coding dataset.¹⁷ For each model, a linear-nonlinear mapping was fitted to predict the response of V1 units to natural movie stimuli ("Natural Movie One" and "Natural Movies Three", 150 seconds total) from the pre-trained model's hidden unit activity. We divided the dataset into 15x10 second clips, taking 3 representative clips (30s total) as the final held-out test set, with the remaining 120 seconds taken as the training and validation sets with hyperparameter selection by k-fold cross-validation. We included all recorded V1 units from wildtype mice whose noise-to-signal power ratio⁵⁸ in response to the natural movies was below 60.

For each model, the neural fitting process consisted of first fitting a linear mapping using Lasso regression before fitting a rectified sigmoidal non-linearity.⁵⁹ Prior to fitting, the dimensionality of the model unit activity that was regressed on was reduced to the first 200 components of a principal component analysis (PCA) fitted on the training-set model responses. PCA was used to equate the number of parameters across models and increase the efficiency of model fitting.⁷ The non-linearity $\hat{r}(x)$ was defined as:

$$\hat{r}(x) = \text{ReLU} \left(\frac{a}{1 + e^{\frac{c-x}{b}}} + d \right)$$

where the parameters a , b , c and d were optimized to minimize the mean squared error between the true and predicted neural firing rate using the SciPy “curve_fit” function.⁶⁰ The L1-regularization strength (α) of the Lasso was chosen via cross-validation from 40 values log-spaced between 10^1 and 10^{-5} to maximize the average normalized correlation coefficient (CC_{norm})^{61,62} across each fold’s validation set for the combined linear-nonlinear mapping. The reported values are finally taken as the performance on the held-out test set.

QUANTIFICATION AND STATISTICAL ANALYSES

Receptive field mapping

Model unit receptive fields were estimated using their response-weighted average. In brief, the responses of model units to 25,000 frames of random Gaussian noise ($\mu=0$, $\sigma=1$) were produced. Each noise frame was then weighted by the unit’s response to give the receptive field estimate. Model receptive fields were subsequently parameterized by a Gabor function to extract the receptive field centers and 2D extent.

We calculated the unit’s spatiotemporal response-weighted average similarly to the standard response-weighted average, but included the past 7 frames. The temporal power was then taken as the mean squared value over space for each time step, normalized by the total power. Similarly, to calculate the projective center of mass, we took the mean squared weight for each unit for each predicted future-frame time step. For each of these, we then determined the center of mass as the average of the time values weighted by that time step’s corresponding normalized power.

Unit inclusion criteria

To maintain consistency across analyses, only those units whose receptive fields that were spatially well defined and which could be well modeled as Gabors were included for analysis. To that end, units whose receptive fields were less than 0.5 pixels in size and therefore had little spatial extent (19% of total units) or which were poorly fitted by the Gabor function were excluded ($r < 0.7$, 12% of total units; 30% including both criteria). Short-range connections were defined as those less than 15° (2.5 pixels) and long-range as greater than 30° (5 pixels).¹ Connections greater than 9.17 pixels were excluded because of the experimental constraints imposed by screen size. In the case of Ko et al.,³ connections were not explicitly defined according to the distance between receptive fields, but we use the same short-range convention as in Iacuruso et al.¹ that, under the assumption of retinotopy, physically short-range connections ($< 50 \mu\text{m}$) are likely to be close in visual space.

Unit tuning characteristics

To measure the model units’ tuning properties, each unit’s response to sinusoidal gratings was recorded. Gratings varied in temporal frequency (0.02–0.25 cycles/frame), spatial frequency (0.03–0.5 cycles/pixel), and orientation (0–360 degrees) with an amplitude of ± 1 . Each unit’s preferred temporal frequency, spatial frequency and orientation were taken as the parameter or parameter combination that maximized the unit’s mean response across 50 frames. For those analyses dependent on the unit’s spatial location, only units within the central 16x16 pixel bounds of the visual fields were included to avoid edge effects.

Orientation and direction selectivity were quantified as OSI and DSI, respectively:

$$\text{OSI} = \frac{R_{\text{pref}}^{\text{Or}} - R_{\text{orth}}^{\text{Or}}}{R_{\text{pref}}^{\text{Or}} + R_{\text{orth}}^{\text{Or}}}$$

$$\text{DSI} = \frac{R_{\text{pref}}^{\text{Dir}} - R_{\text{opp}}^{\text{Dir}}}{R_{\text{pref}}^{\text{Dir}} + R_{\text{opp}}^{\text{Dir}}}$$

where $R_{\text{pref}}^{\text{Or}}$ and $R_{\text{orth}}^{\text{Or}}$ are the unit responses at the preferred and orthogonal orientations, and $R_{\text{pref}}^{\text{Dir}}$ and $R_{\text{opp}}^{\text{Dir}}$ are the unit responses at the preferred and opposite ($+180$ degrees) directions. For Figure 2, we take the same thresholds as Ko et al.,³ where direction selective units were defined as those with OSI values exceeding 0.4 and DSI values exceeding 0.3. For Figure 3, where no threshold is given in Rossi et al.,⁵ we take the more stringent threshold of 0.8 for direction-selective units.

Model units were classified as simple- or complex-like based on their phase-responsiveness to drifting grating stimuli. Quantitatively, units with a modulation ratio $F > 1$ were classed as simple-cell-like or as complex-cell-like for a modulation ratio < 1 . The

modulation ratio was defined as $F = \frac{F_1}{F_0}$ where F_0 is the mean response of the neuron to its preferred stimulus and F_1 is the amplitude of the fitted sinusoid to the neuron's response to its preferred stimulus. Where the correlation between the fitted sinusoid and the true response was < 0.9 , we did not calculate the modulation ratio for that unit.

Natural movie response correlations

One hundred 50-frame clips were randomly selected from the validation set and the response recorded for each model unit. The response correlation was then taken as the average correlation across the set of clips for each pair of units in the network.

Prediction errors analyses

We used two paradigms to assess the presence of sensitivity to prediction error in the network hidden units: oddball stimuli and omission stimuli. In the oddball paradigm, a deviant stimulus interrupts the pattern generated by a preceding set of standard stimuli. In the omission paradigm, the violating stimulus consists of an omission – i.e. the absence of an expected stimulus presentation.

In the odd-ball paradigm, we presented the model with stimuli consisting of two full-field gratings of orientations A and B that each spanned 0° , 45° , 90° or 135° (Figure S2A). The omission paradigm was similar but consisted of a single full-field grating A and a blank stimulus B. A and B stimuli alternated for 25 frames, while the omission or deviant position was varied to occur after 5–25 standard frames. We constructed two sets of control stimuli to ensure that prediction-error responses could not be explained by differences in stimulus tuning or unrelated network dynamics. First, we compared the response to the violating stimulus with that generated by the same set of stimuli using the standard stimulus (i.e. ABABB vs. ABABA). Second, we constructed a 'shifted' set of stimuli, where the deviant position was matched but without violating the pattern (i.e. ABABB vs. BABAB).

We chose a relatively conservative set of criteria to avoid miscategorizing non-prediction-error responses. Specifically, a unit was defined as prediction-error-like only if for a particular orientation or orientation pair it had no response in either control condition and responded to the violating stimulus across at least 5 different deviant positions. We also adopted slightly looser criteria (Figure S2C), whereby prediction-error-like responses were defined when responses to deviant or omission stimuli exceeded three times the response to control stimuli.

Unit connectivity

Units were defined as connected if their connection strength exceeded the 95th percentile of connection weights (\mathbf{W}_{rec}) across all pairs of units. Due to the sparse nature of the recurrent weight connectivity matrix, this threshold equated to rejecting the very low or zero weight connections, while retaining the smaller subset of highly connected units. Thus, varying this threshold across a range of values (92.5–99 percentile) did not qualitatively change the results.

Visual space-dependent connectivity

For each presynaptic ensemble, visual space was normalized according to the receptive field center and preferred orientation of the postsynaptic unit. Receptive field centers were first translated such that the postsynaptic unit receptive field was centered at the origin:

$$\begin{bmatrix} x'_{\text{pre}} \\ y'_{\text{pre}} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -x_{\text{post}} \\ 0 & 1 & -y_{\text{post}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{pre}} \\ y_{\text{pre}} \\ 1 \end{bmatrix}$$

Next, receptive field centers were rotated according to the postsynaptic unit's preferred orientation θ :

$$\begin{bmatrix} x'_{\text{pre}} \\ y'_{\text{pre}} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\left(\theta - \frac{\pi}{2}\right) & -\sin\left(\theta - \frac{\pi}{2}\right) & 0 \\ \sin\left(\theta - \frac{\pi}{2}\right) & \cos\left(\theta - \frac{\pi}{2}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{pre}} \\ y_{\text{pre}} \\ 1 \end{bmatrix}$$

For the comparison with Iacaruso et al.,¹ presynaptic units were binned into co-orthogonal and co-axial receptive field centers according to whether they fell in one of the four quadrants orthogonal to or parallel with the postsynaptic unit's preferred orientation (i.e., defined by $y = -x$ and $y = x$). For the comparison with Rossi et al.,⁵ presynaptic units were binned according to whether they fell opposite to or ahead of the postsynaptic unit's preferred direction of motion (i.e., defined by $x = 0$). Presynaptic densities were computed by binning the presynaptic receptive field centers (taken as the center of the Gabor fit to the unit's receptive field, as described above) across visual space.

Ablation experiments

For the ablation experiments, we randomly selected n connections belonging to the relevant class of units. For each of these units, the connection was 'ablated' by setting its value to the median connection weight for recurrent weights between the excitatory subpopulation of the network. This was repeated 1,000 times and the average taken to obtain the MSE values in Figure 6.

For the silhouette score analyses, we presented three classes of grating stimuli whose orientations were offset by 0° , 11.25° and 22.5° , repeating the analysis across the complete span of orientations (0 – 360°). Over these classes of stimuli, we then calculated the silhouette score to measure the level of distinctness of the three classes in the hidden representations. The silhouette score is bounded between -1 and 1 , with a high score indicating that the clusters are well separated according to the stimulus (orientation) class.³⁹

Control analyses

The distribution of preferred orientations and directions among model units was not uniform (Figure S1). To control for the possibility that the observed model connectivity distributions resulted from this overrepresentation of particular orientation and direction tuning preferences, we compared the true model results to those after randomly shuffling across model weights. Specifically, we took the total set of model units fulfilling the relevant criteria for each analysis (e.g. orientation selectivity, receptive field distance, etc.) and randomly shuffled the recurrent weights connecting these units. It is important to note that we used these shuffled weights only for the connectivity analyses, not for any other part of any analysis, such as getting unit responses for response weighted averages. We repeated this process 1,000 times, taking the mean value of the resulting distribution to compare to the true unshuffled model results.

Mouse V1 comparisons data

Mouse data for comparison were either extracted from published figures using WebPlotDigitizer (Figures 1F and 2J), taken as the exact statistics from the published paper (Figures 2B, 2C, 2F, 2G, and 2I) or computed directly from the Neuropixels Visual Coding dataset from the Allen Institute¹⁷ (Figures 1C, 1E, 4C, and S1). For the V1 receptive fields (Figure 1C), these were estimated by fitting a linear filter to predict the responses of V1 single units to natural movies as described above.

To compare how well each model captured the connectivity patterns described for mouse V1, we calculated a model connectivity score as the average of the Pearson correlation coefficients between the described neural connectivity profiles (data in Figures 2B, 2C, 2F, 2G, 2I, and 2J) and the corresponding connectivity profiles of the given model.