Nonlinear processing of shape information in rat lateral extrastriate cortex

Giulio Matteucci[1], Rosilari Bellacosa Marotti[1], Federica B. Rosselli[1,2,*] and Davide Zoccolan[1,*]

1. *Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Trieste, 34136, Italy*

2. *Department of Behavior and Brain Organization, Center of Advanced European Studies and Research (Caesar), an Institute of the Max Planck Society, Bonn, 53175, Germany*

*Co-corresponding authors

Corresponding authors

Davide Zoccolan
International School for Advanced Studies (SISSA)
Via Bonomea, 265 34136 Trieste (TS) ITALY
Phone: +39 040 3787 701 E-mail: zoccolan@sissa.it

Federica B. Rosselli
Dept. of Behavior and Brain Organization
Center of Advanced European Studies and Research (Caesar)
An Institute of the Max-Planck Society
Ludwig-Erhard-Allee 2 53175, Bonn (Germany)
Phone: +49 (0)228 9656 317 E-mail: federica.rosselli@caesar.de

Abstract

We previously reported that the progression of rat extrastriate areas located laterally to primary visual cortex (V1) is specialized for supporting visual object discrimination despite view changes – the hallmark ability of a ventral-like pathway (Tafazoli et al., 2017). Here, we sought to clarify what computations are at the base of such specialization, by comparing neurons in V1 and laterolateral area LL in terms of the nonlinearity of the stimulus-response relationship and the tendency to be tuned for multiple stimulus orientations. Critically, the trends of variation observed for these properties were consistent with the nonlinear summation of visual inputs that is expected to take place along the ventral stream, according to the predictions of a hierarchical model of ventral computations and a meta-analysis of the monkey literature. This suggests an intriguing homology between the mechanisms responsible for building up shape selectivity and transformation tolerance in rodent and primate visual cortex.

Introduction

In rodents, primary visual cortex (V1) is bordered by several extrastriate visual areas, whose involvement in dorsal and ventral stream processing [i.e., extraction of, respectively, motion and shape information (Maunsell and Newsome, 1987; Orban, 2008)] has become the object of intense investigation (Gao et al., 2010; Van den Bergh et al., 2010; Andermann et al., 2011; Marshel et al., 2011; Roth et al., 2012; Matsui and Ohki, 2013; Glickfeld et al., 2013; Tohmi et al., 2014; Vermaercke et al., 2014; Ji et al., 2015; Vinken et al., 2016, 2017; Tafazoli et al., 2017). Inspired by previous anatomical (Wang and Burkhalter, 2007; Wang et al., 2011a, 2012a) and lesion studies (Gallardo et al., 1979; McDaniel et al., 1982; Wörtwein et al., 1993; Aggleton et al., 1997; Sánchez et al., 1997; Tees, 1999), we recently looked for functional evidence of ventral processing along the progression of extrastriate areas that, in the rat brain, run laterally to V1 (Tafazoli et al., 2017). Our experiments showed that the neuronal populations along this progression become increasingly capable of supporting discrimination of visual objects despite transformations in their appearance (e.g., due to translation and scaling), with the largest transformation tolerance achieved in the most ventral region: the laterolateral area (LL). This trend closely matches the one found along the primate ventral pathway (Li et al., 2009; Rust and DiCarlo, 2010; DiCarlo et al., 2012; Pagan et al., 2013; Yamins et al., 2014; Hong et al., 2016), thus supporting the central role of LL in

processing object information. At the same time, these findings raise the question of what tuning mechanisms are responsible for building up such a specialization for ventral processing in LL.

Here, we address this question by comparing V1 and LL in terms of two key properties that are highly discriminant of higher-order vs. lower-order shape processing: the nonlinearity of the stimulus-response relationship and the tendency of a neuron to be selective for multiple oriented features. The first property was assessed by measuring the degree to which neuronal responses were entrained to the phase of drifting gratings (Movshon et al., 1978; Skottun et al., 1991; Wypych et al., 2012), thus distinguishing linear, phase-locked neurons with a *simple-like* tuning from nonlinear, phase-invariant units with a *complex-like* behavior – where *simple* and *complex* refer here to the classes of cells originally discovered by Hubel and Wiesel in cat primary visual cortex (Hubel and Wiesel, 1968). In addition, we directly measured the extent to which the spatiotemporal structure of the receptive fields (RFs) in V1 and LL could be inferred using a linear reverse correlation approach, such as the *spike-triggered average* (STA) method (Schwartz et al., 2006; Sharpee, 2013). The second property was assessed by detecting the presence of multiple peaks in the orientation tuning curves (David et al., 2006; Cadieu et al., 2007) and by measuring their sharpness.

We found that, compared to V1 neurons, LL units displayed stronger phase-invariance of the response, while the linear structure of their RFs was much less defined. In addition, LL neurons showed weaker orientation tuning and a stronger tendency to respond to multiple orientations of the gratings. Critically, these findings are consistent with the processing that is expected to take place along the ventral stream, where nonlinear pooling of presynaptic inputs with either similar or different feature selectivity (e.g., orientation preference) is thought to mediate the emergence of, respectively, transformation tolerance and shape selectivity (Riesenhuber and Poggio, 1999; DiCarlo et al., 2012). These conclusions were confirmed by the agreement of our findings with the predictions derived from a hierarchical model of ventral computations and a meta-analysis of the monkey literature. As such, our results suggest a strong homology between the mechanisms that are responsible for supporting higher-order processing of shape information in rodents, primates and brain-inspired machine vision systems.

Results

We used 32-channel silicon probes to perform extracellular recordings from areas V1 and LL of anaesthetized rats that were presented with drifting gratings of various orientations, spatial (SF) and temporal (TF) frequencies, as well as with 80 30s-long movies of spatially and temporally correlated noise (Materials and Methods). A total of 105 and 104 neurons were recorded, respectively, from V1 and LL. With "neurons" we refer here to well-isolated single units that were effectively and reproducibly driven by one or more of the grating stimuli across repeated presentations (Materials and Methods). The assignment of the recorded units to a specific area was carried out by tracking the retinotopy of the RFs, combined with the histological reconstruction of the probe insertion track, as described previously (Tafazoli et al., 2017).

Examples of tuning curves and spatiotemporal receptive fields in V1 and LL

In both V1 and LL, neuronal responses to drifting gratings displayed a variety of tuning profiles that matched well the RF structures inferred through STA. This is illustrated in Figure 1, which reports, for each of two representative V1 (A-B) and LL (C-D) neurons: i) the tuning curve as a function of the direction of the grating, when presented at the most effective SF and TF; ii) the modulation of the firing rate elicited by the grating at the preferred direction; and iii) the sequence of STA images showing the spatiotemporal evolution of the RF structure. Each panel also reports the values of the metrics used to quantify the tuning properties of the neurons (Materials and Methods). Orientation and direction tuning were measured using the *orientation* and *direction selectivity indexes* (OSI and DSI), whose values range from 1 (maximal selectivity) to 0 (no selectivity), depending on the difference between the responses to the preferred orientation/direction and the orthogonal/opposite one. We also computed a *bimodal selectivity index* (BSI) to assess whether the tuning curve for orientation had multiple peaks. This index ranges from 0 (perfectly unimodal curve) to 1 (equal responsiveness to two non-adjacent orientations), and can be taken as a measure of complexity of tuning, with large BSI indicating tuning for multiple oriented features (David et al., 2006; Cadieu et al., 2007). Finally, we computed a *modulation index* (MI) to quantify the extent to which the firing rate of a neuron was entrained to the phase of the gratings (Movshon et al., 1978; Skottun et al., 1991; Wypych et al., 2012), with MI > 3 indicating responses that were strongly phase-locked to the stimulus temporal frequency. As for the spatiotemporal filter estimated via STA, we used it as the

linear stage of a Linear-Nonlinear (LN) model of stimulus-response mapping (Schwartz et al., 2006; Sharpee, 2013) to predict the tuning curve of the neuron over the direction axis and the time course of its response to the preferred grating (in both cases, prediction accuracy was quantified as the fraction of response variance that was explained by the model).

Figure 1A shows an example V1 neuron with sharp orientation tuning, good direction selectivity, and a response that was strongly modulated at the temporal frequency (4 Hz) of the preferred grating (green curves/dots). These properties match those of a *simple cell* (Hubel and Wiesel, 1968), detecting the presence of an oriented edge at a specific visual field location – hence, the modulation of the firing rate, produced by the phasic alternation of light and dark oriented stripes, while the preferred grating swept across the neuron's RF. Such position-dependent neuronal firing suggests that the response of the neuron to a visual pattern can be approximated as the sum of its responses to the individual constituent elements (the "pixels") of the stimulus. Accordingly, the STA-reconstructed RF (Figure 1A.iii) had a Gabor-like, double-lobed appearance, with an orientation matching the one corresponding to the peak of the tuning curve (Figure 1A.i). When used as a filter in a LN model, the STA-based RF accurately predicted the tuning of the neuron over the direction and time axes (Figure 1A.i and ii, respectively; black curves), thus confirming the linearity of the stimulus-response relationship. By contrast, the example unit shown in Figure 1B, while still being selective for a single orientation, displayed a highly non-linear behavior, which is typical of the class of neurons known as *complex cells* (Hubel and Wiesel, 1968). This is shown by the lack of modulation of the response to the preferred grating (Figure 1B.ii), by the poorly defined spatiotemporal structure of the STA-based RF (Figure 1B.iii), and by the failure of the latter to account for the tuning and time-modulation of the response (Figure 1B.i-ii; compare green and black curves).

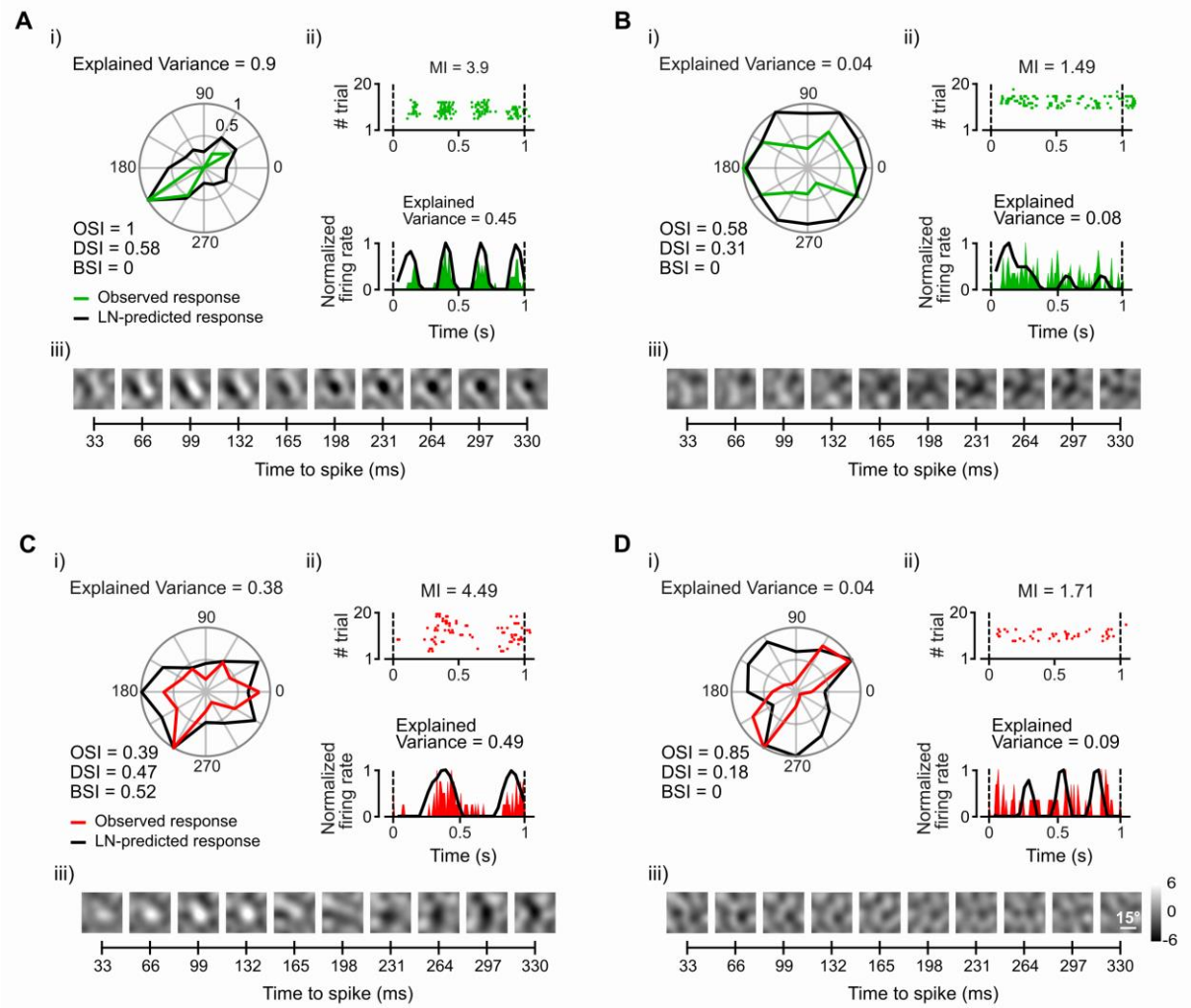

Figure 1 Examples of neuronal tuning in V1 and LL. Representative V1 simple (**A**) and complex (**B**) cells are shown in the upper panels (green/black lines). Representative LL simple-like (**C**) and complex-like (**D**) cells are shown in the lower panels (red/black lines). In each panel, subpanels i-iii illustrate different aspects of neuronal tuning. (i) The polar plot reports the observed (colored curve) and predicted (black curve) tuning of the neuron for the direction of a drifting grating, when shown at the preferred spatial (SF) and temporal (TF) frequencies of the unit. Predictions were obtained using a linear-nonlinear (LN) model, derived from the spike-triggered average (STA) images shown in iii. For ease of comparison, both the observed and predicted curves were normalized to their maximum. The plot also reports the fraction of variance of the tuning curve that was explained by the LN-model, as well as the values of several metrics that quantifies neuronal tuning: the orientation (OSI), direction (DSI) and bimodal (BSI) selectivity indexes. (ii) Raster plot (top) and peri-stimulus time histogram (PSTH; bottom; colored curve), showing the time course of the neuronal response to a drifting grating, when presented at the preferred SF, TF and direction. The bottom plot also reports the time course of the response, as predicted by the LN model (black curve), along with the fraction of variance explained by the model. For ease of

comparison, both the observed and LN-predicted responses were normalized to their maximum. The top plot reports the value of the modulation index (MI), which measured the extent to which the neuronal response was time-locked to the phase of the drifting grating. (iii) Spatial structure and time-evolution of the STA-estimated RF of the neuron. In every STA image, each pixel intensity value was independently z-scored, based on the null distribution of STA values obtained through a permutation test (Materials and Methods). As such, all RF intensity values are reported as distances (in units of standard deviation σ; gray scale bar in panel D.iii) from what expected in the case of no frame-related, linear information carried by the spikes.

Interestingly, simple-like and complex-like units were also found in area LL. For instance, Figure 1C shows an LL neuron with a simple-like behavior, as indicated by the phase modulation of its response (Figure 1C.ii; red dots and curve), which was predicted with reasonable accuracy (black line) by the STA-reconstructed RF. This is consistent with the well-defined spatiotemporal structure of the RF (Figure 1C.iii), which, however, was not able to explain as much variance of the direction-tuning curve (Figure 1C.i; black vs. red curve) as in the case of the example V1 simple cell (Figure 1A.i). This could be due to the sensitivity of the neuron for multiple stimulus orientations, as shown by the large BSI value (0.52), which reflects the presence of two peaks (at 0° and 240°) in the direction-tuning curve (Figure 1C.i; red curve) – a behavior that cannot be captured by a RF with a linear, Gabor-like structure. Figure 1D shows instead an example LL unit that was sharply tuned for a single orientation, but displayed a typical complex-like behavior: its response to the preferred grating was unmodulated (Figure 1D.ii), and both the direction-tuning curve (Figure 1D.i) and the response dynamics (Figure 1D.ii) were poorly explained by the STA-reconstructed RF (red vs. black curves), which, in turn, lacked any clear spatiotemporal structure (Figure 1D.iii).

These examples illustrate how it is possible to compare the level of linearity of the stimulus-response relationship for the neuronal populations recorded in V1 and LL, along with the sharpness and complexity (e.g., bimodality) of their tuning for orientation. To guide such comparison, we derived predictions about the evolution of neuronal tuning along an object processing hierarchy, by assessing how our grating stimuli were represented in HMAX – a neural network model of the ventral stream that is able to account for several qualitative trends in the tuning properties of ventral neurons (Riesenhuber and Poggio, 1999; Serre et al., 2007b; Cadieu et al., 2007; Zoccolan et al., 2007; Cox and Riesenhuber, 2015; Leibo et al., 2017). In its simplest implementation, HMAX is a 4-layered, feed-forward network, with

each layer consisting of either simple-like (S) or complex-like (C) units (Figure 2). The first layer (S1) is a bank of Gabor filters that simulate V1 simple cells. In the second layer (C1), each unit performs a max pooling over a set of S1 afferents with the same orientation tuning, but slightly different RF positions or sizes. This yields orientation-tuned C1 units that are similar to V1 complex cells – i.e., they have larger RFs (dashed circles in Figure 2), with increased position- and size-tolerance, as compared to their S1 afferents. In the third layer (S2), each unit performs a non-linear template-matching computation over a set of C1 afferent units with different preferred orientations. As a result, the S2 units are tuned for complex visual patterns, made of multiple oriented features, like the neurons in higher-level stages of the primate ventral stream, such as cortical area V4 and the inferotemporal cortex (IT). Finally, in the fourth layer (C2), each unit performs again a max pooling over S2 afferents that are tuned for the same visual pattern, but with slightly different RF positions and sizes. This results in C2 units with complex shape selectivity and large tolerance to position and size changes. Here, we used this basic 4-layered architecture to predict how the build-up of shape tuning and invariance should affect neuronal sensitivity to the orientation and temporal frequency of the gratings. In our simulations, layers S1 and C1 correspond to the simple and complex cells of V1, while layers S2 and, above all, C2, correspond to a downstream, higher-level area (such as LL).

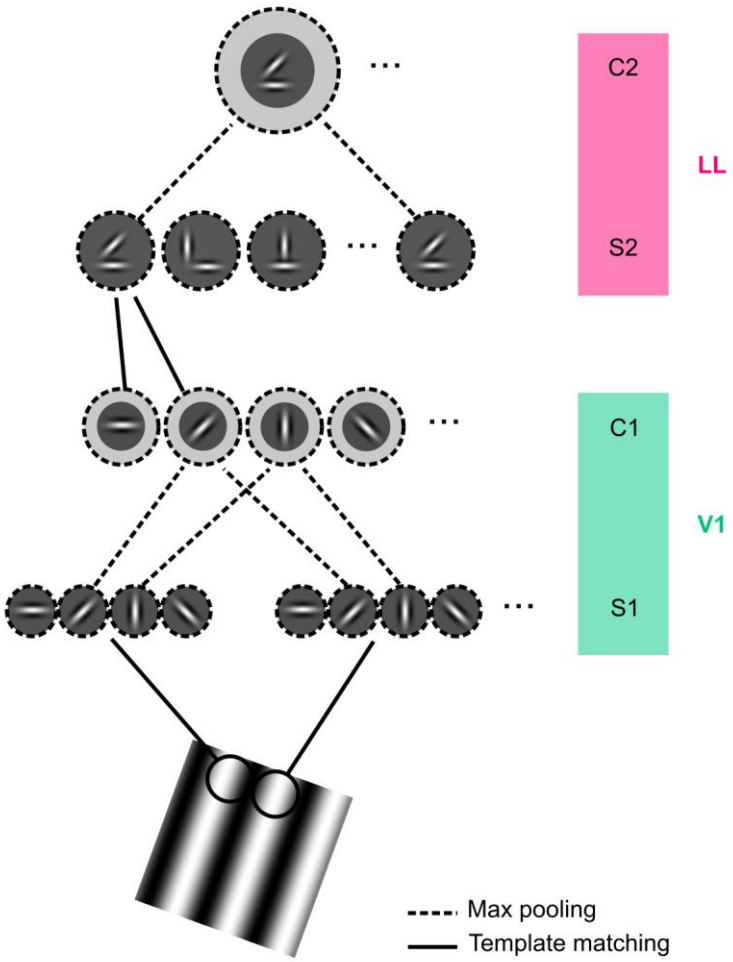

Figure 2 Sketch of the HMAX model implementation used in this study. The model is a feedforward, 4-layered neural network that processes arbitrary input images through a cascade of computations. The first layer (S1) consists of a bank of Gabor filters performing local edge detection on small input image patches (dark gray circles), similar to V1 simple cells. In the second layer (C2), each unit pools over many iso-oriented S1 units via a max-like operation (dashed arrows), thus gaining some amount of position and scale tolerance, similar to V1 complex cells (such expansion of the invariance field is represented by the light gray rings). In the third layer (S2), each unit acquires tuning for more complex visual patterns by combining the output of many C1 units with different orientation via a non-linear template-matching operation (solid arrows). In the fourth layer, each C2 unit applies again the max-pooling operation to S2 units with the same shape tuning but different RF positions and sizes, in order to further gain tolerance to such transformations. In our application, the input images were drifting gratings, similar to those used in the neurophysiology experiment, so as to obtain predictions about the evolution of the relevant tuning indexes (MI, OSI and BSI) across consecutive stages of a ventral-like hierarchy.

The linearity of the stimulus-response relationship decreases from V1 to LL

To establish whether V1 and LL neurons differed in terms of the linearity of the stimulus-response relationship, we first compared the extent to which neuronal responses in the two areas were modulated by the phase of the drifting gratings. Despite the presence of both strongly phase-modulated and unmodulated units in both areas, the distributions of the modulation index (MI) were strikingly different in V1 and LL (Figure 3A). In agreement with earlier studies of rat (Girman et al., 1999) and mouse (Niell and Stryker, 2008; Van den Bergh et al., 2010) primary visual cortex, the MI distribution obtained in V1 (green curve) revealed the existence of two roughly equi-populated classes of neurons, falling, respectively, below and above the MI = 3 threshold that distinguishes complex-like from simple-like cells. By contrast, in LL, most MI values were lower than 3, following a distribution (red curve) with a prominent peak at MI < 1, which is indicative of a strong prevalence of complex-like units. As a result, the median MI was much lower in LL than in V1 (1.28 vs. 2.55; red vs. green line) and this difference was highly significant ($p < 0.001$; Mann-Whitney U-test). Similarly, the fraction of simple-like cells in LL was half as large as in V1 (Figure 3B; $p < 0.001$, χ^2 test).

As illustrated in Figure 1D.ii, the poor sensitivity to the phase of the drifting gratings observed in LL indicates that most neurons in this area respond to phasic stimuli with a sustained firing. This suggests that, in LL, most neurons are tolerant to position changes of their preferred visual stimuli. This interpretation was supported by the evolution of the modulation index observed in HMAX (Figure 2C), where phase-sensitive cells were only found in the bottom layer (S1), corresponding to V1 simple cells. By contrast, from the C1 layer onward, all units displayed a fully phase-invariant behavior, because of the nonlinear max pooling that, in C1, implements position and size invariance. As a result, the combined distribution of MI values obtained in S1 and C1 was in qualitative agreement with the mixture of simple and complex cells observed in V1 (compare to Figure 1A; green curve). Similarly, the extreme phase invariance observed in S2 and C2 was qualitatively consistent with the stark prevalence of complex-like cells found in LL (> 80%; compare to Figure 1A and B; red curve/bar).

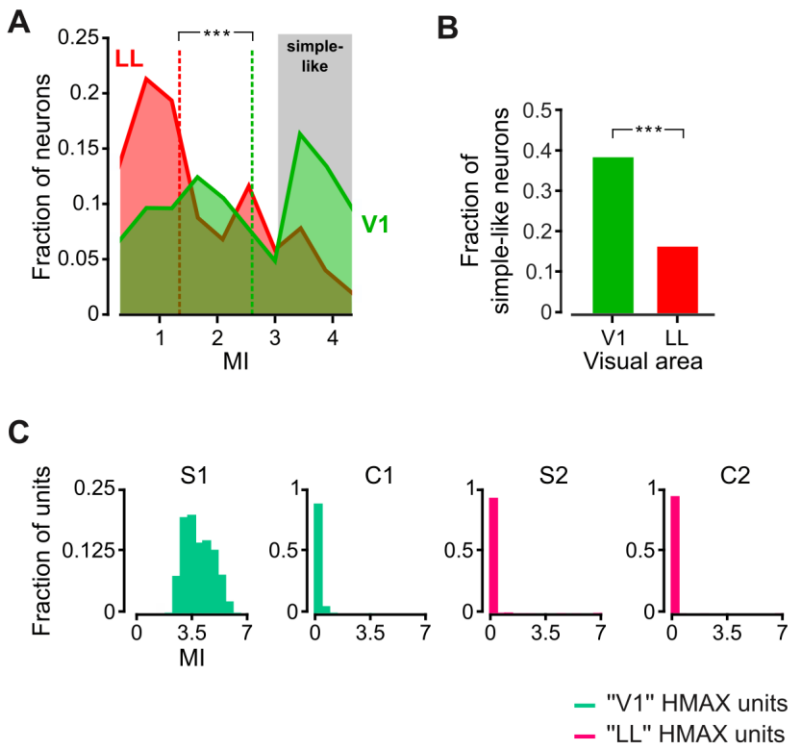

Figure 3 The entrainment of the neuronal response to the phase of drifting gratings decreases from V1 to LL. **A.** Distributions of the modulation index (MI) for the populations of V1 (green; n = 105) and LL (red; n = 104) neurons. Medians are shown as dashed lines (***p < 0.001; one-tailed Mann-Whitney U-test). The gray-shaded area highlights the region of the MI axis (MI > 3) corresponding to strongly phase-modulated units, i.e., simple-like cells. **B.** Fraction of simple-like cells (i.e., units with MI > 3) in V1 (green) and LL (red; ***p < 0.001; χ^2 test). **c.** Evolution of the MI distribution across the layers of the HMAX model (Figure 2). Note the different scale on the ordinate axis for the S1 layer, as compared to downstream layers.

To further assess the linearity of the stimulus-response relationship in V1 and LL, we evaluated the quality of the RFs inferred through the STA method. To this aim, each pixel intensity value in a STA image was z-scored, based on the null distribution of STA values obtained for that pixel, after randomly permuting the association between frames of the movie and spike times 50 times (Materials and Methods). This allowed expressing the intensity values of the STA image in terms of their difference (in units of standard deviation) from what expected in the case of no frame-related information carried by the spikes. We then computed a *contrast index* (CI; Materials and Methods and Figure 4-figure supplement 1) to measure the amount of signal contained in the image and establish a conservative

threshold (CI > 5.5), beyond which the linear RF structure of a neuron could be considered as well defined.

Figure 4A shows the 10 best STA-based RFs in each area, sorted according to the magnitude of their CI, from largest to smallest. In V1, most of these RFs featured equally prominent excitatory and inhibitory regions, often displaying a sharp, Gabor-like structure, while, in LL, the best STA images typically presented a single lobe, with much lower contrast. As a result, the median CI was higher for V1 than for LL neurons (green vs. red dashed line in Figure 4B; $p < 0.001$; one-tailed Mann-Whitney U-test), and the fraction of units with well-defined linear RFs (i.e., with CI > 5.5) was 5 times larger in V1 than in LL (Figure 4C; $p < 0.001$, χ^2 test). We also observed a stark difference between the two areas in terms of the number of lobes contained in a STA image. To count the lobes, we binarized each image, by applying a threshold to the modulus of its intensity values (Materials and Methods and Figure 4-figure supplement 1; this analysis was restricted to the units with well-defined linear RFs). Regardless of the choice of the binarization threshold, the distribution of lobe counts peaked at 1 in both areas, but had a heavier tail in V1 than in LL (compare matching rows in Figure 4D), resulting, in most cases, in a statistically significant difference between the two areas (Fisher's exact test; * $p < 0.05$; ** $p < 0.01$). Overall, this indicates that, in many cases, the responses of V1 neurons could be well approximated as a linear, weighted sum of the luminance intensity values falling inside their RFs – hence the sharp, multi-lobed STA images. By comparison, the lower-contrast, structurally simpler RFs yielded by STA in LL suggest a dominant role of non-linear computations in establishing the stimulus-response relationship in this area. In fact, in the case of prevalently non-linear units, the STA method would be able to reconstruct at most their linear sub-parts (Fournier et al., 2011, 2014; Sedigh-Sarvestani et al., 2014) – a sort of linear "residue" of the actual non-linear filter. Such linear residue would still show, in some cases, the position and main orientation of the underlying filter, but would fail short at accounting for the full richness and complexity of its structure. This conclusion was supported by the tendency (although not significant) of the STA-based RFs to explain a lower fraction of variance of the orientation tuning curves in LL, as compared to V1 (Figure 4-figure supplement 2). Taken together with the prevalence of complex-like units in LL (Figure 3), these findings are consistent with the non-linear computations that are expected to build invariance along an object-processing hierarchy.

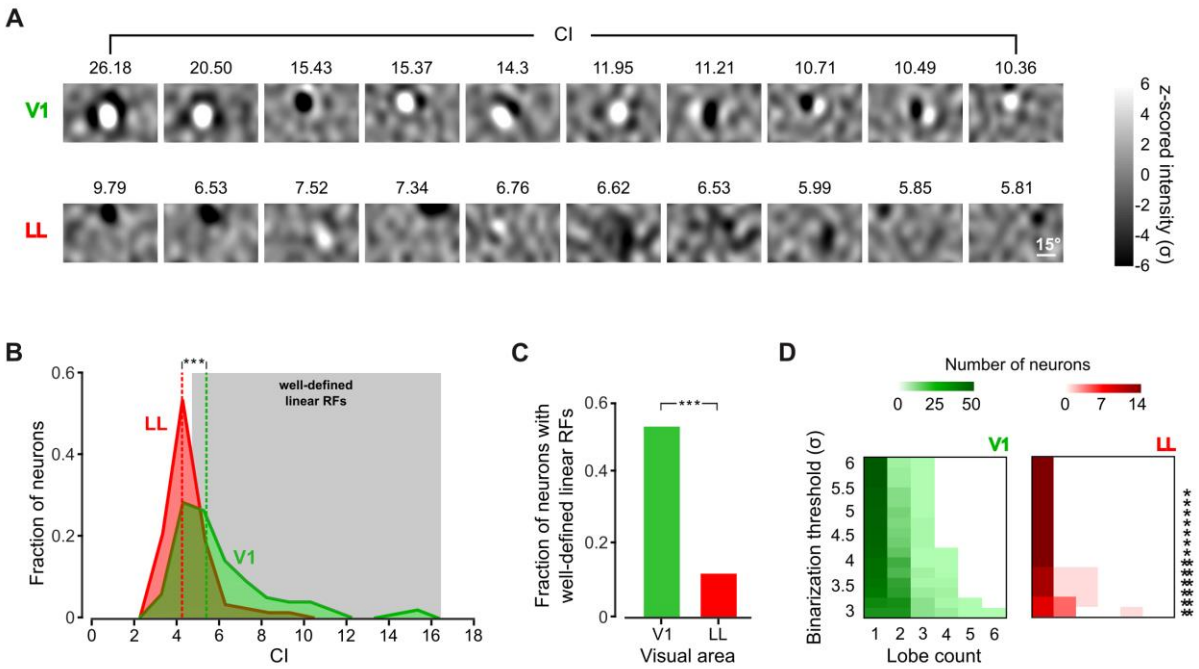

Figure 4 Spike-triggered average yields better-defined and structurally richer receptive fields in V1 than in LL. **A.** Spatial structure of the ten best linear RFs obtained in V1 (top row) and LL (bottom row) using the STA method. The quality of each RF was assessed by computing a contrast index (CI) that measured the amount of signal contained in a STA image (Figure 4-figure supplement 1A and C and Materials and Methods). For each neuron, the RF shown here is taken at the time before the spike when CI was maximum (the corresponding CI value is reported above each STA image). As explained in Figure 1, each pixel intensity value in a STA image was independently z-scored according to the null distribution yielded by a permutation test; the resulting z-scored values are reported here in units of standard deviation σ (gray scale bar). **B.** CI distributions for the populations of V1 (green; n = 105) and LL (red; n = 104) neurons. Medians are shown as dashed lines (***p < 0.001; one-tailed Mann-Whitney U-test). The gray-shaded area highlights the region of the CI axis (CI > 5.5) corresponding to units with well-defined linear RFs. **C.** Fraction of units with well-defined linear RFs (i.e., with CI > 5.5) in V1 (green) and LL (red; ***p < 0.001; χ^2 test). **D.** Heat maps showing the distributions of lobe counts (abscissa) for the units with well-defined linear RFs (i.e., with CI > 5.5) in V1 (green; n = 13) and LL (red; n = 56) as a function of the binarization threshold (ordinate) used by the lobe-counting algorithm (Figure 4-figure supplement 1B and D and Materials and Methods). Stars indicate threshold values for which the two distributions were significantly different (*p < 0.05, **p < 0.01; Fisher exact test). The procedures applied to count the lobes and to compute the contrast index are illustrated in ***Figure 4-figure supplement 1***. The fraction of variance of the direction-tuning curves that was explained by STA-based LN models in V1 and LL is reported in ***Figure 4-figure supplement 2***.

Sharpness and complexity of orientation tuning follow opposite trends from V1 to LL

Next, we compared the V1 and LL neuronal populations in terms of the shape and sharpness of the orientation tuning curves obtained in the two areas. As shown in Figure 5A, in area LL most units had OSI lower than 0.5 (red curve), with a median of 0.41 (red line), while in V1 the peak of the OSI distribution was in the 0.6-0.8 range, with a median of 0.53. The difference between the two OSI distributions was highly significant ($p < 0.01$; Kolmogorov-Smirnov test), as it was the difference between their medians ($p < 0.001$; one-tailed, Mann-Whitney U-test). Following a convention established in previous studies (Figure 7-table supplement 2) we classified as orientation-selective the units with OSI > 0.6, and we compared the fraction of units above such threshold in the two areas (Figure 5B). In V1, the proportion of selective cells was twice as large as in LL (40% vs. 18%), and such difference was, again, highly significant ($p < 0.001$; χ^2 test).

To check the consistency of this finding with ventral-like processing, we looked at the evolution of orientation tuning in HMAX. As shown in Figure 5C, in the model, the OSI distribution underwent an abrupt shift towards much lower values in the transition from the C1 layer, corresponding to V1 complex cells, to the S2 layer, corresponding to a downstream area, where tuning for combinations of multiple oriented elements is achieved using a non-linear template matching operation (see Figure 2). The shape of the OSI distribution was then maintained in the C2 layer, performing the max-pooling operation over the S2 afferents. This trend shows that, in HMAX, building tuning for complex visual patterns leads to a strong reduction of orientation tuning. As such, the drop of orientation selectivity observed from V1 to LL can be taken as indication that LL neurons are tuned to more complex combinations of visual features, as compared to V1 units.

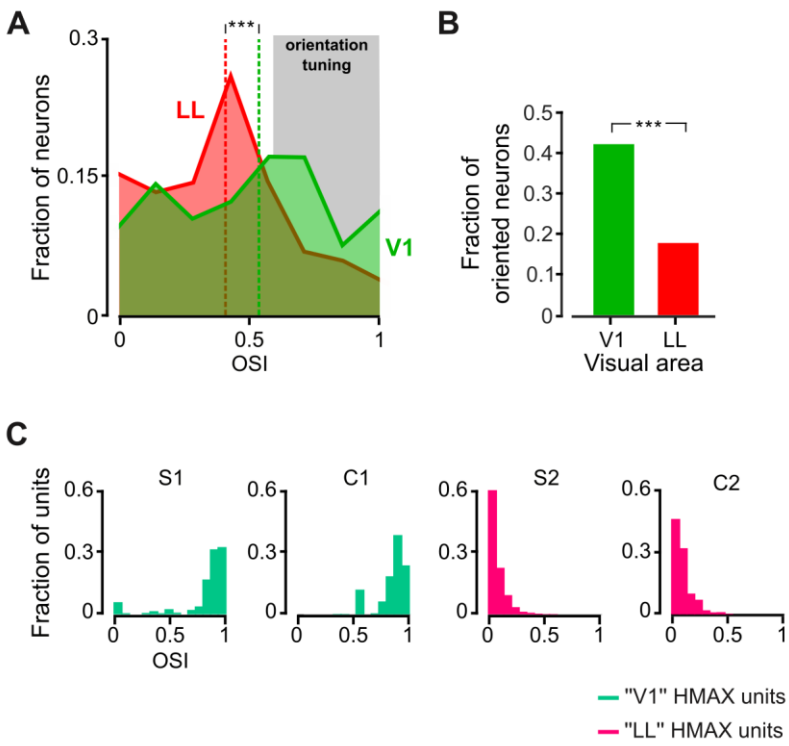

Figure 5 Orientation selectivity decreases from V1 to LL. **A.** Distributions of the orientation selectivity index (OSI) for the populations of V1 (green; n = 105) and LL (red; n = 104) neurons. Medians are shown as dashed lines (**p < 0.01; one-tailed Mann-Whitney U-test). The gray-shaded area highlights the region of the OSI axis (OSI > 0.6) corresponding to orientation-selective units. **B.** Fraction of orientation-selective cells (i.e., units with OSI > 0.6) in V1 (green) and LL (red; ***p < 0.001; χ^2 test). **c.** Evolution of the OSI distribution across the layers of the HMAX model (Figure 2). The distributions of the direction selectivity index (DSI) for the V1 and LL populations are reported in ***Figure 5-figure supplement 1***.

Support to this conclusion came from comparing the two areas in terms of the tendency of the orientation tuning curves to be bimodal (Figure 6A). In V1, most units had BSI equal or close to zero (i.e., virtually perfect unimodal tuning), with the distribution dropping sharply at larger BSI values (green curve). In LL, the BSI distribution also peaked at zero, but featured a heavier tail, with a secondary peak above 0.5 (red curve). As a result, the two distributions and their medians were significantly different (p < 0.05; Kolmogorov-Smirnov and one-tailed, Mann-Whitney U-test, respectively) and the fraction of neurons with bimodal tuning (i.e., with BSI > 0.5) was more than twice as large in LL as in V1 (p < 0.05, χ^2 test; Figure 6B). Once again, a similar trend was observed in HMAX (Figure 6C), where the BSI distribution displayed a much heavier tail in the deeper layers (S2 and C2) than in the early ones (S1 and C1). This supports the interpretation that the decrease of OSI from V1 to

LL, along with the concomitant increase of BSI, results from an increment of the complexity of tuning of LL neurons – a property that is consistent with their involvement in ventral processing. By contrast, no significant differences were found in terms of direction selectivity (Figure 5-figure supplement 1), which suggests a lack of specialization of LL for motion processing, as compared to V1.

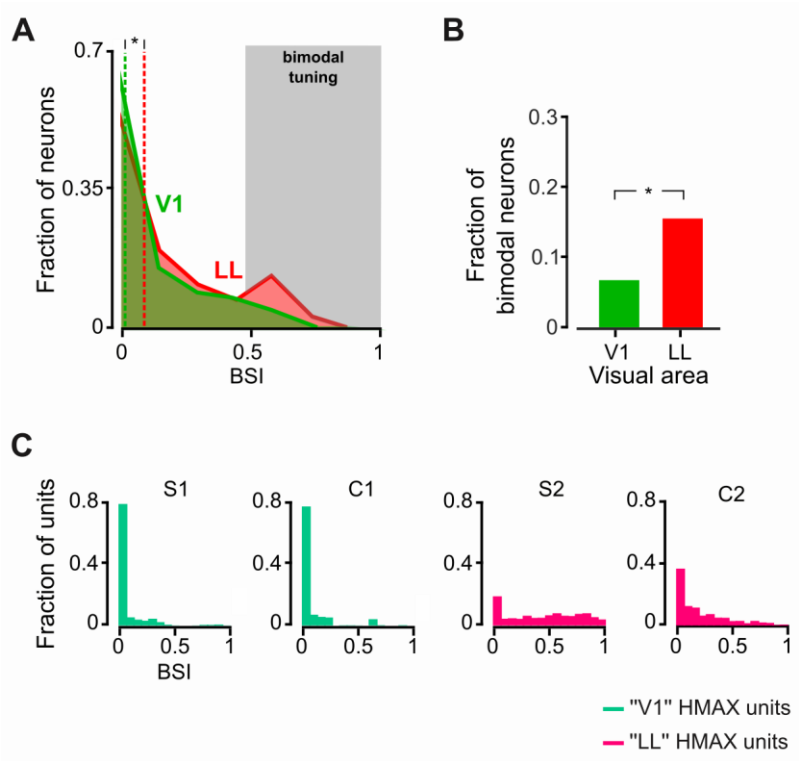

Figure 6 The tendency to be tuned for multiple orientations increases from V1 to LL. **A.** Distributions of the bimodal selectivity index (BSI) for the populations of V1 (green; n = 105) and LL (red; n = 104) neurons. Medians are shown as dashed lines (*p < 0.05; one-tailed Mann-Whitney U-test). The gray-shaded area highlights the region of the BSI axis (BSI > 0.5) corresponding to units with bimodal tuning. **B.** Fraction of cells with bimodal tuning (i.e., with BSI > 0.5) in V1 (green) and LL (red; *p < 0.05; χ^2 test). **C.** Evolution of the BSI distribution across the layers of the HMAX model (Figure 2).

Discussion

In this study, we sought to understand which tuning properties are at the base of the specialization of LL (the most temporal among rat lateral extrastriate areas) for the processing of visual object information, as revealed by our previous investigation of this region and nearby visual cortical areas (Tafazoli et al., 2017). To be fully complementary

with our previous study, we did not probe LL representations using visual objects. With the goal of directly targeting those core tuning processes that are responsible for building up transformation tolerance and feature selectivity, we deployed instead drifting oriented gratings and noise movies, and we compared the V1 and LL neuronal populations in terms of: 1) the phase-sensitivity of their responses; 2) the goodness of the RF structures recovered through STA; and 3) the sharpness and complexity (i.e., bimodality) of their tuning for orientation. This approach required obtaining clear predictions about how these properties should evolve along a ventral-like pathway. In our study, we achieved such predictions based on the simulations of a computational model of the ventral stream (HMAX) and a careful survey of the monkey literature.

HMAX is a feedforward neural network that was originally developed to explain the emergence of selective, yet transformation-tolerant representations of visual objects along the primate ventral stream (Riesenhuber and Poggio, 1999). Over the years, the model has provided useful computational insights about shape tuning in V4 and IT (Riesenhuber and Poggio, 1999; Serre et al., 2007b; Cadieu et al., 2007; Zoccolan et al., 2007; Cox and Riesenhuber, 2015; Leibo et al., 2017), and human perception of visual objects (Serre et al., 2007a; Cox and Riesenhuber, 2015). The model shares the same basic architecture of more recent artificial vision systems [e.g., deep convolutional neural networks (LeCun et al., 2015)], with alternating layers of selectivity-increasing and tolerance-increasing pooling, but it is computationally simpler, since, in its basic instantiation, it has only 4 layers and does not involve any supervised learning of feature representations. While this can easily explain why HMAX has been surpassed by deep-learning networks as a state of the art machine vision system (Khaligh-Razavi and Kriegeskorte, 2014; Cadieu et al., 2014; Kheradpisheh et al., 2016), its simpler and static structure makes it an ideal conceptual model to isolate the role of the selectivity- and invariance-building operations in establishing the tuning properties of ventral neurons. As shown by our simulations, HMAX makes very specific predictions about how the tuning for oriented gratings should evolve from V1 to downstream ventral areas. Because of the tolerance-building operation (the max-pooling), phase-sensitivity should sharply decrease (Figure 3C). At the same time, because of the selectivity-building computation (the template-matching pooling), orientation tuning should become smaller (Figure 5C), while neurons acquire preference for multiple orientations (Figure 6C). Critically, all these trends were matched by the empirical evidence gathered in our study, when comparing V1 to LL (Figures 3A-B, 5A-B and 6A-B), thus suggesting that similar

selectivity- and invariance-building computations are at work along the progression of lateral extrastriate areas that culminates with LL.

To check if this conclusion was also supported by a comparison with the monkey literature, we performed a simple meta-analysis of the trends of variation of orientation tuning and phase-sensitivity to drifting gratings across areas V1, V4 and IT of the ventral stream. In V1, the fraction of simple cells reported by various authors is quite variable (Hubel and Wiesel, 1968; Schiller et al., 1976; De Valois et al., 1982a, 1982b; Foster K H et al., 1985; O'Keefe et al., 1998; Ringach et al., 2002; Kagan et al., 2002; Aronov et al., 2003; Gur et al., 2005), ranging from ~16% to 63% (Figure 7-table supplement 1), with a mean across studies of ~41% (green bar in Figure 7A). According to the few studies that investigated the phase-dependence of neuronal responses to grating stimuli in downstream areas (Figure 7-table supplement 1), the proportion of simple-like cells diminishes to ~14% in V4 (Desimone and Schein, 1987; Gallant et al., 1993, 1996), to become zero in IT (Pollen et al., 1984) (brown and red bars in Figure 7A), resulting in an overall significant decrement along the ventral stream ($p < 0.01$, $F_{2,13} = 7.25$, one-way ANOVA). With regard to orientation selectivity, most authors report a very large fraction of orientation-tuned cells in monkey V1 (Schiller et al., 1976; De Valois et al., 1982a, 1982b; Vogels and Orban, 1991; Zhou et al., 2000; Ringach et al., 2002; David et al., 2006) (Figure 7-table supplement 2), with a mean across studies close to 89% (green bar in Figure 7B). This proportion significantly decreases along downstream areas ($p < 0.001$, $F_{2,14} = 46.17$, one-way ANOVA), with ~65% and ~19% of orientation-tuned neurons reported, respectively, in V4 (Desimone and Schein, 1987; Pasupathy and Connor, 1999; Zhou et al., 2000; David et al., 2006) and IT (Desimone et al., 1984; Vogels and Orban, 1993, 1994; Sary et al., 1995) (brown and red bars in Figure 7B; see also Figure 6-table supplement 2). These trends are confirmed by the few studies that applied the same experimental design to directly compare orientation tuning in V1 and either V4 (David et al., 2006) or IT (Vogels and Orban, 1991, 1993). Vogels and Orban found ~85% of orientation-tuned neurons in V1, compared to just ~14% in IT. David and colleagues reported a similar figure for V1, while in V4 they found that only 50% of neurons were orientation-tuned. In addition, they found that the fraction of cells with bimodal tuning was more than twice as large in V4 (~28%) as in V1 (~11%), in agreement with the increasing proportion of neurons that, in V4, are tuned for the combinations of oriented elements found in corners and curved boundaries (Gallant et al., 1993, 1996; Pasupathy and Connor, 1999, 2002).

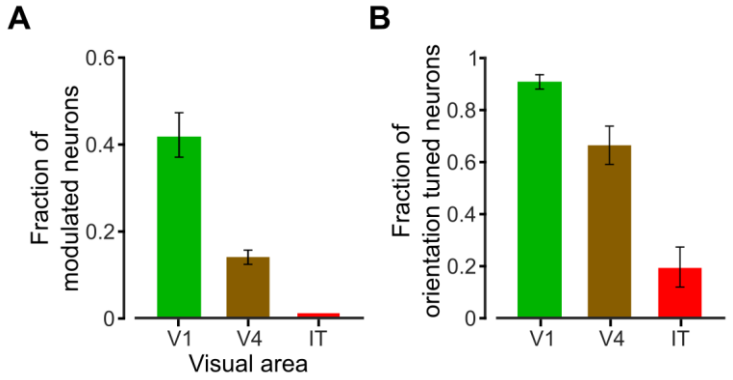

Figure 7 Meta-analysis of the trends of variation of phase sensitivity and orientation tuning to drifting gratings across the primate ventral stream. **A.** Fraction of cells that have been reported as being strongly modulated by the phase of drifting gratings (i.e., simple-like cells) in areas V1 (red; n = 10), V4 (brown; n = 3) and IT (red; n = 1) of the monkey ventral stream (mean across n studies in each area ± SE; see Figure 7-table supplement 1). The drop of phase modulation across the three areas was statistically significant (p < 0.01, one-way ANOVA). **B.** Fraction of cells that have been reported as being orientation tuned in areas V1 (red; n = 7), V4 (brown; n = 4) and IT (red; n = 4) of the monkey ventral stream (mean across n studies in each area ± SE; see Figure 7-table supplement 2). The drop of orientation tuning across the three areas was statistically significant (p < 0.001, one-way ANOVA). The full list of research papers included in the meta-analysis summarized by panel A is reported in ***Figure 7-table supplement 1***. The full list of research papers included in the meta-analysis summarized by panel B is reported in ***Figure 7-table supplement 2***.

Among rodent studies of visual extrastriate areas, our results are the first that are in complete agreement with the monkey literature reviewed above and are fully consistent with the computational principles thought to be at the root of ventral processing (as implemented in the HMAX simulations). In mice, area LL has not been identified, but areas LM (lateromedial) and LI (laterointermediate), which, as in rats, are located progressively more lateral to V1 [as a reference, see Figure 1A of our previous study (Tafazoli et al., 2017)], have been attributed to the ventral stream based on anatomical considerations (Wang et al., 2011b, 2012b) and have been tested with drifting gratings in several electrophysiological and imaging studies, often along with other medial and parietal visual areas (Van den Bergh et al., 2010; Andermann et al., 2011; Marshel et al., 2011; Roth et al., 2012; Smith et al., 2017). Consistently with our findings (Figure 3), the only study that measured the phase dependence of neuronal responses in V1 and lateral visual areas of the mouse found that the latter had a lower proportion of simple cells (15% vs. 38% in V1) (Van den Bergh et al., 2010).

Conversely, the authors reported no difference between the areas in terms of OSI. This result contrasts with the general increase of orientation tuning that was found in nearly all extrastriate areas (compared to V1) by (Marshel et al., 2011), although the increment of OSI was much larger in putative dorsal-stream areas than in LM and LI. A significant increase of OSI in LM, relative to V1, was also observed by (Tohmi et al., 2014), while other authors found no differences between V1 and some of the putative dorsal-stream areas in terms of orientation tuning (Andermann et al., 2011; Roth et al., 2012). As for rat cortex, (Vermaercke et al., 2014) found the opposite trend of that observed in our study – i.e., a progressive increase of OSI from V1 to LL and beyond.

Different causes can be at the root of such discrepancies among rodent studies. In mice, many experiments were based on calcium imaging, performed using different indicators [e.g., OGB-1 (Marshel et al., 2011) vs. GCaMP3 (Andermann et al., 2011)], which diverge from a perfectly linear relationship between fluorescence changes and underlying spiking activity in an-indicator specific way (Akerboom et al., 2012). As pointed out by (Niell, 2011), this could potentially distort the orientation tuning curves and bias the OSI measurements, as compared to approaches where well-isolated single-units are recorded with micro-electrodes, as in our study. Other important differences could concern the specific stimuli that were used by different authors. For instance, while we presented very large (60ºx110º), full-field gratings, (Vermaercke et al., 2014) used smaller (33º) circular gratings, shown at the center of each recorded receptive field. Our presentation protocol ensured that, for most neurons in both areas, the stimuli covered the entire *classical* receptive field and, very likely, a large fraction of the *extra-classical* receptive field, where surround-suppression contributes to shape tuning both in monkey (Maunsell and Newsome, 1987; Orban, 2008) and rodent (Adesnik et al., 2012; Vaiceliunaite et al., 2013; Self et al., 2014; Pecka et al., 2014; Alwis et al., 2016) visual cortex. By contrast, the smaller and more localized gratings used in (Vermaercke et al., 2014) likely covered an increasingly smaller fraction of both the classical and extra-classical receptive fields along the V1-LM-LI-LL progression, given that, as shown in our previous study (Tafazoli et al., 2017), RF size gradually increases along rat lateral extrastriate areas, being twice as large in LL as in V1. This, in turn, may have prevented the tuning for multiple oriented elements, located at different RF locations, to emerge in neurons with larger RFs, thus artificially boosting OSI in higher-order areas. Finally, another factor that is worth considering is whether neuronal recordings were performed in anesthetized or awake animals. However, as discussed at length in our previous

study (Tafazoli et al., 2017), most studies comparing neuronal activity in mouse visual cortex under anesthesia and during active wakefulness found that several key tuning properties, both in V1 and extrastriate areas, were very similar in the two conditions (Niell and Stryker, 2008, 2010; Froudarakis et al., 2014; Juavinett and Callaway, 2015; Durand et al., 2016).

To summarize, it is difficult to reconcile the variety of findings about the trends of variation of orientation tuning across rodent visual cortical areas, without performing dedicated experiments (which are beyond the scope of our study) – e.g., imaging vs. micro-electrode recordings, full-filed vs. localized stimuli, or anesthetized vs. awake recordings. Ultimately, five considerations support the strength and reliability of our conclusions, as compared to previous studies of rodent visual cortex based on gratings. First, the full agreement with the trends found across the monkey ventral stream. Second, the full agreement with the predictions derived by simulating the basic computational architecture of the ventral stream. Third, the self-consistency of our findings – for instance, the reduction of orientation tuning from V1 to LL was matched by a concomitant increase of bimodal tuning, with both trends pointing to the development of selectivity for complex visual patterns in LL. Fourth, the agreement with our previous investigation of rat lateral visual areas, where we found the signature of specialized processing of visual objects (Tafazoli et al., 2017). Fifth, the consistency with the advanced processing of shape information reported, for rats, by several behavioral studies (Alemi-Neissi et al., 2013; Rosselli et al., 2015; Djurdjevic et al., 2018).

In the light of these arguments, we believe that our results nicely complement, corroborate and extend those of our previous study (Tafazoli et al., 2017). In fact, they show how some key neuronal tuning properties, which are reveling of the build-up of transformation tolerance and shape selectivity along an object-processing pathway, naturally evolve, from V1 to LL, in a direction that is consistent with the previously demonstrated ability of LL to support transformation-tolerant object recognition. In addition, being our experimental approach based on the presentation of widely-used parametric stimuli and the application of intuitive metrics, it provides a direct, standardized and easily interpretable way to carry out a large-scale screening of ventral processing across rodent visual cortical areas. This could help optical imaging studies finding the yet missing functional signature of ventral processing in mouse visual cortex – e.g., by analyzing existing large datasets of stimulus-response associations, as the one made available by the Allen Brain Observatory (Hawrylycz et al., 2016). In this regard, having provided clear predictions about the trends of variation of

the modulation, orientation and bimodal selectivity indexes along the ventral stream, based on computational modeling and an exhaustive survey of the monkey literature, is another major contribution of our study. Such predictions, which were missing from previous screenings of rodent visual cortex, will likely serve as a reference to guide future explorations of ventral functions in both mice ad rats.

Methods

Animal preparation and surgery, neuronal recordings and histology. All animal procedures were in agreement with international and institutional standards for the care and use of animals in research and were approved by the Italian Ministry of Health: project N. DGSAF 22791-A, submitted on Sep. 7, 2015 and approved on Dec. 10, 2015 (approval N. 1254/ 2015-PR). We performed extracellular recordings from either primary visual cortex (V1) or laterolateral (LL) extrastriate visual cortex of 18 naïve, anesthetized Long-Evans male rats (Charles River Laboratories), with age 3-12 months and weight 300-700 grams, using single-shank 32-channel silicon probes (NeuroNexus Technologies) with site recording area of 775 μm^2 and 25 μm of inter-site spacing. The apparatus and procedures for anesthesia, surgery, neuronal recordings and histology were the same as described in our previous study (Tafazoli et al., 2017).

Visual stimuli. During a recording session, two kinds of visual stimulation protocols were administered to the rat.

A 15 min-long receptive field (RF) mapping procedure was used to precisely identify the visual area each neuron was recorded from (see details below) and optimize the location of the RFs for the main presentation protocol. This procedure was identical to the one described in our previous study [see Figure 1C in (Tafazoli et al., 2017)], with the only difference that no tangent screen projection was applied to the stimuli.

Once the probe was positioned in the final recording location, the main presentation protocol was administered. The rat was presented with: 1) 1s-long drifting gratings, made of all possible combinations of 3 spatial frequencies (SF; 0.02, 0.04 and 0.08 cpd), 3 temporal frequencies (TF; 2, 4 and 8 Hz), and 12 directions (from 0° to 330°, in 30° increments); and 2) 30s-long movies of spatially and temporally correlated noise. To allow for a precise

estimate of neuronal tuning properties, each grating stimulus was presented in 20 repeated trials, while 80 distinct noise movies were shown for an overall duration of 40 minutes. All stimulus conditions were randomly interleaved, with a 1s-long inter stimulus interval (ISI), during which the display was set to a uniform, middle-gray luminance level (later used to estimate spontaneous firing rates; see below). To generate the movies, random white noise movies were spatially correlated by convolving them with a Gaussian kernel. The kernel full width at half maximum (FWHM) was chosen to match a 0.02 cpd SF, which was found to elicit a robust spiking response in a series of pilot experiments. For the same reason (i.e., to increase the chance of evoking spikes), temporal correlations were also introduced, by convolving the movies with a causal exponential kernel with a 33 ms decay time-constant.

Stimuli were generated and controlled in MATLAB (The MathWorks) using the Psychophysics Toolbox package and displayed with gamma correction on a 47-inch LCD monitor (SHARP PNE471R) with 1920x1080 pixel resolution, 220 cd/m^2 maximum brightness and spanning a visual angle of 110° azimuth and 60° elevation. Grating stimuli were presented at 60 Hz refresh rate, whereas noise movies were played at 30 Hz.

Selection of the single units included in the analyses. Responses of single units were isolated offline by applying the spike sorting package KlustaKwik-Phy (Rossant et al., 2016). Automated spike detection, feature extraction and expectation maximization (EM) clustering were followed by manual refinement of the sorting using a customized version of the "Kwik-GUI" interface. Specifically, the manual curation of the automatic output was performed by taking into consideration many features of the candidate clusters: 1) the distance between their centroids and their compactness in the space of the principal components of the waveforms (a key measure of goodness of spike isolation); 2) the shape of the auto- and cross-correlograms (important to decide whether to merge or not two clusters); 3) the variation, over time, of the principal component coefficients of the waveform (important to detect and take into account possible electrode drifts); and 4) the shape of the average waveform (to exclude, as artifacts, clearly non-physiological signals). Clusters suspected to contain a mixture of one or more single units were separated using the "reclustering" feature of the GUI (i.e., by rerunning the EM clustering algorithm on the spikes of these clusters only). At the end of the manual refinement step, only responsive, well-isolated single units, with reproducible firing across repeated stimulus presentations, were included in the dataset used to perform all subsequent analyses. Specifically, such units were defined by having: i)

less than 0.5% of "rogue" spikes within 2 ms in their autocorrelogram (i.e., units displaying a clear refractory period); ii) a mean stimulus response firing rate of 2 spikes/s above baseline (i.e., responsive units); and iii) responsiveness in at least 30% of the trials in one or more stimulus conditions (i.e., stable isolation during the recording and reproducible stimulus-evoked response). We also applied an additional, very loose screening on the reproducibility of the firing rate across repeated stimulus presentations, by including only neurons with a mean correlation coefficient of the direction-tuning curve across trials not lower than 0.03. The average baseline (spontaneous) firing-rate of each well-isolated unit was computed by averaging its spiking activity over every ISI condition. These criteria led to the selection of 105 units in V1 and 104 units in LL.

Analysis of the responses to drifting gratings. The response of a neuron to a drifting grating was computed by counting how many spikes the neuron fired (on average, across repeated presentations of the stimulus) during the whole duration of stimulus presentation, and then subtracting the spontaneous firing rate (see above). To quantify the tuning of a neuron for the orientation and direction of drifting gratings, we computed two standard metrics, the orientation and direction selectivity indexes (OSI and DSI), which are defined as: $\mathrm{OSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{ortho}}\right)/\left(R_{\mathrm{pref}} + R_{\mathrm{ortho}}\right)$, and $\mathrm{DSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{opposite}}\right)/\left(R_{\mathrm{pref}} + R_{\mathrm{opposite}}\right)$, where R_{pref} is the response of the neuron to the preferred direction, R_{ortho} is the response to the orthogonal direction, relative to the preferred one (i.e., $R_{\mathrm{ortho}} = R_{\mathrm{pref}} + \pi/2$), and R_{opposite} is the response to the opposite direction, relative to the preferred one (i.e., $R_{\mathrm{opposite}} = R_{\mathrm{pref}} + \pi$). Values close to one indicate very sharp tuning, whereas values close to zero are typical of untuned units.

The modulation of the spiking response to drifting gratings at the temporal frequency f_1 of the stimulus was quantified by a modulation index (MI) adapted from (Wypych et al., 2012) and defined as:

$$\mathrm{MI} = \left| \frac{PS(f_1) - \langle PS \rangle_f}{\sqrt{\langle PS^2 \rangle_f - \langle PS \rangle_f^2}} \right|,$$

where PS indicates the power spectral density of the stimulus-evoked response, i.e., of the peri-stimulus time histogram (PSTH), and $\langle\ \rangle_f$ denotes the average over frequencies. This

metric measures the difference between the power of the response at the stimulus frequency and the average value of the power spectrum in units of its standard deviation. The power spectrum was computed by applying the Blackman-Tukey estimation method to the baseline-subtracted, 10ms-binned PSTH. Being MI a standardized measure, values greater than 3 can be interpreted as signaling a strong modulation of the firing rate at the stimulus frequency (typical of simple-like cells), whereas values smaller than 3 indicate poor modulation (typical of complex-like cells).

Bimodal tuning (i.e., the presence of two distinct peaks in the orientation-tuning curve) was quantified by the Bimodal Selectivity Index (David et al., 2006; Cadieu et al., 2007), defined as $\mathrm{BSI} = \left(R_{\mathrm{2nd\ peak}} - R_{\mathrm{2nd\ trough}}\right)/\left(R_{\mathrm{1st\ peak}} - R_{\mathrm{1st\ trough}}\right)$, where $R_{\mathrm{1st\ peak}}$ and $R_{\mathrm{2nd\ peak}}$ indicate the two largest peaks in the orientation-tuning curve (with $R_{\mathrm{1st\ peak}} > R_{\mathrm{2nd\ peak}}$), and $R_{\mathrm{1st\ trough}}$ and $R_{\mathrm{2nd\ trough}}$ indicate the deepest troughs (with $R_{\mathrm{1st\ trough}} < R_{\mathrm{2nd\ trough}}$). This index takes values close to one for tuning curves displaying two equally high peaks and close to 0 for tuning curves with a single peak.

Analysis of the responses to the noise movies. We used the spike-triggered average (STA) analysis (Schwartz et al., 2006; Sharpee, 2013) to estimate the linear RF structure of each recorded neuron. The method was applied to the spike trains fired by the neuron in response to the spatiotemporally correlated noise (see above). The method yields an ordered sequence of images, each showing the linear approximation of the spatial RF at different times before spike generation (see examples in Figure 1). To take into account the correlation structure of our stimulus ensemble and prevent artifactual blurring of the reconstructed filters, we "decorrelated" the resulting STA images by dividing them by the covariance matrix of the whole stimulus ensemble (Schwartz et al., 2006; Sharpee, 2013), using Tikhonov regularization to handle covariance matrix inversion. Statistical significance of the STA images was then assessed pixelwise via a permutation test. After randomly reshuffling the spike times, the STA analysis was repeated multiple times (n=50) to derive a null distribution of intensity values for the case of no linear stimulus-spike relationship. This allowed z-scoring the actual STA intensity values using the mean and standard deviation of this null distribution (Figure 1 and 4). The temporal span of the spatiotemporal linear kernel reconstructed via STA extended till to 330 ms before spike generation, corresponding to a duration of 10 frames of the noise movie played at 30 Hz (Figure 1).

After estimating the linear spatiotemporal structure of a RF with STA, we used it as a filter in the input stage of a classical Linear-Nonlinear (LN) model of stimulus-response relationship (Schwartz et al., 2006; Sharpee, 2013). To obtain a prediction of the tuning of the neuron over the direction axis, the sequence of frames of each drifting grating was fed as an input to the STA-estimated linear filter. The output of the filter was then passed through a rectifying nonlinearity with unit gain to obtain the final response of the model to each stimulus frame. We finally integrated the response of the LN model through time to predict the activity of the neuron for each direction of the tuning curve. The agreement between the LN-predicted and observed tuning curves was quantified by computing the fraction of explained variance (dark vs. light tuning curves in Figure 1.i). In addition, we also quantified how well the time course of the neuronal response to the preferred grating was explained by the time course of the LN-model prediction (dark vs. light dynamics in Figure 1.ii), again, in terms of explained variance.

We also estimated the amount of signal contained in a given STA image by defining a quality metric based on a measure of maximal local contrast. Specifically, we used MATLAB "rangefilt" function (MATLAB Image Processing toolbox) to compute the local contrast in a STA image over convolutional patches with size equal to one fifth of the total image size and stride 1 (the local contrast in every patch was measured as the difference between the maximum and the minimum image intensity values within the patch itself). From the resulting local-contrast map (Figure 4-figure supplement 1 A and C, middle plot), we obtained a distribution of contrast intensity values (Figure 4-figure supplement 1 A and C, rightmost plot), and we defined a contrast index (CI) by taking the 0.9 quantile of this distribution (dashed line; the 0.9 quantile was chosen, instead of simply using the maximum local-contrast value, to make this measure more noise-robust). This CI metric was then used to quantify the amount of signal contained in the STA image. A STA-based RF was considered as well defined if, among the sequence of STA images corresponding to each pre-spike-generation time lag (see examples in Figure 1), the one with the highest contrast had CI > 5.5 (usually, such highest-contrast image was the third in the sequence, corresponding to 66-99 ms before the spiking event; see examples in Figure 4A). It should be noted that, since the intensity values of the original STA images (Figure 4-figure supplement 1 A and C, leftmost plot) were expressed as z-scores (see above), the 5.5 threshold can be interpreted in terms of peak-to-peak (i.e. white-to-black) distance in sigma units of the z-scored STA values.

To further analyze the spatial structure of each well-defined, STA-based RF we automatically counted how many distinct lobes the highest-contrast STA image obtained for a unit contained. To this aim, the selected image was background-suppressed with the Perona-Malik algorithm (Perona and Malik, 1990) to enhance the high-intensity regions within the RF and reduce noise-driven false lobe detections. We then took the modulus of the resulting image and we set a threshold over its z-score values (ranging from 3 to 6 units of standard deviations; Figure 4-figure supplement 1 B and D). Based on such binarized image, we computed: i) the centroid position of each simply connected region (blue dots); and ii) the area-weighted center of mass (CM) of the image (green dot). Finally, we defined a "region of acceptability" for the centroids of the simply connected regions to be included in the lobe count. This region was a circle, centered on CM, with a radius that was twice as large as the diameter of the largest simply connected region. Setting this acceptability circle allowed excluding spurious, simply connected (often very small) regions that were far away from the CM of the STA image (yellow dots in Figure 4-figure supplement 1 B and D), especially in the case of poorly defined RFs. The total number of accepted centroids at the end of this selection procedure is the lobe count reported in Figure 4D for the different binarization thresholds.

HMAX simulations. The structure of the HMAX model of the ventral stream has been already described in the Results (Figure 2) and motivated in the Discussion. For our application, we have chosen the version of the model described in (Serre et al., 2007b) and downloaded from http://maxlab.neuro.georgetown.edu/docs/hmax/hmaxMatlab.tar. Briefly, the model is a feedforward neural network with alternating layers of units performing: 1) either a max-pooling (invariance-building) operation over input units with the same feature-selectivity, but RFs having different positions/scales (C1 and C2 units; dashed lines in Figure 2); or 2) a template-matching (selectivity-building) operation over input units with different feature selectivity (S1 and S2 units; solid lines). As pointed out in the Discussion, the feature representations built by the model in the S-type layers are not learned, and there is no attempt to maximize classification accuracy in a given discrimination task. The S1 layer is simply a bank of Gabor filters with various orientations, spatial frequencies, positions and scales (similar to V1 simple cells), while the S2 units are tuned to random patches of images taken from various databases (e.g., Caltech 101), with different units having as a template the same image patch, but at different positions and scales. Such static, hardwired architecture makes the model very suitable to isolate the role of the max-pooling and template-matching

computations in determining the tuning for oriented gratings along a ventral-like processing hierarchy.

To this aim, we fed as an input to the network (i.e., to the S1 layer) a set of drifting gratings spanning the same range of directions used in the neurophysiology experiment. Each drifting grating consisted of a sequence of progressively phase-shifted frames, and the activations of the units of the network in response to such frames were computed to simulate the dynamics of the neuronal responses during stimulus presentation. For each sampled unit, its preferred stimulus direction was selected and the power spectrum of its time-dependent (i.e., frame-dependent) response was computed to estimate the modulation index (MI), thus quantifying its phase sensitivity. We also integrated the activation of each unit over the whole stimulus presentation to compute its overall mean response for a given grating direction. By iterating this procedure for each stimulus direction, we built orientation and direction tuning curves and computed OSI and BSI values. All these indexes were computed for a randomly selected subset of units (n = 250) in each layer. The resulting distributions are those shown in Figures 3C, 5C and 6C.

Statistical analysis. Differences between median values of two distributions were quantified using a Mann-Whitney U-test. Differences in the faction of units above or below a certain threshold index value (for MI, OSI, BSI or CI) were quantified using the χ^2 test for homogeneity with one degree of freedom. When needed (i.e., in case of small sample size), Fisher exact test was used to compare two distributions (e.g., the lobe count distributions in Figure 4D). Finally, in the meta-analysis shown in Figure 7, we compared the mean fractions of modulated (Figure 7A) and orientation-tuned (Figure 7B) units across primate visual cortical areas V1, V4 and IT using a one-way ANOVA.

Acknowledgments. We thank Margherita Riggi for performing the histological procedures.

Competing interests. None of the authors has financial and non-financial competing interests of any nature.

References

Adesnik H, Bruns W, Taniguchi H, Huang ZJ, Scanziani M (2012) A neural circuit for spatial summation in visual cortex. Nature 490:226–231.

Aggleton JP, Keen S, Warburton EC, Bussey TJ (1997) Extensive cytotoxic lesions involving both the rhinal cortices and area TE impair recognition but spare spatial alternation in the rat. Brain Res Bull 43:279–87.

Akerboom J et al. (2012) Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. J Neurosci 32:13819–13840.

Alemi-Neissi A, Rosselli FB, Zoccolan D (2013) Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition. J Neurosci 33:5939–5956.

Alwis DS, Richards KL, Price NSC (2016) Masking reduces orientation selectivity in rat visual cortex. J Neurophysiol 116:2331–2341.

Andermann ML, Kerlin AM, Roumis DK, Glickfeld LL, Reid RC (2011) Functional Specialization of Mouse Higher Visual Cortical Areas. Neuron 72:1025–1039.

Aronov D, Reich DS, Mechler F, Victor JD (2003) Neural Coding of Spatial Phase in V1 of the Macaque Monkey. J Neurophysiol 89:3304–3327.

Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. J Neurophysiol 98:1733–50.

Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, Solomon EA, Majaj NJ, DiCarlo JJ (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLOS Comput Biol 10:e1003963.

Cox PH, Riesenhuber M (2015) There Is a “U” in Clutter: Evidence for Robust Sparse Codes Underlying Clutter Tolerance in Human Vision. J Neurosci 35:14148–14159.

David SV, Hayden BY, Gallant JL (2006) Spectral Receptive Field Properties Explain Shape Selectivity in Area V4. J Neurophysiol 96:3492–3505.

De Valois RL, Albrecht DG, Thorell LG (1982a) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res 22:545–559.

De Valois RL, William Yund E, Hepler N (1982b) The orientation and direction selectivity of cells in macaque visual cortex. Vision Res 22:531–544.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–62.

Desimone R, Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J Neurophysiol 57:835–868.

DiCarlo JJ, Zoccolan D, Rust NC (2012) How Does the Brain Solve Visual Object Recognition? Neuron 73:415–434.

Djurdjevic V, Ansuini A, Bertolini D, Macke JH, Zoccolan D (2018) Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy. Curr Biol 28:1005-1015.e5.

Durand S, Iyer R, Mizuseki K, Vries S de, Mihalas S, Reid RC (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.

Foster K H, Gaska J P, Nagler M, Pollen D A (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. J Physiol 365:331–363.

Fournier J, Monier C, Levy M, Marre O, Sári K, Kisvárday ZF, Frégnac Y (2014) Hidden Complexity of Synaptic Receptive Fields in Cat V1. J Neurosci 34:5515–5528.

Fournier J, Monier C, Pananceau M, Frégnac Y (2011) Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. Nat Neurosci 14:1053–1060.

Froudarakis E, Berens P, Ecker AS, Cotton RJ, Sinz FH, Yatsenko D, Saggau P, Bethge M, Tolias AS (2014) Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. Nat Neurosci 17:851–857.

Gallant JL, Braun J, Essen DV (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science 259:100–103.

Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC (1996) Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. J Neurophysiol 76:2718–2739.

Gallardo L, Mottles M, Vera L, Carrasco MA, Torrealba F, Montero VM, Pinto-Hamuy T (1979) Failure by rats to learn a visual conditional discrimination after lateral peristriate cortical lesions. Physiol Psychol 7:173–177.

Gao E, DeAngelis GC, Burkhalter A (2010) Parallel Input Channels to Mouse Primary Visual Cortex. J Neurosci 30:5912–5926.

Girman SV, Sauvé Y, Lund RD (1999) Receptive Field Properties of Single Neurons in Rat Primary Visual Cortex. J Neurophysiol 82:301–311.

Glickfeld LL, Andermann ML, Bonin V, Reid RC (2013) Cortico-cortical projections in mouse visual cortex are functionally target specific. Nat Neurosci 16:219–226.

Gur M, Kagan I, Snodderly DM (2005) Orientation and Direction Selectivity of Neurons in V1 of Alert Monkeys: Functional Relationships and Laminar Distributions. Cereb Cortex 15:1207–1221.

Hawrylycz M et al. (2016) Inferring cortical function in the mouse visual system through large-scale systems neuroscience. Proc Natl Acad Sci 113:7337–7344.

Hong H, Yamins DLK, Majaj NJ, DiCarlo JJ (2016) Explicit information for category-orthogonal object properties increases along the ventral stream. Nat Neurosci 19:613–622.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215–43.

Ji W, Gămănuţ R, Bista P, D'Souza RD, Wang Q, Burkhalter A (2015) Modularity in the Organization of Mouse Primary Visual Cortex. Neuron 87:632–643.

Juavinett AL, Callaway EM (2015) Pattern and Component Motion Responses in Mouse Visual Cortical Areas. Curr Biol 25:1759–1764.

Kagan I, Gur M, Snodderly DM (2002) Spatial Organization of Receptive Fields of V1 Neurons of Alert Monkeys: Comparison With Responses to Gratings. J Neurophysiol 88:2557–2574.

Khaligh-Razavi S-M, Kriegeskorte N (2014) Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLOS Comput Biol 10:e1003915.

Kheradpisheh SR, Ghodrati M, Ganjtabesh M, Masquelier T (2016) Deep Networks Can Resemble Human Feed-forward Vision in Invariant Object Recognition. Sci Rep 6:32672.

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

Leibo JZ, Liao Q, Anselmi F, Freiwald WA, Poggio T (2017) View-Tolerant Face Recognition and Hebbian Learning Imply Mirror-Symmetric Neural Tuning to Head Orientation. Curr Biol 27:62–67.

Li N, Cox DD, Zoccolan D, DiCarlo JJ (2009) What Response Properties Do Individual Neurons Need to Underlie Position and Clutter "Invariant" Object Recognition? J Neurophysiol 102:360–376.

Marshel JH, Garrett ME, Nauhaus I, Callaway EM (2011) Functional Specialization of Seven Mouse Visual Cortical Areas. Neuron 72:1040–1054.

Matsui T, Ohki K (2013) Target dependence of orientation and direction selectivity of corticocortical projection neurons in the mouse V1. Front Neural Circuits 7 Available at: https://www.frontiersin.org/articles/10.3389/fncir.2013.00143/full [Accessed January 26, 2018].

Maunsell JHR, Newsome and WT (1987) Visual Processing in Monkey Extrastriate Cortex. Annu Rev Neurosci 10:363–401.

McDaniel WF, Coleman J, Lindsay Jr. JF (1982) A comparison of lateral peristriate and striate neocortical ablations in the rat. Behav Brain Res 6:249–272.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283:79–99.

Niell CM (2011) Exploring the Next Frontier of Mouse Vision. Neuron 72:889–892.

Niell CM, Stryker MP (2008) Highly selective receptive fields in mouse visual cortex. J Neurosci 28:7520–36.

Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. Neuron 65:472–9.

O'Keefe LP, Levitt JB, Kiper DC, Shapley RM, Movshon JA (1998) Functional Organization of Owl Monkey Lateral Geniculate Nucleus and Visual Cortex. J Neurophysiol 80:594–609.

Orban GA (2008) Higher Order Visual Processing in Macaque Extrastriate Cortex. Physiol Rev 88:59.

Pagan M, Urban LS, Wohl MP, Rust NC (2013) Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. Nat Neurosci 16:1132–1139.

Pasupathy A, Connor CE (1999) Responses to Contour Features in Macaque Area V4. J Neurophysiol 82:2490–2502.

Pasupathy A, Connor CE (2002) Population coding of shape in area V4. Nat Neurosci 5:1332–1338.

Pecka M, Han Y, Sader E, Mrsic-Flogel TD (2014) Experience-Dependent Specialization of Receptive Field Surround for Selective Coding of Natural Scenes. Neuron 84:457–469.

Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12:629–639.

Pollen DA, Nagler M, Daugman J, Kronauer R, Cavanagh P (1984) Use of gabor elementary functions to probe receptive field substructure of posterior inferotemporal neurons in the owl monkey. Vision Res 24:233–241.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–25.

Ringach DL, Shapley RM, Hawken MJ (2002) Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. J Neurosci 22:5639–5651.

Rossant C, Kadir SN, Goodman DFM, Schulman J, Hunter MLD, Saleem AB, Grosmark A, Belluscio M, Denfield GH, Ecker AS, Tolias AS, Solomon S, Buzsáki G, Carandini M, Harris KD (2016) Spike sorting for large, dense electrode arrays. Nat Neurosci 19:634–641.

Rosselli FB, Alemi A, Ansuini A, Zoccolan D (2015) Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. Front Neural Circuits 9:10.

Roth MM, Helmchen F, Kampa BM (2012) Distinct Functional Properties of Primary and Posteromedial Visual Area of Mouse Neocortex. J Neurosci 32:9716–9726.

Rust NC, DiCarlo JJ (2010) Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT. J Neurosci 30:12978–12995.

Sánchez RF, Montero VM, Espinoza SG, Díaz E, Canitrot M, Pinto-Hamuy T (1997) Visuospatial Discrimination Deficit in Rats after Ibotenate Lesions in Anteromedial Visual Cortex. Physiol Behav 62:989–994.

Sary G, Vogels R, Kovacs G, Orban GA (1995) Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. J Neurophysiol 73:1341–1354.

Schiller PH, Finlay BL, Volman SF (1976) Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. J Neurophysiol 39:1288–1319.

Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. J Vis 6 Available at: http://www.journalofvision.org/content/6/4/13 [Accessed May 22, 2013].

Sedigh-Sarvestani M, Fernandez-Lamo I, Jaegle A, Taylor MM (2014) Second Order Receptive Field Properties of Simple and Complex Cells Support a New Standard Model of Thalamocortical Circuitry in V1. J Neurosci 34:11177–11179.

Self MW, Lorteije JAM, Vangeneugden J, Beest EH van, Grigore ME, Levelt CN, Heimel JA, Roelfsema PR (2014) Orientation-Tuned Surround Suppression in Mouse Visual Cortex. J Neurosci 34:9290–9304.

Serre T, Oliva A, Poggio T (2007a) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U A Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17404214.

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007b) Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 29:411–26.

Sharpee TO (2013) Computational Identification of Receptive Fields. Annu Rev Neurosci 36:103–120.

Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. Vision Res 31:1078–1086.

Smith IT, Townsend LB, Huh R, Zhu H, Smith SL (2017) Stream-dependent development of higher visual cortical areas. Nat Neurosci 20:200.

Tafazoli S, Safaai H, Franceschi GD, Rosselli FB, Vanzella W, Riggi M, Buffolo F, Panzeri S, Zoccolan D (2017) Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. eLife 6:e22794.

Tees RC (1999) The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. Behav Brain Res 106:55–73.

Tohmi M, Meguro R, Tsukano H, Hishida R, Shibuki K (2014) The Extrageniculate Visual Pathway Generates Distinct Response Properties in the Higher Visual Areas of Mice. Curr Biol 24:587–597.

Vaiceliunaite A, Erisken S, Franzen F, Katzner S, Busse L (2013) Spatial integration in mouse primary visual cortex. J Neurophysiol 110:964–972.

Van den Bergh G, Zhang B, Arckens L, Chino YM (2010) Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. J Comp Neurol 518:2051–2070.

Vermaercke B, Gerich FJ, Ytebrouck E, Arckens L, Beeck HPO de, Bergh GV den (2014) Functional specialization in rat occipital and temporal visual cortex. J Neurophysiol:jn.00737.2013.

Vinken K, Van den Bergh G, Vermaercke B, Op de Beeck HP (2016) Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. Cereb Cortex N Y NY 26:3310–3322.

Vinken K, Vogels R, Op de Beeck H (2017) Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. Curr Biol 27:914–919.

Vogels R, Orban GA (1991) Quantitative study of striate single unit responses in monkeys performing an orientation discrimination task. Exp Brain Res 84:1–11.

Vogels R, Orban GA (1993) Responses of monkey infero-temporal units in an orientation discrimination task In T. P. Hicks, S. Molotchnikoff, & T. Ono, eds. Progress in Brain Research Elsevier, p. 427–444. Available at: http://www.sciencedirect.com/science/article/pii/S0079612308603866 [Accessed May 23, 2018].

Vogels R, Orban GA (1994) Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. J Neurophysiol 71:1428–1451.

Wang Q, Burkhalter A (2007) Area map of mouse visual cortex. J Comp Neurol 502:339–357.

Wang Q, Gao E, Burkhalter A (2011a) Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. J Neurosci 31:1905–1918.

Wang Q, Gao E, Burkhalter A (2011b) Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. J Neurosci 31:1905–1918.

Wang Q, Sporns O, Burkhalter A (2012a) Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. J Neurosci 32:4386–4399.

Wang Q, Sporns O, Burkhalter A (2012b) Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. J Neurosci 32:4386–4399.

Wörtwein G, Mogensen J, Williams G, Carlos J, Divac I (1993) Cortical area in the rat that mediates visual pattern discrimination. Acta Neurobiol Exp (Warsz) 54:365–376.

Wypych M, Wang C, Nagy A, Benedek G, Dreher B, Waleszczyk WJ (2012) Standardized F1 – A consistent measure of strength of modulation of visual responses to sine-wave drifting gratings. Vision Res 72:14–33.

Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci 111:8619–8624.

Zhou H, Friedman HS, Heydt R von der (2000) Coding of Border Ownership in Monkey Visual Cortex. J Neurosci 20:6594–6611.

Zoccolan D, Kouh M, Poggio T, Dicarlo J (2007) Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. J Neurosci 27:12292–307.