SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Neuroscience Area – PhD course in

Cognitive Neuroscience

Multidisciplinary investigation of shape and motion processing in the rat visual cortex

Candidate:

Giulio Matteucci

Advisor:

Davide Zoccolan

Academic Year 2019-20

Acknowledgements

First, I want to express all my gratitude to my supervisor, Prof. Davide Zoccolan. Even before the start of my PhD, during my master studies, he has always been a guide and a reference point towards my dream of becoming a neuroscientist. Then, throughout the 5 years I spent in the lab, his constructive critiques, friendly encouragements and thoughtful discussions had a huge impact on my development as a researcher.

I want to say thanks to all my lab mates, present and past, that shared the of hopes and disappointments, joys and sorrows entailed by this 5-years-long scientific and personal journey. Together with Davide they became a sort of second family.

A special thanks goes to Rosilari for having been an invaluable mentor, co-worker and friend, sharing so many morning cafellatte with me. Half of the work described in this thesis started with her. Another big one goes to Benedetta, who will always be my first master student and a true friend. Her enthusiasm and hard-work have been essential in pushing forward the behavioral project described in this thesis. Another special thanks goes to Anna whose thought-provoking discussion and deep scientific passion boosted my motivation at the end of this journey and also gifted me with some of the most beautiful moments I lived here in Trieste.

Thanks to Margherita who is the true soul of the lab. To Eis for saving my day so many times thanks to his amazing technical skills. Riccardo for adding some laughter to the heaviest days. Thanks to Natalia for her friendship and for sharing with me her incredibly extensive lab experience. Thanks to Alessio for having introduced me to the art of neural data analysis. And thanks to Federica, my first mentor, for having generously taught me her hard-won surgical knowledge.

Thanks to Giulia that accompanied me along a big chunk of my PhD student life.

Finally, thanks to my parents, who always nurtured my passion for science, always believed in me, even when I didn’t, and supported me in a way that no words could tell. Without them, me and this thesis wouldn’t be here.

Table of contents

Introduction

Extracting information useful to guide behavior from the raw retinal input is what the visual system was shaped for through the scope of evolution. In the complex ecological niches occupied by mammals, such information can clearly be quite abstract in nature: is there a predator in view? Or is it a conspecific? Where is it going? Is that food in reach?

The answers to these biologically relevant questions are clearly very far apart from the low-level description of the momentary visual input provided by the pattern of photoreceptor activations in the retina of an animal.

As the historical struggle of computer vision to succeed in confronting these problems demonstrates, an impressive amount of nontrivial computations is needed to infer answers to those questions through the process that we may call "high-level" vision[1,2] (a term commonly used to stress the distance from the local-luminance-based descriptions of the input). Nonetheless, modern computer vision systems based on deep learning[3] are starting to achieve and, in some cases surpass, human level performance in some well-circumscribed high-level vision tasks: from object recognition[4] to face verification[5] or image captioning[6]. These recent technological breakthroughs are renewing inside the broader scientific community the interest in understanding how biological brains manage to solve these same problems. The present moment is, indeed, particularly ripe for the study of high-level vision: for the first time we have access to computational models managing to replicate (at least to some extent) the sensory and cognitive functions we are interested in. Those artificial systems are starting to provide something akin to the "ideal observer models" of classical psychophysics, extending their reach to a wider range of interesting tasks (albeit relaxing any claim of full optimality). The huge potential for cross-fertilization existing between neuroscience and machine learning implied by the above-mentioned considerations make the experimental testing of computational theories about biological high-level sensory processing more relevant than ever.

On the other hand, another revolution − this time of experimental nature − is currently transforming neuroscience. The development of optogenetics[7] (to reversibly excite or inhibit neurons in vivo with sub millisecond precision), two-photon imaging with genetically encoded calcium indicators[8] (to monitor large neural population across different brain areas) and transgenic animal lines expressing genetically encoded sensors or effectors with cell-type specificity[9] are providing unprecedented ways to precisely monitor and perturb information flow both at the microcircuit level and across areas in awake behaving animals. Crucially, this expanded neuroscientific toolset is currently available in rodents and not readily applicable to primates.

To tackle the challenge of understanding how the brain performs high-level vision it will be crucial to leverage on both the computational and the experimental advances described above. In order to do so, exploiting the superior experimental accessibility offered by rodent animal models will be of paramount importance[10]. Rats, in particular, have been found to rapidly learn and reliably perform many complex visual tasks, emerging as the most promising candidates for this role[11,12]. Until very recently however, primates have been the dominant animal model for high-level vision research. Such sustained scientific attention

focused on these animals trough decades led to a thorough anatomical and physiological characterization of the visual system of these species.

The recurring theme of my PhD work is the comparison between rat and monkey visual cortex coding properties in the attempt to abstract some general principles of high-level visual computation. More specifically, my work has been subdivided in 5 distinct subprojects. Each one will be described in a specific chapter of the thesis:

I --- "***Nonlinear processing of shape information in rat lateral extrastriate cortex***". In this chapter, the progression of neuronal tuning properties across the putative rat ventral stream homologue (V1→LM→LI→LL) is characterized. The result of such characterization is then compared with the monkey literature and with what observed "in silico" (both in a deep neural network and in a classical hierarchical object recognition model) providing new support to the ventral homology hypothesis. The paper describing this work has been published in Journal of Neuroscience.

II --- "***A template-matching algorithm for laminar identification of cortical recording sites from evoked response potentials***". In this chapter, the development and validation of an automated method to identify the laminar position of recording sites of a multichannel silicon probe is described. The proposed method is based on matching the waveform of recorded local field potentials (LFPs) in response to visual stimulation (VEP) to a reference template (deduced from a dataset of histologically annotated laminar VEP recordings). The paper describing this work is available on bioRxiv.

III --- "***Causal adaptation to visual input dynamics governs the development of complex cells in V1***". In this chapter, a causal test of the so called "unsupervised temporal learning" (UTL) hypothesis is provided. Such computational principle postulates a critical role for the temporal continuity of the visual-input experienced during early postnatal development in building-up position tolerance of V1 complex cells. The paper describing this work is available on bioRxiv.

IV --- "***Behavioral evidence of spontaneous motion integration in rats***". In this chapter, rats are shown to be able to spontaneously perceive global motion of plaids (i.e. stimuli composed of multiple oriented elements). To this aim, a motion direction discrimination priming task was designed. This behavioral paradigm enables investigating the spontaneous (i.e. non-rewarded) representation of grating and plaid stimuli. This, in turn, allows to infer whether gratings and plaids are represented by the same neural populations and whether the organizing principle of the representation is the local or the global direction of the stimulus.

V --- "*Linear receptive field structure does not account for pattern motion responses in rat visual cortex*". In this chapter, the neural substrate of motion integration in rat visual cortex is investigated. Neurons recorded in V1, LM and RL areas are classified as global (i.e. "pattern") or local (i.e. "component") direction selective cells and their linear receptive fields (RFs) are reconstructed. This enables testing whether observed global-motion sensitive

responses observed throughout rodent visual cortex are predictable on the basis of their linear RFs or whether they originate from some (less trivial) nonlinear integration mechanism.

Introduction		
Subproject I	*"Nonlinear processing of shape information in rat lateral extrastriate cortex"*	shape processing
Subproject II	*"A template-matching algorithm for laminar identification of cortical recording sites from evoked response potentials"*	laminar identification
Subproject III	*"Causal adaptation to visual input dynamics governs the development of complex cells in V1"*	learning of invariance
Subproject IV	*"Behavioral evidence of motion integration in rats"*	motion processing
Subproject V	*"Linear receptive field structure does not account for pattern motion responses in rat visual cortex"*	
Conclusion		

Table 1. Layout of the thesis displaying titles and topic of each subproject.

In the next introductory sections, we will briefly review the structure and function of monkey "ventral" and "dorsal" streams for shape and motion processing. This is, in fact, the necessary background to frame the comparison between primates and rats. Next, we will introduce the importance of studying postnatal developmental processes to attain a satisfactory understanding of sensory systems, by uncovering the learning principles underlying their function. Lastly, we will briefly review the functional and anatomical organization of the rodent visual system.

Ventral and Dorsal streams

The first lines of evidence that primate visual cortex contains two distinguishable processing streams was based on contrasting the effects of inferior temporal (IT) and posterior parietal cortex (PPC) lesions in monkeys. IT lesions caused severe deficits in visual discrimination tasks, without affecting animal performance in visuospatial tasks (such as visually guided reaching or judging which of two objects lies closer to a visual landmark). On the other hand, PPC lesions spared visual discrimination capabilities while causing instead severe deficits in visuospatial performance. To explain such findings, Ungerleider and Mishkin[13] proposed the existence of two segregated streams of processing (see Fig. 1).

The first of those streams is the "occipitotemporal" or "ventral" one, traveling into the temporal lobe from V1 and V2 to V4 and finally reaching inferotemporal (IT) cortex, which is thought to mediate the visual recognition of objects (i.e. encoding information about "what" an object is). The second stream is the "occipitoparietal" or "dorsal" one, starting from V1 to reach MT (middle-temporal), MST (medial superior temporal) as well as LIP (lateral intraparietal) and VIP (ventral intraparietal) cortex, which is thought to mediate the

perception of spatial relationships among objects as well as visual guidance towards them (i.e. encoding information about "where" an object is).

Through almost three decades of research, thanks to the support of many functional studies carried out with different techniques (from single unit electrophysiology to fMRI and psychophysics), such ideas took hold as the dominant theory of primate visual system organization[14]. The consolidation of this research field was accompanied by a revision of the role of the dorsal pathway towards supporting motor control rather than spatial representations per se[15,16].

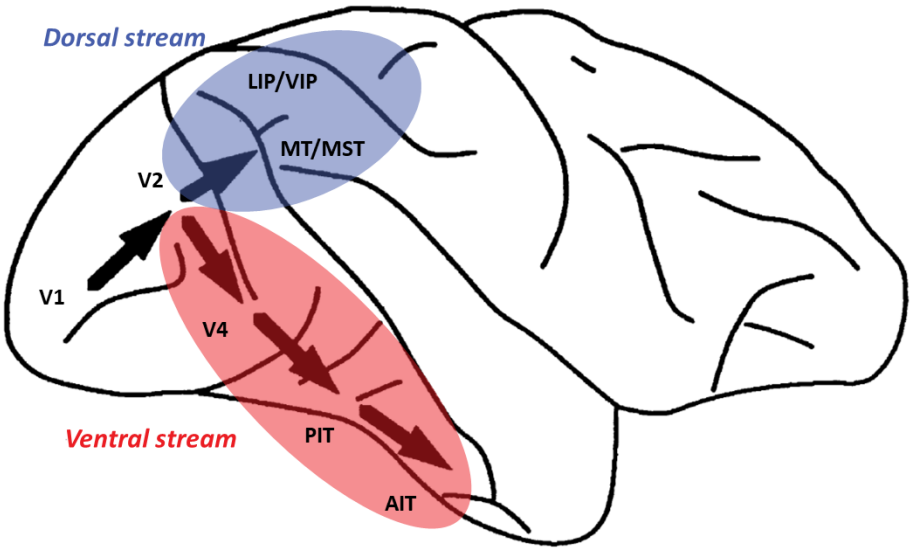

Figure 1. Ventral and Dorsal stream in the monkey brain. Both streams start in V1, in the occipital lobe. The dorsal stream (in blue) continues trough the parietal lobe reaching middle temporal (MT), middle superior temporal (MST), lateral intraparietal (LIP) and ventral intraparietal (VIP) area. The ventral stream (in red) travels into the temporal lobe through V4 to reach first posterior- and later anterior-inferotemporal cortex (PIT and AIT). Adapted from Goodale et al.[16].

More recently, however, researchers started to draw a more nuanced picture of the functional specialization of monkey extrastriate areas. In the dorsal stream a finer subdivision between a purely visual motion processing dorsolateral pathway (including MT and MST) and a dorsomedial visuomotor one (including LIP and VIP) has been proposed. The first pathway would be devoted to optic flow analyses (e.g. to discriminate object from self-motion or recognize specific motion patterns), while the second one would be involved in monitoring the spatial location of objects to guide actions[17] (e.g. reaching and grasping). The view of the ventral stream became more nuanced too with the realization that spatial properties of visual objects, including their position, size, and pose, can be reliably decoded from ventral cortex[18].

Recent years also saw the research community cautioning against an exaggeratedly rigid distinction among visual streams[19] (and one to one mapping areas and function/stream assignment). For instance, the existence of direct connections from MT/MST to V4/IT is now known to play an important role in underlying the extraction of shape information from motion and depth cues[20,21], effectively violating a rigid ventral vs. dorsal segregation[22]. Consistently, a view of ventral and dorsal areas as parts of a larger network of areas devoted to build physical object representations as building blocks of causal generative models of the

environment (i.e. models of how objects move, interact among them and with the agent: something key for planning complex behavior) has been proposed[23].

Despite all those caveats the ventral vs. dorsal distinction remains one of the most important conceptualizations in cognitive neuroscience. This is why many efforts are being devoted to map that subdivision to the rodent brain in order to improve our understanding of extrastriate areas of these animals[24].

Shape processing

Arguably, the most widely studied high-level ventral visual function is object recognition. Perceiving the shape of objects we encounter and recognizing their identity is usually such an effortless and automatic process for us that we may be intuitively tempted to underestimate its computational difficulty. Each object we encounter throughout the course of our life will produce, in every different moment, an almost unique projection on our retina due to differences in the many nuisance variables influencing the specific pattern of light projected on our photoreceptor arrays. Crucially, changing nuisance variables such as object pose, position, size, illumination, presence of clutter, etc… does not change the identity of the object itself, while heavily affecting its retinal representation. Becoming tolerant to those identity-preserving transformations of the appearance of objects (i.e. changes of the nuisance variables) is the very hard challenge that our visual system has to solve to achieve accurate and robust visual object recognition[25].

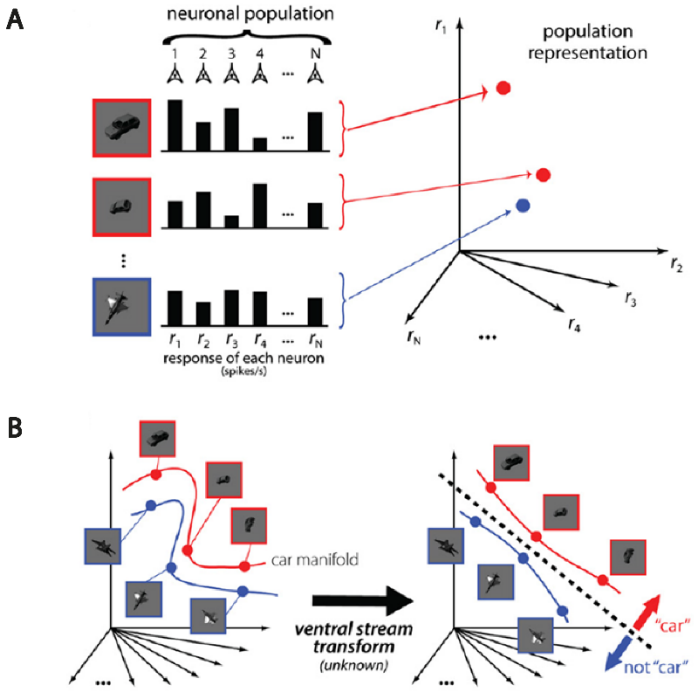

Figure 2. Representational untangling. A) The response of a population of N neurons to a given visual stimulus can be conceptualized as a point in the high-dimensional space spanned by the activities of each neuron in the population (i.e. where each axis represents the response firing rate r_i of each neuron). Neural representations of three different images are shown: two views of a car (squared in red) and a view of a plane (squared in blue). B) All possible identity-preserving transformations of a given object will lie on a lower-dimensional subspace called "object manifold". Manifolds of different objects will be highly curved and tangled, as shown in figure. Car (red) and plane (blue) manifolds are shown in figure as curves for simplicity. The computational role of ventral processing is to successively transform object representations, smoothing manifolds up to the point when they become separable by a hyperplane (represented in figure by a dashed line). Reproduced from DiCarlo et al.[26].

A solution to such computational challenge, also known as the "invariance problem", has been shown to exist in monkey IT. In fact, neuronal representations in this area (probed via extracellular population recordings) are able to support transformation tolerant linear decoding of object identity (thus making computationally trivial to perform invariant object recognition on them). This property of IT representation is thought to be achieved through a progressive reformatting of visual representations throughout the ventral stream with a progressive untangling of the manifolds spanned by different objects in the neural population space as they undergo identity preserving transformations[26] (see Fig. 2). The core of this process is thought to be implemented, in the primate brain, via a cascade of largely feedforward canonical computations (but see DiCarlo recent work on the role of recurrency[27] in object recognition).

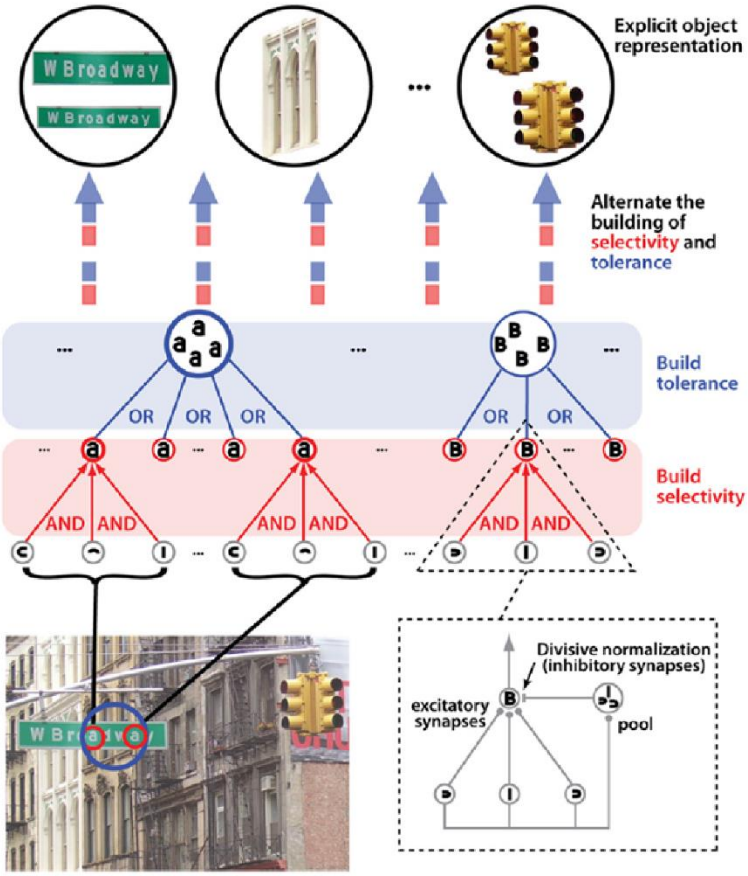

Figure 3. Hierarchical build-up of selectivity and invariance. Biologically-inspired models of object recognition such as HMAX and modern deep neuronal networks aim at achieving a gradual untangling of object manifolds by stacking layers of neuronal units in a feedforward hierarchy. The example depicted in this figure shows an alternation of layers performing AND-like (e.g. filtering, in red) and OR-like (e.g. max-pooling, blue) operations in a hierarchical stack. AND-like operations that build up tuning for new combinations of visual features are inspired by V1 simple cells. OR-like operations that build-up tolerance to changes in (e.g.) position and size by pooling over units of the previous layer (with identical feature selectivity but slightly different center locations and sizes) are inspired by V1 complex cells. Reproduced from DiCarlo et al.[26].

More specifically, the main hypothesis about how representations of objects become more and more explicit (i.e. object identity gets more and more linearly decodable) as they get transformed along the ventral stream, is through a hierarchical stacking of AND-like selectivity-building and OR-like tolerance-building operations. Such idea has been inspired

by the seminal work by Hubel and Wiesel on simple and complex cells[28]. The transfer function of those two functional cell types, in fact, have been conceptualized as computing a logical AND (i.e. combination of simple feature detectors into more complex feature detectors to get new selectivity) or logical OR (i.e., max pooling to get new position tolerance) between their afferents (see Fig. 3).

Historically, influential models such as the Neocognitron[29] or HMAX[30] provided a first proof of the effectiveness of such architectural motif in explaining many experimentally observed phenomena in object perception. Later, these same ideas also led to the development of modern deep neural networks that for the first time successfully tackled the challenge of invariant object recognition in real world setting[3]. Their success corroborated even further the hypothesis that the principles outlined above may indeed represent the key computational building blocks of ventral computations.

The first subproject of my PhD work has been devoted to the search for new evidence that such principles are implemented in rat lateral extrastriate cortex, thereby strengthening the functional homology between such hierarchy of areas and the monkey ventral stream.

Motion processing

The landmark, purely-perceptual high-level dorsal function is arguably visual motion processing. One of the first steps entailed by it is the solution to what is known in the psychophysics literature as the "aperture problem". Such problem consists in the fact that local motion direction measurements are insufficient on their own to unambiguously specify the global direction of motion of a visual object causing them. In order to do so, at least two local directions have to be taken into account in a process known as "motion integration".

The need for such computation stems from the fact that when looking at the visual scene trough a small aperture (such as the small RF of a V1 neuron) we easily end up observing only straight edges moving through it (i.e. small details of the contours of the visual objects that are present in the scene). The critical consideration to understand the aperture problem is that, in such a situation, the visual pattern visible through the aperture is invariant to translation in directions parallel to the edge. This causes the parallel component of its motion to be undetectable, effectively enabling only the perception of the edge-perpendicular component of the object global velocity. Only combining the information from at least two aperture-limited samples of local motion (of different edges) it is possible to determine the object global velocity causing them (ideally solving the system of two equations describing the projection of global velocity on the edge orthogonal vectors).

A neurally plausible implementation of such computation can be devised moving to the frequency space. The starting point for understanding this is the appreciation of the fact that the spatiotemporal Fourier spectrum of a moving edge has all its power concentrated in two blobs located symmetrically around the origin, and that the one of a rigid, non-rotating, object moving in a given direction is bound to lie on a plane in that same space. Thinking to the problem of building a global velocity sensitive detector in this space, the possibility of doing so by combining the output of many edge detectors to fully tile the object velocity plane becomes intuitive. This is exactly the intuition behind the influential model proposed

by Simoncelli & Heeger[31] to get global (i.e. 2D) motion detectors hierarchically combining local ones (i.e. 1D). This is illustrated in Fig. 4.

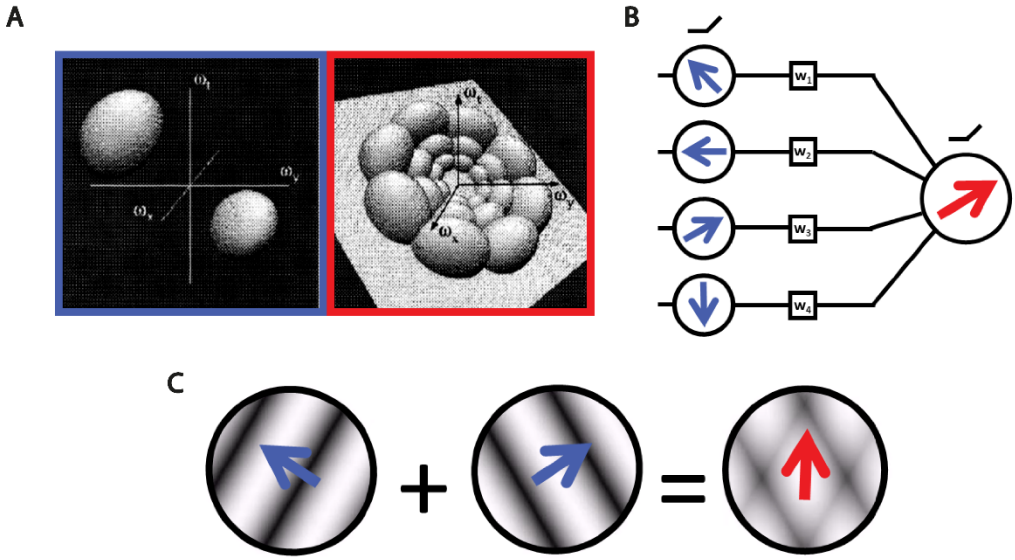

Figure 4. Construction of pattern cell selectivity. A) Left subpanel shows in the Fourier domain the selectivity (i.e. the filter) of one of the hypothesized V1-like input units (assumed as a directionally selective edge detector). As can be seen in the drawing, such feature selectivity corresponds, in Fourier space, to a pair of localized spatiotemporal frequency bands, symmetrically arranged about the origin. Right subpanel depicts in the same way selectivity of 12 V1-like afferents tuned for four orientations and three spatial scales, each consistent with a single common velocity. The plane depicted in the same subpanel identifies the region of the space where spectrum of a rigid, non-rotating object moving at this common velocity is bound to lie. Notice how selectivity of the 12 units effectively tile (part of) the plane. B) Simoncelli & Heeger model prescribes to sum the responses of those local direction selective neurons (blue arrows) using positive (i.e. excitatory) weights (boxed w_i) to yield a response selective for the global direction associated to the plane (red arrow). Not shown are a set of V1 neurons whose tuning bands lie outside the plane: these should be combined using negative (inhibitory) weights to increase the sharpness of the selectivity. Also not highlighted in figure is the fact that, to yield a physiologically realistic result the summation should be performed over V1 neurons with receptive fields distributed over a local spatial region. C) This panel depicts the construction of a plaid stimulus (with global direction specified by the red arrow) by combining two component gratings (drifting in the direction specified by the blue arrow). The plaid is the most frequently used stimulus to study pattern and component cells. Adapted from Simoncelli and Heeger[31].

Pioneering studies on cats and primates in the dorsal area MT discovered neurons capable of such kind of integration[32] (dubbed "pattern cells") and neurons signalling local direction only (dubbed "component cells"). As such kind of experimental investigations on the neural substrate of motion integration progressed through the years, Simoncelli & Heeger[31] model emerged as the most successful account of the computation underlying pattern selectivity[33,34].

Investigating behaviourally the ability of rats to integrate motion as well as the presence and mechanism of pattern and component cells in their visual cortex was the aim of the fourth and fifth subprojects, respectively, included in my PhD work.

Learning to see

Since the pioneering work of Attneave and Barlow[35,36], generations of neuroscientists were inspired by the idea that the tuning of sensory neurons is determined by an adaptation to the statistics of the input they need to encode. Such process should also happen in an

unsupervised manner, requiring only continuous, passive, exposure to the spatiotemporal regularities of natural sensory environment. Over the years, such intuition gained momentum, increasing its popularity and gaining strong support within the computational neuroscience field. A crucial turning point was reached with the now classic work by Olshausen and Field[37] on unsupervised learning based on maximizing the sparseness of the representation. The sparseness of a neural code can be intuitively defined in terms of coding density (i.e. the fraction of neurons that are strongly active at any one time): the lower the density, the higher the sparseness. Olshausen and Field demonstrated that adapting the selectivity of idealized units to produce the sparsest possible code for natural images (while maximally preserving the image information) makes their visual selectivity remarkably similar to that of V1 simple cells (see Fig. 5).

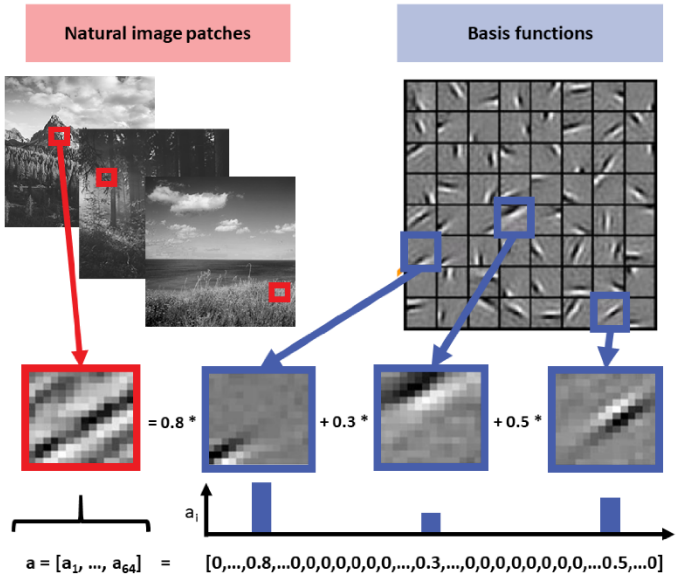

Figure 5. Sparseness. In searching for an efficient way to represent natural images (i.e. a way to represent them with sparse coefficients) one could try to directly optimize a basis set to this aim. This would mean searching, in the space of image patches (16x16 pixel in the example in figure), a suitable set of basis patches to represent natural ones in the sparsest possible way (without losing the content of the original image). In the top-left part of the figure three grayscale natural images provide a graphical impression of the natural image set used for a computational experiment of this kind (some patches taken from those natural images are highlighted in red). In the top-right part the basis set obtained as a result of the sparsity maximization procedure is shown. Notice their Gabor-like appearance similar to the shape of RFs experimentally observed for simple cells. An example of the reconstruction of a specific natural image patch (boxed in red) as a linear combination of the optimal basis patches (boxed in blue) is given in the bottom part of the figure. Notice how the vector of combination coefficients "a" representing the natural image patch in the optimal base is sparse, being full of zeros except for three of its elements corresponding to the three basis patches shown above. Adapted from Olshausen and Field[37].

Such a theoretical breakthrough prompted many efforts directed at showing that unsupervised learning can also account for the emergence of the other key ingredient of high-level vision: invariance. Most notably, another unsupervised learning principle known as slowness maximization[38] became the most prominent computational account of how visual neurons can learn invariance[39]. Intuitively, slowness quantifies how slowly in time the activation of a neuron changes in response to a time-varying input (see Fig. 6). Wiskott and Sejnowski[39] demonstrated that optimizing the input-output function of artificial units, to make their output change as slowly as possible in response to movies of transforming visual

objects, produces units that are able to support transformation tolerant decoding of object identity. This can be thought as a way to exploit the temporal structure of the input (i.e. the tendency of different object views to occur nearby in time) to factor out high-level visual attributes (such as object identity) from other faster-varying, lower-level ones (such as local luminance patterns).

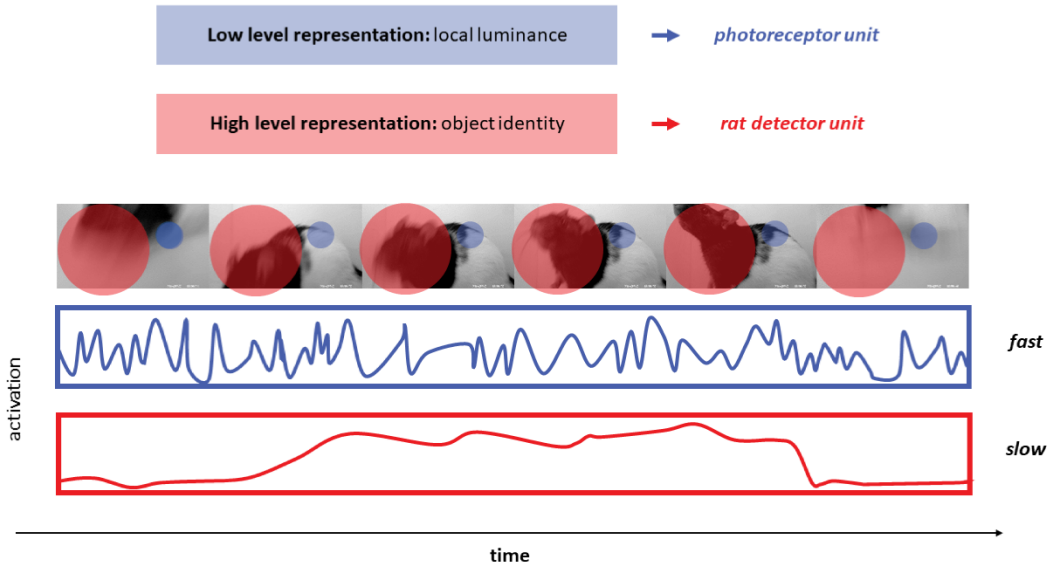

Figure 6. Slowness. The maximization of response slowness has been proposed as a candidate learning principle to explain the transformation tolerance of neurons in mammalian visual cortex. The intuitive relationship between slowness and tolerance is illustrated in this figure. Let's consider two idealized visual neurons: a "photoreceptor unit" (blue) simply detecting local luminance level in a small region of the visual space and a "rat detector unit" (red) detecting the presence of a rat in a larger region of the visual space. Those regions of the input space to which the two neurons are sensitive to (i.e. receptive fields - RF) are depicted as coloured shaded circles over the sequence of frames represented in the central subpanel (a footage of a rat moving around in an arena, first entering and then exiting from the field of view of the camera). The responses (i.e. activations) of the two neurons as the movie is played are depicted in the bottom subpanels (blue and red respectively). The photoreceptor responds to the fast luminance changes within its local RF with the jittering, fast-varying activity depicted in blue. On the contrary the rat detector activity change on the, much slower, timescale of the red trace following the presence of its preferred object (the rat) in the RF. This illustrates how the response activity of a high-level unit (such as the rat detector) activity varies on a longer timescale than a low level one (such as the photoreceptor).

On the experimental side, decades of research have documented the existence of so called "critical periods"[40,41] during the early postnatal life of mammals in which synaptic plasticity is boosted and the connectivity and function of sensory circuits are highly sensitive to manipulation of the input. Classically, this has been largely studied in the context of ocular dominance plasticity, with the monocular deprivation paradigm: closing an eye early in life makes it less effective in driving V1 cells whereas the nondeprived eye gains influence. As long as the cortex remains in a highly plastic state it is possible to reverse this condition by switching the identity of the deprived eye. However, with time, the effectiveness of such manipulations tends to decrease as the critical period closes. Decades of research on this topic demonstrated that the opening and closure of the critical period is controlled by the maturation of intracortical GABAergic inhibition[42,43] and modulated by neurotrophins[44] expression levels as well as by the onset of visual experience (see Fig. 7). Crucially, such early susceptibility to the structure of sensory environment is exactly what is needed to

implement the above-mentioned adaptations to the statistics of sensory input. Throughout the years, many researchers started targeting more subtle coding properties than ocular dominance (such as orientation and direction selectivity and surround-suppression). Such efforts convincingly demonstrated the susceptibility of a very wide array of stimulus coding properties to the early postnatal input[45–50].

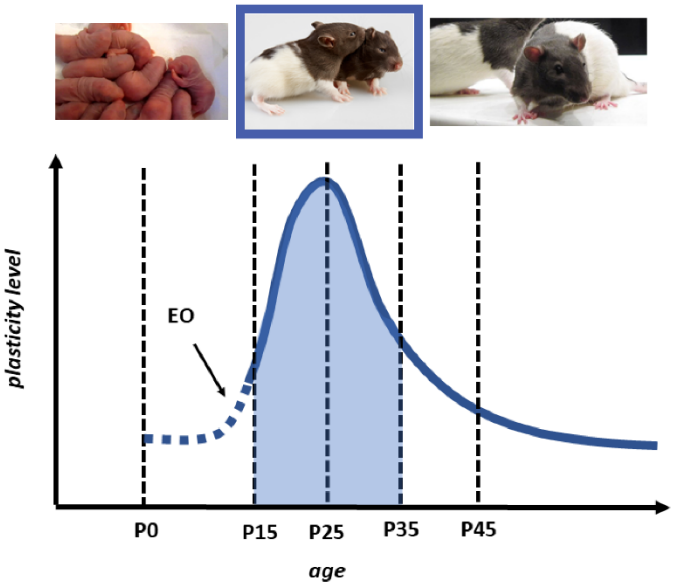

Figure 7. Plasticity during the critical period. One classical way of probing V1 plasticity level is monocular deprivation (MD): an experimental paradigm in which one eye of an animal gets closed for some days or weeks producing a shift of responsivity of neurons previously preferring that eye to the non-deprived one. The blue curve plotted here ideally represent V1 plasticity level as a function of animal age as assessed by sensitivity to MD. The timespan of high sensitivity (shaded region) that starts a few days after eye opening (EO) and closes a few weeks later is called "critical period". During such period of high plasticity visual cortical circuits are sculpted by the sensory inputs they receive. This is thought to enable adaptation to the statistical regularities of the sensory environment allowing for efficient coding of sensory information. Adapted from Berardi et al.[41].

Only recently, however, awareness is rising about the importance of studying experience-dependent development of brain function, not only as a topic of great scientific interest per se, but also as a key avenue to reach a better understanding of adult brain function. A convincing argument supporting such point of view has been laid out in a recent perspective paper by Richards et al.[51]. The authors argue that it may result easier to understand the few lines of code required to train and run a deep neural network capable of recognizing images rather than the resulting properties of its units. Tuning or connectivity properties of single units in those networks, in fact, often appear not to be easily interpretable. On the other hand, the kind of "understanding" of a neural network that is given in terms of learning objectives and rules (i.e. in terms of everything that is needed to specify how the network will self-organize during training, adapting to the specific training set that is provided to it) may be easier to attain than a mechanistic understanding based on single neuron properties. The success of this "learning principles approach" in the artificial neural network field is a strong encouragement to a greater focus on learning and development for the study of brain function.

Thanks to their relative ease of handling and their very fast lifecycle, rats are the ideal animal models for implementing a research program aiming at investigating the

developmental learning principles underlying adult sensory information processing. The third subproject of my PhD project was indeed aimed at exploiting the strengths of the rat model to take a first step in this direction. More specifically, to provide a causal test to the idea (inherent to the slowness maximization hypothesis) that the temporal structure of visual input experienced during the postnatal period instructs the development of V1 complex cells position tolerance.

The rat visual system

Rodents, way before becoming widespread in visual neuroscience research, have been the workhorse models of physiological and biomedical research for more than a century. One of the main reasons behind their popularity in biological research is that they are not so distant from us on the phylogenetic tree. Our common ancestor with rodents, in fact, dates back to 75 million years ago: 10 million years after the ancestor of other higher mammals (e.g., carnivores), and only 12 million years before the common ancestor of all primates.

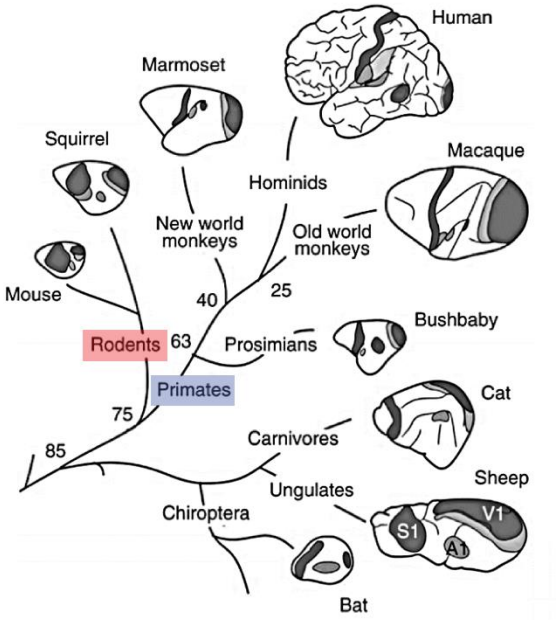

Figure 8. Phylogenetic tree representing mammalian brain evolution. Schematic of gross brain anatomy in eight mammalian species. Primary and secondary visual areas (V1 and V2), the primary auditory area (A1) and the primary somatosensory area (S1) are highlighted in shades of gray. Brain drawings are arranged along the mammalian phylogenetic tree: numbers at branching points indicate the age of last common ancestors, expressed in millions of years. Adapted from Carandini et al.[10].

This explains why, despite many evident morphological differences, rodents share a general common plan for brain development and organization with us, human primates[52]. Even if the rodent cortex is not folded and shows clear differences in its laminar structure with respect to the monkey one, its parcellation in functionally and anatomically distinct areas still shows a clear anatomical and functional homology with the primate one[53]. Obviously, however, it is not always easy (and likely, sometimes not possible) to find a precise one-to-one mapping between cortical areas of rodents and monkeys. This is particularly true for smaller and more specialized secondary sensory cortices, where structural and functional divergence may strongly reflect the different behavioral demands faced by each species in their own peculiar ecological niche. Despite this, attempting to

establish such correspondence can be a very productive scientific endeavour with the potential of highlighting the conserved core computational mechanisms underlying sensory processing while, at the same time, establishing a new advantageous model system for their study. In fact, besides their amenability to the ever-growing array of cutting-edge experimental techniques mentioned previously, rodents are much easier to house and handle with respect to non-human primates. Their reduced infrastructural burden translates into a critical scientific advantage since it allows experimenters to reach larger number of subjects getting to higher statistical powers in testing their hypotheses. Furthermore, bigger subject numbers coupled to their smaller physical dimensions fostered, in recent years, a proliferation of different rodent behavioral training paradigms and apparata pushing towards higher automation and a reduction of trainer manual intervention in the process. This is very promising for increasing data yield and reproducibility[54]. Moreover, as already mentioned, their very fast lifecycle (with a 21 days pregnancy duration and adulthood reached at postnatal day 60) makes rats perfect for developmental studies that would require years to be carried out in primates.

As crepuscular animals, rodents have often been presumed to have poor vision and weak visually driven behavior. It is this widespread belief that limited for years their adoption as visual neuroscience models. However, even if their spatial resolution is indeed ~50 times lower than that of primates[55], their natural behavior is frequently guided or modulated by vision. It has been now widely established that rodents strongly rely on vision to navigate their environment[56], anchoring their spatial representations to visual landmarks[57]. Other recent evidence shows how vision drives both defensive[58] and prey capture behavior in mice[59] and how rats can use vision as an emotional communication channel with conspecifics[60]. On a more fundamental level, the spontaneous ability of rats to solve invariant object recognition, (i.e. to keep recognizing visual objects even in spite of identity-preserving transformations such as pose, scale and position) has been thoroughly demonstrated[61–63].

On the other hand, a direct demonstration that rat spontaneously solve the aperture problem (i.e. integrate local motion cues to perceive object global motion direction) is still lacking. Some studies demonstrated the ability of mice to discriminate the direction of random dot kinematograms[64,65] by relying on V1[66]. Such behavioral evidences however, while requiring the spatial integration of local motion signals to happen in the brain of trained animals, do not imply the existence of integrated motion representation in the visual cortex of naïve animals. The fourth subproject described in this thesis was aimed at filling this gap.

Let's now briefly review the anatomy and function of the rat visual system.

Rodents lack a fovea[67]: centro-peripheral retinal ganglion cell (RGC) density ratio in rats is 0.2 (from 3000 in the area centralis to 600 cells/mm^2 in the periphery) against factors of the order of approximately 0.001 characterizing foveated monkey retinas[68]. The lack of a proper fovea makes exploratory fixation likely not useful for rodents (i.e. gaze shifts ought to bring the image of an object of interest into the higher resolution-region of the visual field). However, studies characterizing the dynamics of rat eye movements in head-fixed rats reported the presence of spontaneous stereotypic saccades, distributed mainly along the horizontal axis, with little vertical components[69,70]. In freely moving condition on the other hand, the dynamics of their eye movements are complex, deconjugate, often asymmetrical, and frequently driven by vestibulo-ocular reflexes[71]. Differently from primates and other

mammals, rodents are also unable to accommodate[72] (i.e. perform adjustment of the eye optics to keep an object in focus on the retina as its distance from the eye varies). However, in high pupil constriction state, rat's eye shows a considerable depth of focus[73], way exceeding human one. Such property enables them to form an acceptably sharp image on the retina over a span of tens of cm without needing accommodation.

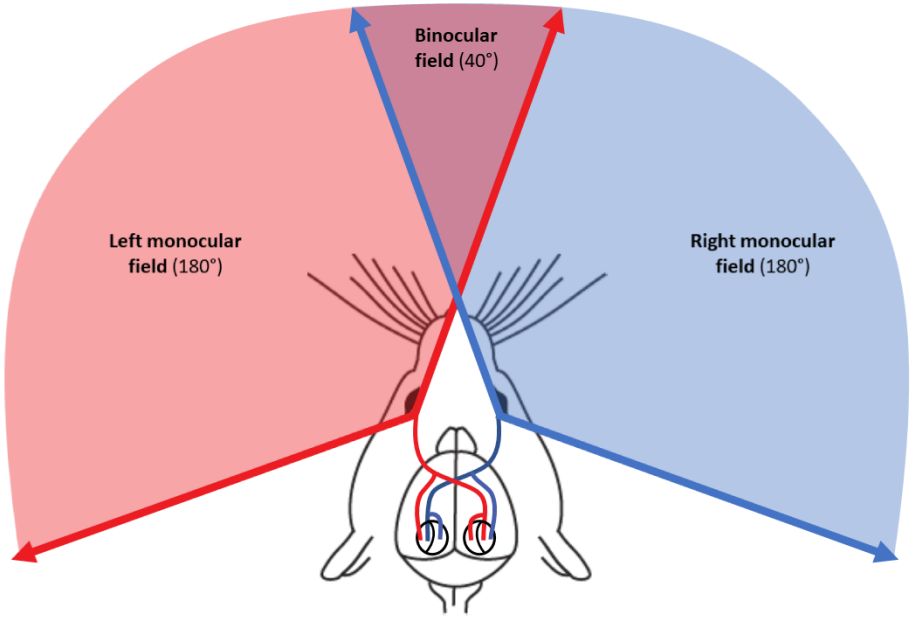

Figure 9. Binocularity in the rat. The lateral placement of rat eyes results in each eye viewing largely nonoverlapping regions of the visual scene (left and right monocular fields, in red and blue respectively). The fields of both eyes overlap in the central (i.e. rostral) 40° of visual space enabling some stereopsis (binocular field, in violet). Red and blue wires connecting the eyes with the occipital part of the brain represent optic nerve and optic radiation carrying visual information to V1 (represented by the black ellipse); notice the crossing at the center of the brain representing the optic chiasm. Curved black segment depicted within the V1 ellipse represents the boundary between the monocular (receiving input from the contralateral eye only) and binocular (receiving input from both eyes) region of V1. Adapted from Seabrook et al.[87].

Consistently with their natural non-diurnal lifestyle their retina is composed at 99% by rods[74]. The remaining 1% of photoreceptors are cones, out of which 93% show a peak in absorption spectrum at 510 nm (blue-green) and 7% at 360 nm (ultraviolet)[75]. In this respect it's interesting to note that, due to the above mentioned color sensitivities of cones in the rat retina, the "white" emitted by monitors (i.e. a combination of red, green and blue light adjusted for human perception) would likely not appear as white to rats[76].

Due to the lateral placement of their eyes on their head, rats possess a very wide-angle field of view characterized by a small frontal region of binocular overlap measuring 40° on the horizontal plane to 60° degrees on the vertical one[71] (see Fig. 9).

Finally, as already mentioned, rats are very low acuity animals: their vision have 50 folds poorer spatial resolution than human one. While human visual acuity is close to 50 cycle/°, rats acuity varies from 0.5 to 1.5 cycle/° depending on their strain[77–79] and according to different behavioral and electrophysiological measurements (see Fig. 10).

When the light pattern of an image travels through the eye of a rat and finally gets absorbed by photoreceptor cells in the retina, visual information is encoded in the pattern of electrical activity of retinal circuits. Here interneurons like horizontal, bipolar and amacrine

cells take part in shaping the output encoded by the RGCs. In the mouse retina 33 different distinct subtypes of RGC have been identified[80] each coding for distinct spatiotemporal features in the input image (including oriented and directional ones). Such reformatted visual information leaves the eye travelling through the optic nerve towards more than 50 retino-recipient subcortical brain areas[81].

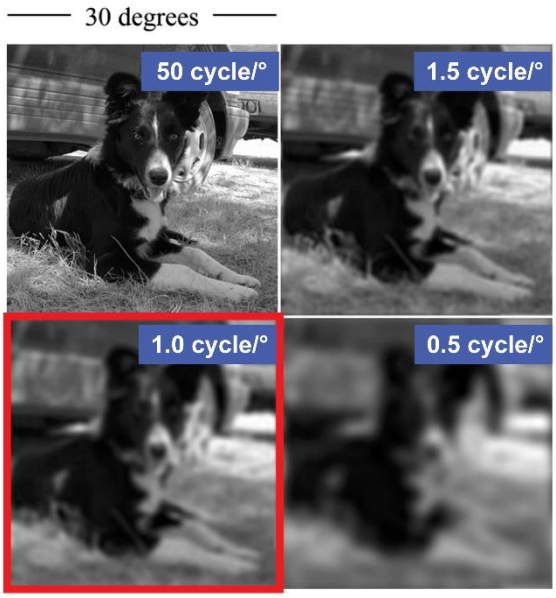

Figure 10. Visual acuity of different rat strains. The original image (top-left, tagged 60 cycle/° to represent human visual acuity) has been gaussianly blurred to give a visual impression of the way rats perceive the visual stimuli (assumed of 30° of width). Acuity levels reported in tags and visualized by smoothing where obtained testing Fisher-Norway (top-right), Long Evans (bottom-left, squared in red) and Wistar (bottom-right) rats in a visual water maze task. Adapted from Prusky et al.[77].

Among those, the key one in transmitting information to primary visual cortex (V1) (and therefore for visual perception) is the dorsolateral geniculate nucleus (dLGN) of the thalamus. Given the above-mentioned functional diversity of mouse RGCs one could expect a similarly rich functional representation in mouse dLGN. In contrast, the majority of mouse dLGN neurons has been reported to have non-oriented, circularly symmetric, RFs[82,83] and to perform linear spatial summation[84,85], similar to what can be observed in primates (i.e. conveying a representation that could still be conceptualized as a "pixel-based" one). In mouse, however, a dedicated pathway (anatomically segregated from the classical retino-geniculo-cortical one carrying non-direction-tuned information to L4 of V1) has been discovered relaying some directional and orientation selective signals from retina directly to V1[86]. Even if subcortical visual pathways in rats are much less well characterized than those of mice[87] to which we referred above, similar organizational principles likely hold true for this closely related species.

As final subcortical information processing step, dLGN representations are relayed to the primary visual cortex trough the optic radiation. Primary visual area (V1) is the largest and most studied visual area in rodents (and more generally in mammals). Rat V1 occupies about 10% of the total neocortex[53] covering the most posterior surface of both hemispheres. It spans a length of about 5mm in the anteroposterior (AP) direction (from AP -4.4 mm to AP -9.36 mm) and it reaches a maximal mediolateral (ML) width of about 4 mm width at bregma -7.92 mm[88] (where it spans from ML ~1.4 mm to ML ~5.8 mm). The most medial part of it is

often called monocular V1 (V1M) since it receives thalamic input originating from the contralateral eye only (and therefore contains monocular neurons only). The most lateral part is on the other hand often called binocular V1 (V1B) since it receives thalamic input originating from both eyes (and therefore contains binocular neurons). As in other mammals V1 displays a retinotopic organization: visual space is smoothly mapped over the cortical surface (i.e. physically close neurons in it will respond to stimuli at adjacent positions in the visual field). In contrast with what has been found in other species such as monkey and cats, however, no clear orientation columns have been found in rat[89] and mouse[90] visual cortex. Interestingly, a more recent study however reported a significant degree of spatial clustering of iso-oriented neurons suggesting such organization to be a degraded version of the one observable in other mammals, hinting at a possible common origin[91].

V1 (as any cerebral cortex more in general) is well known to be structured in a complex, laminar (i.e. subdivided in layers) microcircuit where different neuron types give specific contributions to the computation implemented by the tissue. The canonical view[92] is that thalamic afferents synapse mainly on L4 neurons and in L5-L6 border (even if, to lesser extent, in all cortical layers). L4 principal cells project in a mainly feedforward way to all layers, but most strongly to L2/3, receiving little feedback in return. L2/3 principal cells on the other hand project locally to L5 as well as distal cortical targets (contralateral cortex and high-order cortices). L5 in turn contains two distinct principal cell classes: "intratelencefalic neurons" projecting locally back to L2/3 and distally to the high-order (and contralateral) cortex, as well as "pyramidal tract neurons" exhibiting a connectivity pattern similar to the inverse of L4 principal cells (i.e. receiving input from multiple cortical layers and projecting to brain stem and spinal cord but also to high-order thalamus). L6 on the other hand receives input from all other layers and sends projections targeting primary sensory thalamus and L4 inhibitory interneurons (i.e. "corticothalamic neurons"; likely with gain control functions) as well as long-range cortical targets (i.e. "corticocortical neurons"). Lastly L1 does not contain principal cell bodies but rather apical dendritic tufts of L5 and L2/3 neurons onto which axons from high-order cortex and thalamus synapse (carrying high order feedback and contextual modulation signals) (see Fig. 11).

Given the rich laminar structure of the cortex it is easy to understand how helpful it is to know the laminar location of channels one is recording from. Usually this is attained by means of time-consuming histological analyses. The second subproject described in this thesis originated from the will to ease such burden. In order to do so we developed an automated tool for extracting information about the laminar location of a multichannel silicon probe from the shape of the local field potential recorded there in response to sensory stimulation.

First investigators who carried out extracellular recordings in rat V1 reported neurons responding to spots, bars, edges and flashes of light with an average RF size of about 15°[93]. Others assessed that most of V1 neurons were at least broadly tuned for orientation, reporting that about 70% of the orientation-selective cells display an orientation tuning bandwidth lesser then 60°. On the other hand, the fraction of direction selective units was less than half of that reported for orientation[79]. In this early stage of the exploration of rat V1 functional properties, single neuron's firing temporal modulation in response to drifting gratings was also analysed. This enabled to evidence the existence of distinct simple and complex

functional cell classes in rat V1 similar to the ones observed in other mammals such as cats and monkeys[79,94]. Such latter report was critical in motivating my PhD work since, as already mentioned, simple and complex cells are great testbed for understanding the principles underlying high-level vision.

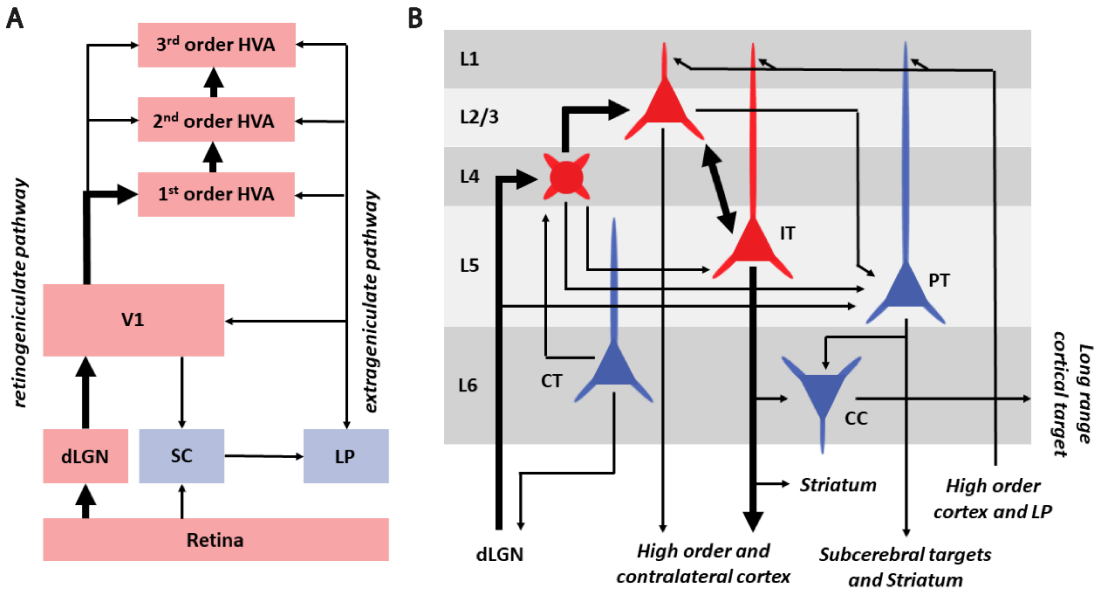

Figure 11. Schematics of the rodent visual system and V1 microcircuit. A) Schematic of feedforward connectivity in the retinogeniculate and extrageniculate pathways of the rodent visual system. Areas participating the retinogeniculate pathway are highlighted in red, areas participating the extrageniculate pathway are highlighted in blue. Thick black arrows depict the feedforward canonical flow path of information from retina trough the visual cortical hierarchy. Axes label keys: dLGN, dorsal lateral geniculate nucleus of the thalamus; SC, superior colliculus; LP, lateral posterior nucleus of the thalamus; V1 primary visual; HVAs, higher-order visual areas. Adapted from Glickfeld et al.[24]. B) Schematic of the laminar structure of the V1 cortical microcircuit (only principal cells are shown) . Alternating gray bands represent cortical layers (labelled on the left). Arrows represent connections between two principal cell classes. Thick black arrows depict the feedforward canonical flow path of information trough V1 layers from thalamus to high-order cortex (as HVAs). Principal cells involved are highlighted in red, other cell classes in blue. Area names on the border of the scheme represent projection source/target of entering/exiting arrows. Cell classes label keys: IT, intratelencephalic; PT, pyramidal tract; CT, corticothalamic, CC, corticocortical. Adapted from Harris et al.[92].

Simple cells were classically characterized since the pioneering work by Hubel & Wiesel[28] as position and orientation selective edge detector units responding to a light bar of a defined orientation at a given position. Complex cells, on the other end, where defined as orientation selective but, at least locally, position invariant edge detectors, insensitive to small changes in position of their preferred bar.

The identity of an unknown neuron can therefore be assessed with drifting gratings by quantifying how locked the neuronal response is to the phase of the stimulus. Simple cells, performing linear spatial summation, will exhibit phase-sensitive behavior, responding strongly only when a white stripe of the grating is correctly overlapping their receptive field. Complex cells on the other hand, performing non-linear spatial summation, will show a uniform phase-insensitive response (see Fig. 12).

In hypothesizing a circuit mechanism to explain their response properties Hubel & Wiesel originally suggested those two cell types to represent two consecutive (i.e. hierarchical) stages of processing. Complex cells where deemed to build their response

properties by pooling over many iso-oriented but slightly displaced simple afferents. Simple cells on the other hand were thought to build their oriented receptive fields by integrating input from many aligned center-surround thalamic afferents.

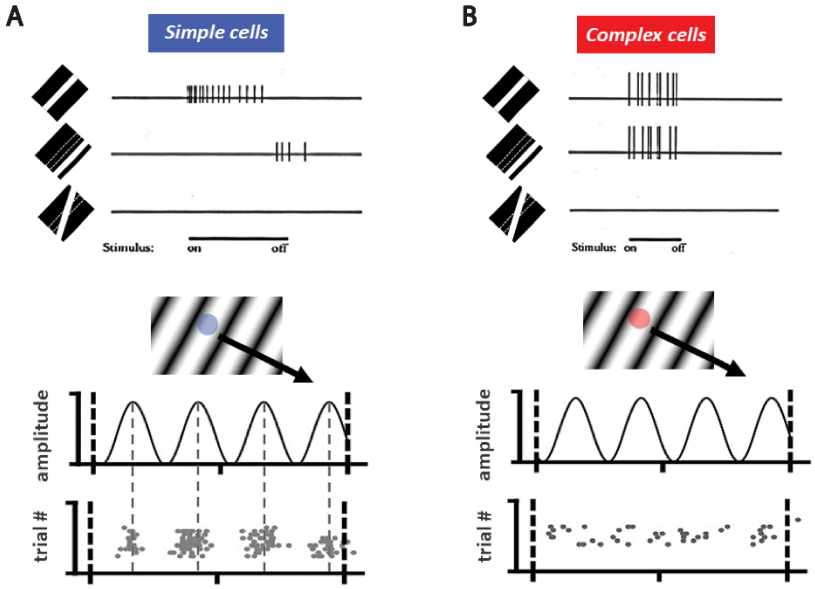

Figure 12. Simple and complex cells. A) Simple cells are functionally defined as position and orientation selective edge detector neurons. As visible in the top part of the panel such neurons respond optimally (i.e. producing more spikes, represented by vertical ticks) to a light bar of correct orientation placed at a given position in the visual field. Changing both position and orientation of the preferred bar reduces the response. This is revealing of simple cells acting like linear filters on the input image. When probed with moving gratings (such as the one represented in the central subpanel) such linear units exhibit a phase-locked firing pattern responding strongly only when a white stripe of the grating is correctly overlapping their RF. B) Complex cells are also functionally defined as orientation selective but, at least locally, position tollerant edge detector neurons. As visible in the top part of the panel they respond optimally (i.e. producing more spikes, represented by vertical ticks) to light bars of a defined orientation placed at a given position in the visual field, however, differently from simple cells, small displacements of the preferred stimulus will not affect the response. Such property requires some form of nonlinearity in the processing of the input image. When probed with moving gratings (such as the one represented in the central subpanel) such non-linear units are able to exhibit a phase-independent, uniform, firing pattern as long as their preferred stimulus remains in their RF. Adapted from Hubel & Wiesel[28].

The ’80s and ’90s saw different studies in cats directed to validate those circuits models[95]. This was earlier achieved for L4 simple cells[96] whereas for complex cells, validation of the hierarchical anatomical model remained debated for longer. The consensus model that emerged throughout the years included both thalamic-recipient L4 “first-order” complex cells (built by pooling directly over thalamic afferents of reversed RF contrast polarity) as well as L2/3 “second-order” L4-simple-cells-recipient ones[95] (supported by a circuit similar to the one proposed by Hubel & Wiesel − see Fig. 13).

Those ideas however, originated mainly from cat neurophysiology, have never been put to the test in rodents. Some studies[97,98] that investigated the laminar distributions of complex cells in adult mouse however reported a strong predominance of complex cells in L5 of mouse V1 and a predominance of simple cells not only in L4 and L6 (as reported in other mammals[99]), but also in L2/3. This is suggestive of the possibility that rodents may display a shallower laminar processing hierarchy where the first “complexification” (i.e. the built up of

second-order complex cells by integrating simple cells input) of RFs happens later (i.e. in L5) than in cats and primates.

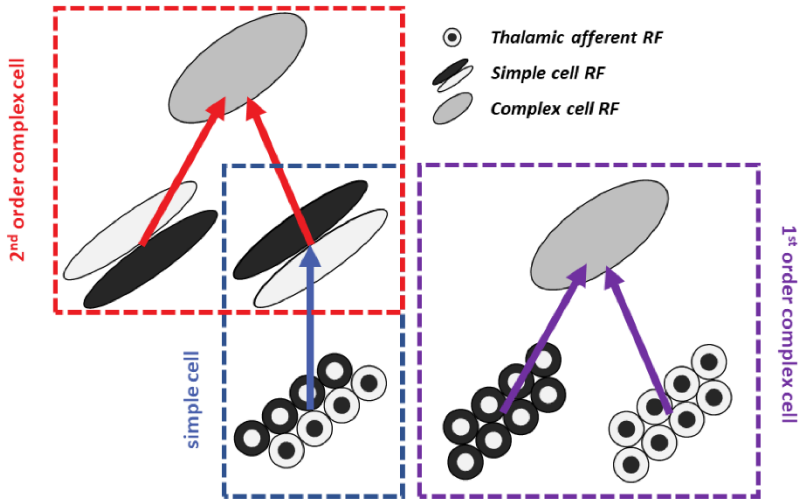

Figure 13. Proposed circuit mechanisms underlying simple and complex cell. The hierarchical model proposed by Hubel & Wiesel put forth the idea that simple cells and complex cells represent two successive stages of cortical processing. To reconcile the evidence for this model with some contradictory findings gathered throughout the years Martinez & Alonso proposed the existence of two subclasses of complex cells with similar RF properties stemming from different circuit mechanism: "first order " (i.e. corticothalamic, mainly located in L4) and "second order" (i.e. corticocortical, mainly located in L2/3) complex cells. Second order complex-cells together with simple-cells form a two-step hierarchy towards tolerant receptive field build up. On the other hand, first order complex cells would build their tolerant receptive fields up directly frpom geniculate inputs pooling over overlapping on and off thalamic afferents. Adapted from Martinez & Alonso[95].

In rodents V1 is bordered by a collection of more than 10 high-order visual areas[24] (HVA). However, determining a parcellation of visual cortex in distinct areas is not a straightforward process that saw many changes of naming conventions across the years. Visual areas can be defined, in fact, according to a convergence of multiple functional and anatomical features. Nowadays, existence of a full (even if biased) retinotopic representation of the visual field together with a distinct projection pattern is the most commonly accepted standard to this aim. Whereas V1 is characterized by its robust retinogeniculate input[100] as well as landmark cytoarchitectural features (from which the name "striate" cortex), extrastriate areas lack them (from which the name "extrastriate" cortices). On the other hand, all extrastriate areas receive visual input from two main pathways: an intracortical one (i.e. collecting input from V1 and the other HVAs) and the so called "extrageniculate" one, connecting them to the lateral posterior nucleus (LP) (receiving input from high-order visual thalamus, in turn connected to the superior colliculus). Since the first pioneering electrophysiological explorations of the rodent visual cortex, retinotopic mapping has been an invaluable tool to discriminate between HVAs[93,101,102]. In particular, Espinoza & Thomas first individuated six areas around V1 containing a full representation of visual space by performing serial electrode recordings[93]. An early tracer-injection study demonstrated that V1 strongly projects to extrastriate areas in a retinotopically[103] organized fashion whereas another one demonstrated that extrastriate areas send feedback connections back to V1 as well as to LP[104]. Still using tracers, Coogan and Burkhalter identified laminarly asymmetric

projection patterns between pairs of connected areas enabling to distinguish feedforward from feedback connectivity[105]. This allowed them, for the first time, to claim the existence of a hierarchy among extrastriate areas. More recently, in mice, the progress towards better definition of extrastriate areas borders was driven by the adoption of intrinsic imaging[106,107] and wide field calcium imaging[108] techniques, leading to high resolution retinotopic mapping of visual cortex (yielding the identification of 16 distinct retinotopically-organized regions).

At the same time anatomical characterization progressed too. One study in mice proposed, on the basis of chemoarchitectonic (i.e. m2AChR expression pattern), cytoarchitectonic (i.e. cell density) and projection criteria, lateromedial (LM) and anterolateral (AL) areas as the "gateways" of ventral and dorsal stream[109] respectively (i.e. as input hubs of the two putative pathways). A year later, the same group refined this picture applying optical densitometry and anterograde tracing to quantitatively determine the projection strength of 10 main visual areas towards 39 cortical targets (including the other visual areas). Such a rich characterization of the visual cortical graph enabled to analyse its community structure, evidencing two clear subnetworks. The first one included medial/anterior extrastriate areas and was more strongly linked to parietal, motor, and limbic cortices, strongly resembling the dorsal stream in primates[110]. The second one included lateral extrastriate areas and was preferentially connected to temporal and parahippocampal cortices, strongly resembling the ventral stream in primates. This result provided a very solid anatomical ground for the existence of ventral and dorsal pathways homologues in the rodent cortex.

Evidence for the existence of functionally distinct temporal "ventral-like" and parietal "dorsal-like" pathways in rodent extrastriate cortex came from lesion studies too. In a seminal work, Tees showed that damage to posterior temporal extrastriate areas (likely including LM, LI, LL, AL, and POR) induces in rats a deficit in the ability to detect change in the identity of an object located in a familiar location. On the other hand, damage to posterior parietal extrastriate areas (likely including RL, A, AM and PM) induces the opposite deficit: animals are unable to recognize the change of position of a familiar object while remaining able to detect change in its identity at a familiar location[111]. Other groups showed that damaging temporal and perirhinal cortex (downstream of LM and LI) destroys object recognition while sparing spatial memory[112–115]. Conversely other investigators reported impairment in visuospatial tasks after lesioning the posterior parietal cortex[116,117].

A few groups tried to characterize the functional specialization of extrastriate areas in order to bring some neurophysiological support to the proposed ventral/dorsal distinction. The approach adopted in first mice studies was grounded in the assumption of being able to distinguish dorsal and ventral areas from their spatial and temporal frequency preferences (high SF / low TF = ventral; low SF / high TF = dorsal). Such assumption makes sense in light of the differential routing of detailed & static vs. coarse & moving visual information that is known to take place across monkey streams[118]. This approach worked as expected for part of the putative dorsal stream (AL, RL and AM) that exhibited the SF and TF preferences compatible with optic flow processing areas. PM on the other hand, while firmly in the dorsal stream from anatomical grounds, displayed apparently more ventral characteristics, similarly to LI. LM broke the anatomical expectations too, preferring more "dorsal" SF and TF[119,120]. Even if showing an interesting degree of specialization and target-specific routing among

extrastriate areas[121], the above-mentioned studies yielded an inconclusive and conflicted picture.

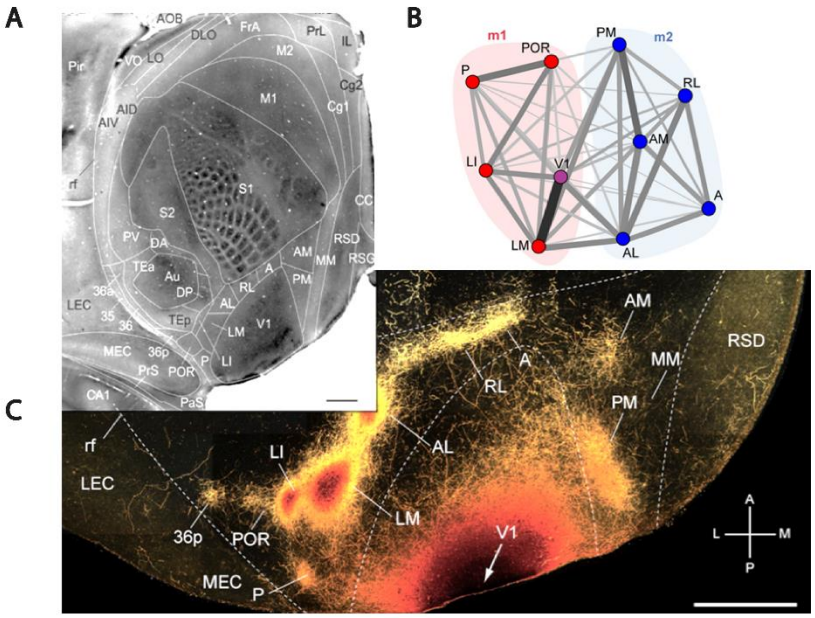

Figure 14. Anatomical evidence for ventral and dorsal networks in the mouse visual cortex. A) CO (cytochrome C oxidase) stained tangential section through layer 4, showing differential expression across the whole cerebral cortex displaying position and size of V1 relative to other areas. B) Modularity structure of the inter-areal connectivity graph obtained from tracing experiments via optical densitometry. Red areas belong to the ventral module. Blue areas belong to the dorsal module. V1 is coloured in purple. Connections between pairs of areas are shown as the sum of their reciprocal projection density (darker gray/thicker lines symbolize stronger projections). C) Dark-field image of anterogradely labelled axonal projections of V1 (high-density regions are red, even higher ones are red-brown) with an injection of BDA tracer (biotinylated dextran amines; arrow). Section through flat-mounted mouse cerebral cortex. Scale bars: 1 mm. Axes label keys: A, anterior; P, posterior; M, medial; L, lateral. A, Anterior. Area label keys: AID, anterior dorsal insula; AIV, anterior ventral insula; AL, anterolateral; AM, anteromedial; Amy, amygdala; AOB, accessory olfactory bulb; Au, auditory; CA1, hippocampus; CC, corpus callosum; Cg1, cingulate 1; Cg2, cingulate 2; DA, dorsal anterior; DLO, dorsal lateral orbital; DP, dorsal posterior; FrA, frontal association; IL, infralimbic; LEC, lateral entorhinal; LI, lateral intermediate; LM, lateralmedial; LO, lateral orbital; M1,motor 1; M2, motor 2; MEC, medial entorhinal; MM, mediomedial; MO, medial orbital; OB, olfactory bulb; P, posterior; PaS, parasubiculum; Pir, piriform; PM, posterior medial; POR, postrhinal; PrL, prelimbic; PrS, presubiculum; PV, parietal ventral; rf, rhinal fissure; RL, rostrolateral; RSD, retrosplenial dysgranular; RSG, retrosplenial granular; S1,somatosensory 1; S2, somatosensory 2; TEa, temporal anterior; TEp, temporal posterior; Tu, olfactory tubercle; V1, primary visual; VO, ventral orbital. Adapted from Wang et. al.[110].

More recently Smith et al.[122] tracked the development of visual responsivity across mouse visual cortex from eye opening throughout the critical period. They found that the average visual response to gratings, observed via intrinsic imaging, shows a much higher correlation within the putative dorsal and ventral subnetworks rather than between them. They also reported two different and coherent developmental trajectories for those two clusters of areas. Whereas ventral areas are stable in their response magnitude during postnatal development, dorsal ones increase their responsivity after eye opening in an experience dependant way. The authors then looked at the single neuron level to one representative ventral area (LM) and a dorsal one (PM). Doing so, they observed a refinement of functional properties consisting in a RF size and orientation selectivity increase in the first but not in the second area. Those evidences corroborated the ventral/dorsal clustering suggested by anatomical studies.

Parallelly, some studies marked a significant step forward in the functional characterization of extrastriate areas by pushing the focus beyond mere spatial and temporal frequency preferences toward properly ventral and dorsal computations. On one hand Juavinett et al. used gratings and plaid stimuli to characterize the amount of pattern and component cells in mouse V1 and 4 extrastriate areas (LM, AL, RL and AM). Only RL and, somewhat surprisingly (given its anatomical inclusion in the ventral stream), LM displayed neurons capable of solving the aperture problem signalling global motion direction[123] (i.e. pattern cells). The relatively high fraction of pattern units observed in RL, an area already implicated in visuotactile integration[124], pushed it forward as the most promising candidate to be a rodent homolog of MT. On the other hand, two studies by Vermaercke et al.[125] and Vinken et al.[126] were the first to adopt richer stimulus sets (including 2D objects and natural movies respectively) and to target near- as well as far-lateral extrastriate cortex of rats. Those first rat studies found only weak and contrasted evidence for high level shape processing going on in these lateral areas. A few years later, however, Tafazoli et al., adopting an even richer (3D objects undergoing size, position, luminance and both in-plane and in-depth rotations) stimulus set and more advanced analytical tools, managed to get sharper results. They directly demonstrated an increase in the ability of both single neurons and neuronal populations to support discrimination of visual objects under identity-preserving transformations across the V1→LM→LI→LL hierarchy[127]. Their result pushed forward LI and LL as candidate rodent homologues of the highest stages of the primate ventral stream (V4 and IT). Conversely, LM, in light of its mixed properties, could be reasonably mapped to primate area V2: one step beyond V1 in the processing hierarchy, routing information to both ventral and dorsal streams.

The picture drawn by the studies discussed above, although far from being conclusive, offer a solid standpoint to devise new studies aiming at testing further the proposed ventral and dorsal homology of different extrastriate areas.

Two of the subprojects described in this thesis were aimed at this target. The first one searching for the signatures of complex selectivity built up and position tolerance increase from V1 to LL, and a second one investigating the computational mechanism behind pattern responses in V1 and LM (first and fifth subproject).

N.B. − references for papers cited in the Introduction (above) are listed together with those of Conclusion (from p. 140).

Chapter I

Nonlinear processing of shape information in rat lateral extrastriate cortex

Abstract

In rodents, the progression of extrastriate areas located laterally to primary visual cortex (V1) has been assigned to a putative object-processing pathway (homologous to the primate ventral stream), based on anatomical considerations. Recently, our group found functional support for such attribution, by showing that this cortical progression is specialized for coding object identity despite view changes – the hallmark property of a ventral-like pathway. Here, we sought to clarify what computations are at the base of such specialization. To this aim, we performed multielectrode recordings from V1 and laterolateral area LL (at the apex of the putative ventral-like hierarchy) of male adult rats, during the presentation of drifting gratings and noise movies. We found that the extent to which neuronal responses were entrained to the phase of the gratings sharply dropped from V1 to LL, along with the quality of the receptive fields inferred through reverse correlation. Concomitantly, the tendency of neurons to respond to different oriented gratings increased, while the sharpness of orientation tuning declined. Critically, these trends are consistent with the nonlinear summation of visual inputs that is expected to take place along the ventral stream, according to the predictions of hierarchical models 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 the visual cortex of primates and rodents, reasserting the potential of the latter as models to investigate ventral stream functions at the circuitry level.

Introduction

In rodents, primary visual cortex (V1) is bordered by at least 9 extrastriate visual areas, whose involvement in dorsal and ventral stream processing (i.e., extraction of, respectively, motion and shape information) has become the object of intense investigation. To date, most studies on this front have been inspired by the routing of visual information in the specialized channels that, in primates, selectively relay low temporal/high spatial frequency signals to the ventral stream only, and high temporal/low spatial frequency content to both the ventral and dorsal streams[1,2]. Inspired by a possible homology with primates, and by the existence of distinct V1 subpopulations making (receiving) target-specific projections with (from) downstream (upstream) areas[3–6], several investigators have mapped mouse visual areas with drifting gratings of various spatiotemporal frequencies[7–11], finding the signature of dorsal processing in medial and parietal extrastriate cortex, but only scant evidence of ventral computations in lateral extrastriate cortex. This is not surprising, because the increase in the complexity of shape selectivity and tolerance to image variation that is the signature of ventral processing cannot be routed from upstream areas, but has to emerge as the result of local, non-linear integration of presynaptic inputs[12].

To find the signature of such computations, our group recently investigated how visual objects are represented along rat lateral extrastriate areas[13]. These 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, thus supporting a central role of LL as the apex of the rat object-processing hierarchy. At the same time, studying cortical representations of visual objects has the disadvantage of relying on idiosyncratic choices of the stimulus set (the object conditions) and rather complex interpretative approaches (e.g., information theory and machine learning). This makes the experimental design hard to standardize, and often leads to conclusions that are only partially overlapping across studies. The assignment of rat lateral extrastriate cortex to the ventral stream is no exception, as shown by the much weaker specialization for object processing found in this region by earlier studies[14,15].

This prompted us to verify the higher rank of LL, relative to V1, in ventral processing, by designing an experiment that exploited the benefits of parametric stimuli, but, rather than focusing on spatiotemporal frequency tuning, compared the two areas in terms of the nonlinearity of the stimulus-response relationship and the tendency of a neuron to be selective for different stimulus orientations. The first property was measured by the degree to which neuronal responses were entrained to the phase of drifting gratings, and by the extent to which the structure of the neuronal receptive fields could be inferred through reverse correlation. The second property was assessed by detecting the presence of multiple peaks in the orientation tuning curves and by measuring their sharpness.

Critically, all these properties followed trends of variations that are consistent with the nonlinear summation of visual inputs that is expected to take place along the ventral stream, according to the predictions derived from: i) a meta-analysis of the monkey literature; ii) a conceptual model of hierarchical ventral computations[16]; and iii) a state-of-the-art deep

convolutional network trained for image classification[17]. 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 (Fig. 1A) of anaesthetized rats that were presented with drifting gratings of various orientations, spatial (SF) and temporal (TF) frequencies, as well as with 80 30 s-long movies of spatially and temporally correlated noise (see Materials and methods). A total of 168 and 208 well-isolated single units were recorded, respectively, from V1 and LL. Among these neurons, 63% (i.e., 105 V1 units) and 50% (i.e., 104 LL units) were effectively and reproducibly driven by one or more of the grating stimuli across repeated presentations (see Materials and methods). Before the main stimulus presentation protocol, we also ran a receptive field mapping procedure (Materials and methods) to track the progression of the RFs recorded along a probe (Fig. 1B). Given that, in rodents, the polarity of the retinotopic map reverses at the boundary between each pair of adjacent visual cortical areas[18], tracking the retinotopy of the RFs allowed a precise identification of the units that were recorded from LL. This was possible because, to reach LL, the probe was inserted obliquely (with an angle of 30º relative to the vertical to the surface of the skull), in such a way to first cross the lateral extrastriate areas that are medial to LL – i.e., the lateromedial (LM) and laterointermediate (LI) areas (Fig. 1A). In the case of V1 recordings, the probe was inserted either perpendicularly or with a variable angle (between 10º and 30º).

For each recording session, we also performed the histological reconstruction of the probe insertion track, appositely coated with a fluorescent dye before the penetration (see Fig. 1A). The outline of the probe, with the known location of the recording sites, was then superimposed to an image of the cortical section stained for Nissl substance, in such a way to infer the laminar location of the recorded single units (see Material and methods). In both areas, our recordings targeted the infragranular layers, because the tilted insertion that was necessary to track the retinotopy of lateral areas (allowing a reliable identification of LL − see previous paragraph), naturally brought the recording sites to land mostly in layer 5. Therefore, in order to allow a fair statistical comparison with LL, the probe was aimed at the infragranular laminae also in V1.

Our histological analysis confirmed that the laminar sampling was highly consistent between the two areas, with the vast majority of neurons being recorded from layer 5 in both V1 and LL (Fig. 1C). Critically, this ensures that the tuning properties being compared in our study refer to nearly identical laminar populations. At the same time, the selective targeting of layer 5 does not undermine the generality of our comparison. In fact, previous investigation of V1 and lateral extrastriate areas showed that the same increase in the specialization for ventral processing along the areas' progression was observable across the whole cortical thickness, and was equally sharp in superficial and deep layers[13].

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 Fig. 2, 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).

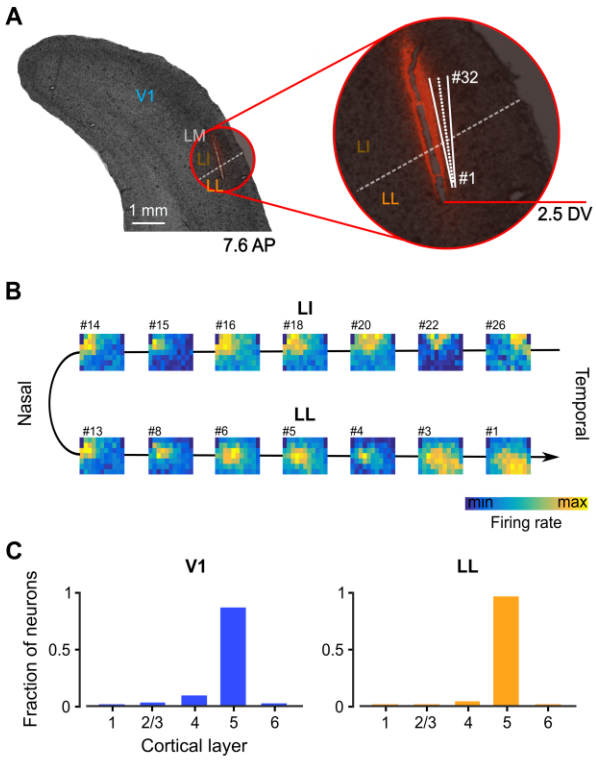

Figure 1. Targeting rat laterolateral (LL) visual cortical area for multi-electrode recordings using a linear silicon probe. A) Representative Nissl-stained coronal slice showing the insertion track of a single-shank silicon probe aimed at area LL. The probe was inserted obliquely into visual cortex, crossing lateromedial (LM) and laterointermediate (LI) visual areas before reaching LL. The dashed line depicts the border between LI and LL, as reconstructed by tracking the retinotopy of the receptive fields (RFs) recorded along the probe (see panel B). The position of the probe during the recordings is marked by the staining with the fluorescent dye (red) that was applied to it before insertion, and by the mechanical lesion produced by the penetration. Based on this information, it was possible to estimate the geometry of the probe (white lines) at the final recording depth (2.5 mm DV), along with the location of the recording sites (white dots), from the tip (site #1) to the base (site #32). B) Firing intensity maps displaying the RFs of the units recorded along the probe shown in A. The numbers refer to the sites each unit was recorded from. The reversal of the retinotopic progression (from leftward to rightward) at sites #13-14 marks the boundary between areas LI and LL. C) Laminar distributions of the responsive and reproducibly-driven V1 (blue) and LL (orange) single units, for which it was possible to establish the cortical layer through histological analysis.

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 selectivity for multiple oriented features[19,20].

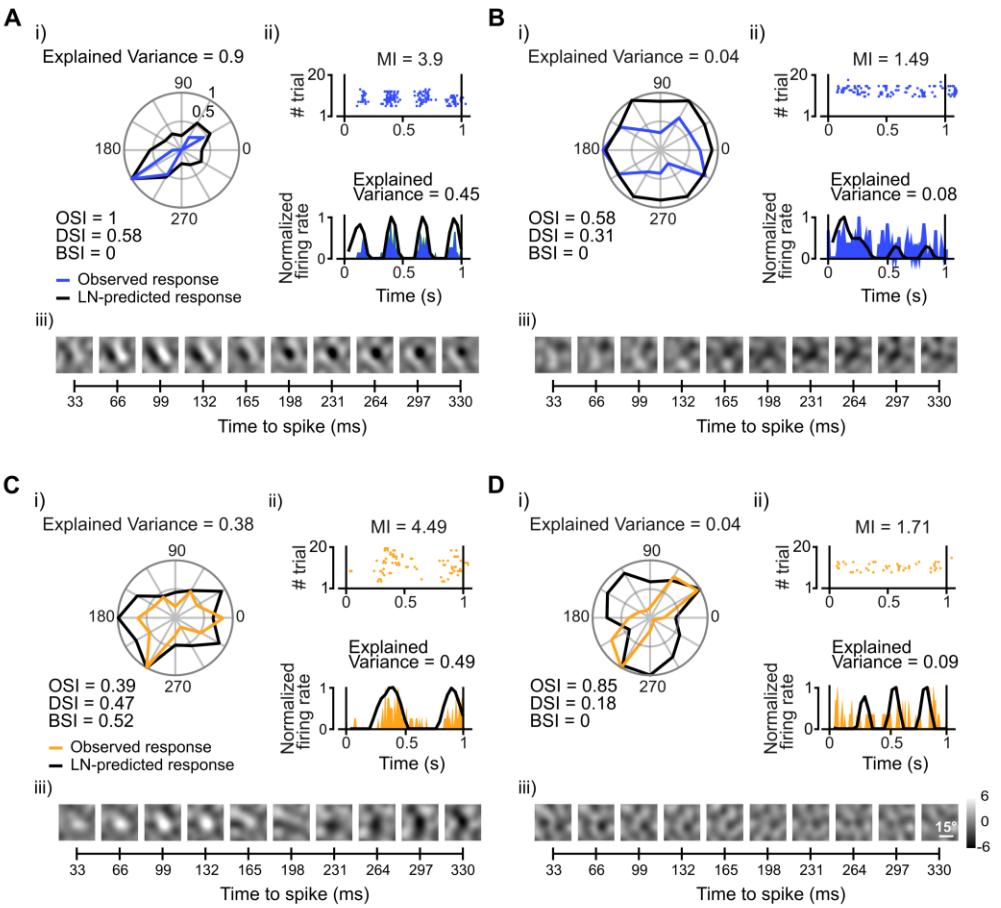

Figure 2. Examples of neuronal tuning in V1 and LL. Representative V1 simple (A) and complex (B) cells are shown in the upper panels (blue lines). Representative LL simple-like (C) and complex-like (D) cells are shown in the lower panels (orange lines). In each panel, subpanels i-iii illustrate different 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 quantify 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.

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[21–23], 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[24,25] 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).

Fig. 2A 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 (blue curves/dots). These properties match those of a "simple cell"[26], 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 (Fig. 2A.iii) had a Gabor-like, double-lobed appearance, with an orientation matching the one corresponding to the peak of the tuning curve (Fig. 2A.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 (Fig. 2A.i and ii, respectively; black curves), thus confirming the linearity of the stimulus-response relationship.

By contrast, the example unit shown in Fig. 2B, 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"[26]. This is shown by the lack of modulation of the response to the preferred grating (Fig. 2B.ii), by the poorly defined spatiotemporal structure of the STA-based RF (Fig. 2B.iii), and by the failure of the latter to account for the tuning and time-modulation of the response (Fig. 2B.i-ii; compare blue and black curves).

Interestingly, simple-like and complex-like units were also found in area LL. For instance, Fig. 2C shows an LL neuron with a simple-like behavior, as indicated by the phase modulation of its response (Fig. 2C.ii; orange 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 (Fig. 2C.iii), which, however, was not able to explain as much variance of the direction-tuning curve (Fig. 2C.i; black vs. orange curve) as in the case of the example V1 simple cell (Fig. 2A.i). This could be due to the sensitivity of the neuron to 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 (Fig. 2C.i; orange curve) – a behavior that cannot be captured by a RF with a linear, Gabor-like structure.

Fig. 2D 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 (Fig. 2D.ii), and both the direction-tuning curve (Fig. 2D.i) and the response dynamics (Fig. 2D.ii) were poorly explained by the STA-reconstructed RF (orange vs. black curves), which, in turn, lacked any clear spatiotemporal structure (Fig. 2D.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[16,20,27–30].

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 (Fig. 3).

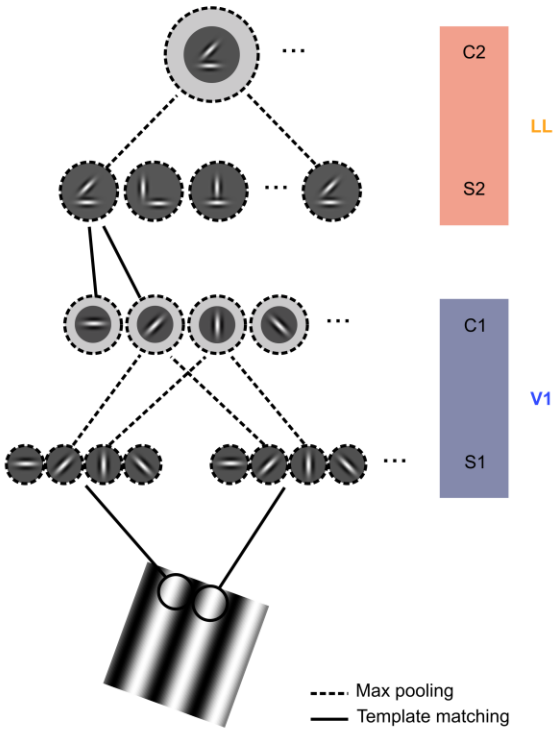

Figure 3. Sketch of the HMAX model implementation used in this study. The model is a feedforward, 4-layered neural network that processes arbitrary input images through a cascade of computations. The first layer (S1) consists of a bank of Gabor filters performing local edge detection on small input image patches (dark gray circles), similar to V1 simple cells. In the second layer (C1), each unit pools over many iso-oriented S1 units via a max-like operation (dashed lines), 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 lines). 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 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 Fig. 3), 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)[12]. 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).

We first turn to investigating the linearity of the stimulus-response relationship.

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 ($p = 5.85*10^{-6}$; Kolmogorov-Smirnov test; Fig. 4A).

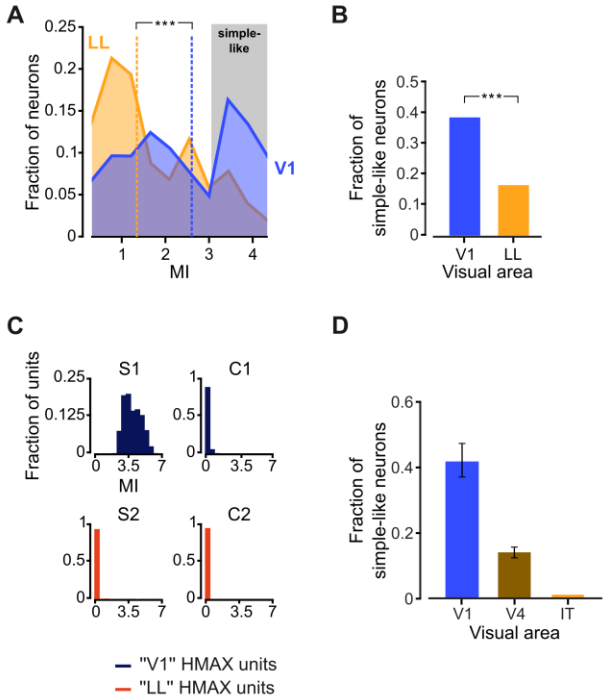

Figure 4. The entrainment of the neuronal response to the phase of drifting gratings decreases from V1 to LL. A) Distributions of the modulation index (MI) for the populations of V1 (blue; n = 105) and LL (orange; n = 104) neurons. Medians are shown as dashed lines (***p < 0.001; one-tailed Mann-Whitney 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 (blue) and LL (orange; ***p < 0.001; $\chi 2$ test). C) Evolution of the MI distribution across the layers of the HMAX model (Fig. 3). Note the different scale on the ordinate axis for the S1 layer, as compared to downstream layers. D) 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 (blue; n = 10), V4 (brown; n = 3) and IT (orange; n = 1) of the monkey ventral stream (mean across n studies in each area ± SE; see Table 1). The drop of phase modulation across the three areas was statistically significant (p < 0.01, one-way ANOVA).

In agreement with earlier studies of rat[31] and mouse[9,32] primary visual cortex, the MI distribution obtained in V1 (blue 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 (orange 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; orange vs. blue line) and this difference was highly significant ($p = 8.04*10^{-6}$; Mann-Whitney U-test). Similarly, the fraction of simple-like cells in LL was half as large as in V1 (Fig. 4B; $p = 1.93*10^{-4}$, χ^2 test, df = 1).

As illustrated in Fig. 2D.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 (Fig. 4C), 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 Fig. 4A; blue 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 Fig. 4A and B; orange curve/bar).

The agreement between the evolution of the modulation index observed in HMAX and in our data suggests the existence of specialized computations that build up transformation tolerance along rat lateral extrastriate cortex. To further support this conclusion, we checked whether the same qualitative trend has been reported in the monkey ventral stream literature. To this aim, we performed a simple meta-analysis of the evolution of phase-sensitivity to drifting gratings across areas V1, V4 and IT of primates. In V1, the fraction of simple cells reported by various authors is quite variable [6,33–41], ranging from ~16% to 63% (Table 1), with a mean across studies of ~41% (blue bar in Fig. 4D).

According to the few studies that investigated the phase-dependence of neuronal responses to grating stimuli in downstream areas (Table 1), the proportion of simple-like cells diminishes to ~14% in V4[42–44], to become zero in IT[45] (brown and orange bars in Fig. 4D), resulting in an overall significant decrement along the ventral stream ($p = 0.0098$, $F_{2,13} = 7.25$; one-way ANOVA). This trend, which is in agreement with the outcome of the HMAX simulations, confirms that the drop in the fraction of simple-like cells across consecutive visual processing stages can be taken as a strong marker of ventral stream computations. As such, its agreement with the trend observed from V1 to LL in our study (compare to Fig. 4B) suggests the existence of specialized machinery to build invariant representations of visual objects along rat lateral extrastriate cortex (see also Fig. 10A).

To further assess the involvement of nonlinear computations in establishing the tuning of LL neurons, we compared the quality of the RFs inferred through the STA method in V1 and LL. 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).

reference	area	number of units	fraction of modulated units	modulated units criterion
Hubel et al. 1968	V1	272	0.45	qualitative (overlap)
Schiller et al 1976	V1	654	0.42	qualitative (overlap)
De Valois et al. 1982a	V1	343	0.61	F1/F0 > 1
De Valois et al. 1982b	V1	222	0.63	qualitative (drift modulation)
Foster 1984	V1	122	0.28	qualitative (overlap)
O'Keefe et al. 1998	V1	210	0.46	F1/F0 > 1
Ringach et al 2002	V1	308	0.40	F1/F0 > 1
Kagan et al 2002	V1	114	0.22	OI < 0.3
Aronov et al. 2003	V1	70	0.53	F1/F0 > 1
Gur et al 2005	V1	339	0.16	OI < 0.3
DeSimone et al. 1987	V4	322	0.16	F1/F0 > 1
Gallant et al 1993	V4	28	0.11	phase effect (ANOVA)
Gallant et al 1996	V4	103	0.14	phase effect (ANOVA)
Pollen et al 1984	IT	43	0.00	qualitative (drift modulation)

Table 1. Full list of research papers included in the meta-analysis shown in Fig. 4D. The first column lists all the articles used in the meta-analysis by the name of the first author and the publication dates. The second column indicates the cortical area that was studied in each individual article. The third column reports the neuronal sample size of each study. The fourth column shows the fraction of phase-modulated units reported in each article. The bars shown in Fig. 4D were obtained by averaging the values in this column across studies within each individual cortical area. The fifth column specifies the criterion adopted in each study to define phase-modulated units (OI stands for Overlap Index, while F1/F0 is a classical measure of temporal modulation of the neuronal response to the frequency of drifting gratings, which is equivalent to the modulation index used in our study).

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; see Fig. 5A and a full description in Materials and methods) 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.

Fig. 5C 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.

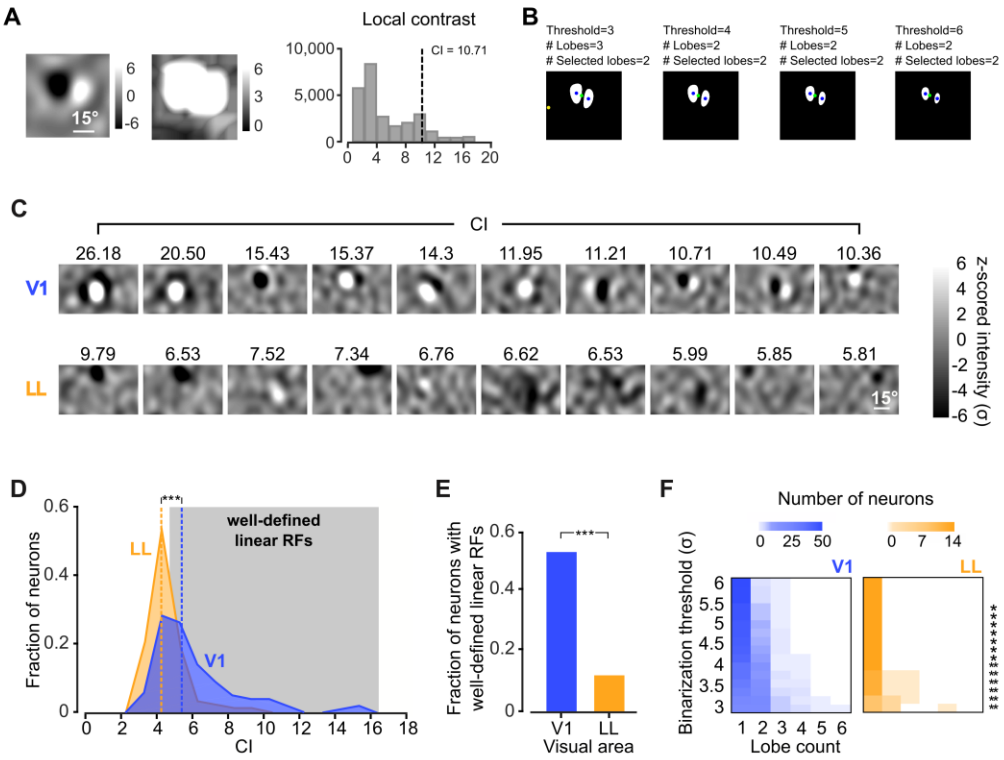

Figure 5. Spike-triggered average yields better-defined and structurally richer receptive fields in V1 than in LL. A) Illustration of the procedure to compute the contrast index of an STA image. The z-scored STA image obtained for a representative V1 neuron (left) is shown along the corresponding local-contrast map (middle), obtained by computing the contrast of the STA image (i.e., peak-to-peak difference) in local convolutional patches, with size equal to one fifth of the total image size (both images are cropped around the RF center). The contrast index (CI) associated to the STA image was computed by considering the resulting contrast distribution (right; bar plot) and taking its 0.9 quantile (dashed line; see Materials and methods). B) Illustration of the procedure to count the lobes of an STA image. The output of the lobe counting algorithm is shown as a function of the binarization threshold (increasing from left to right) that was applied to the modulus of the z-scored STA image of the example neuron shown in A. Green dots show the centers of mass of the binarized images, whereas blue dots represent the centers of "valid" detected lobes (i.e., above-threshold, simply connected regions). Yellow dots mark "invalid" detected lobes (i.e., above-threshold, simply connected regions that were discarded for exceeding a distance limit from the center of mass of the binarized image; see Materials and methods). For this example neuron, a very small spurious lobe was rejected at threshold = 3 (leftmost panel). C) 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 the contrast index CI to quantify the amount of signal contained in a STA image (as illustrated in A). 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 Fig. 2, 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). D) CI distributions for the populations of V1 (blue; n = 105) and LL (orange; 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. E) Fraction of units with well-defined linear RFs (i.e., with CI > 5.5) in V1 (blue) and LL (orange; ***p < 0.001; $\chi 2$ test). F) 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 (blue; n = 56) and LL (orange; n = 13) as a function of the binarization threshold (ordinate) used by the lobe-counting algorithm illustrated in B. Stars indicate threshold values for which the two distributions were significantly different (*p < 0.05, **p < 0.01; Fisher exact-test).

As a result, the median CI was higher for V1 than for LL neurons (blue vs. orange dashed line in Fig. 5D; $p = 8.09*10^{-10}$; one-tailed Mann-Whitney U-test), and the fraction of units with well-defined linear RFs (i.e., with CI > 5.5) was five times larger in V1 than in LL (Fig. 5E; $p = 6.48*10^{-10}$, χ^2 test, df = 1).

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 (see Fig. 5B and a full description in Materials and methods; also note that 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 Fig. 5F), resulting, in most cases, in a statistically significant difference between the two areas (Fisher's exact test; * $p < 0.05$; ** $p < 0.01$; see legend for exact p values).

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 a prevalently non-linear unit, the STA method would be able to reconstruct at most its linear sub-parts[46–48] – 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 fall short at accounting for the full richness and complexity of its structure and for the extension of its invariance field (see next paragraph). 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 (Fig. 6).

The binarized STA images could also be used to compare the size (diameter) of the well-defined linear RFs in the two areas (with the diameter computed as the mean of the major and minor axes of the ellipse that best fitted the area covered by the detected lobes). When a liberal binarization threshold of 3.0 was chosen (with the goal of capturing most of the RF extension), the average diameter in LL (55.4 ± 10º; n = 13) was significantly larger than in V1 (39.1 ± 3.0º; n = 54; p = 0.02; one-tailed, unpaired t-test, df = 65). However, such difference became smaller (34.2 ± 2.9º in V1 vs. 39.4 ± 5.1º in LL) and not significant (p = 0.21; one-tailed, unpaired t-test, df = 65), as soon as the binarization threshold was raised at 3.5.

This result should not be taken as evidence against the increase of RF size that is expected to take place along a ventral-like pathway. Such an increase has already been reported by several authors[14,15,49,50] and was carefully quantified in previous investigation of rat lateral extrastriate cortex[13] by our group, where the RF diameter in LL (~30º) was found to be twice as large as in V1 (~15º). These estimates were obtained using a mapping procedure in which small, high-contrast stimuli (specifically, drifting bars with multiple orientations) were presented over a grid of visual field locations. While such high-contrast, localized stimuli can effectively elicit responses over the full invariance field of a neuron, thus yielding reliable RF estimates also for highly position-tolerant (i.e., nonlinear) units, the linear STA analysis used in our current study cannot, by definition, achieve this goal. In fact,

for a highly tolerant neuron (such as a complex cell), a spike can be triggered both by a given distribution of luminance intensity values (a noise pattern) in a given portion of the visual field, and by its contrast reversal – which results in a failure of the STA method to map that portion of the RF. In other words, the nonlinear stimulus-response relationship that is typical of highly transformation-tolerant neurons prevents an accurate estimate of their RF size using STA.

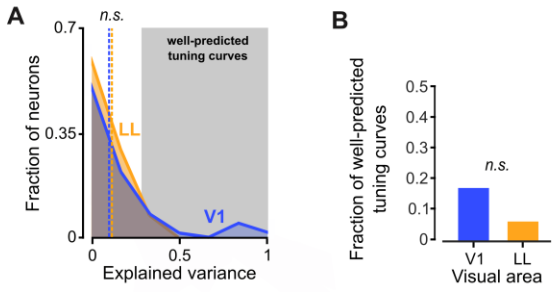

Figure 6. The V1 population includes neurons, whose direction-tuning curves are well predicted by linear LN models. A) Distributions of the variance of the direction-tuning curves that was explained by STA-based LN models for the populations of V1 (blue; n = 56) and LL (orange; n = 13) neurons with well-defined linear RFs (i.e., neurons with STA images having CI > 5.5; see Fig. 5D). Although the medians of the two distributions were not significantly different (p = 0.075; one-tailed Mann-Whitney U-test), a heavier tail can be noticed in the blue curve, which highlights the presence of a subset of units, in V1, whose selectivity was highly predictable using a linear, STA-estimated RF. B) Fraction of neurons, whose direction-tuning curves were well predicted by the STA-based LN models (i.e., units for which at least one third of the variance was explained by the models; gray patch in A) in V1 (blue) and LL (orange). Although twice as many well-predicted tuning curves were found in V1, as compared to LL, the difference was only marginally significant (p = 0.670; Fisher exact-test).

As a result, the RF size estimates yielded by STA in LL are largely underestimated, when compared to those obtained in V1, while the latter are consistent with the RF sizes that can be inferred from previous mouse studies, in which V1 receptive fields were also mapped using a linear reverse correlation method[32,51,52].

To conclude, the poor quality of the STA-based RFs obtained in LL (in terms of contrast, number of lobes and estimated size; Fig. 5), taken together with the prevalence of complex-like units in this area (Fig. 4A-B), is consistent with the non-linear computations that are expected to build invariance along an object-processing hierarchy.

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 Fig. 7A, in area LL most units had OSI lower than 0.5 (orange curve), with a median of 0.41 (dashed orange line), while in V1 the peak of the OSI distribution (blue curve) was in the 0.6-0.8 range, with a median of 0.53 (dashed blue line).

The difference between the two OSI distributions was significant (p = 0.0029; Kolmogorov-Smirnov test), as it was the difference between their medians (p = 0.0084; one-tailed, Mann-Whitney U-test). Following a convention established in previous studies (Table 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 (Fig. 7B). In V1, the proportion of selective cells was twice as large as in LL (40% vs. 18%), and such difference was highly significant ($p = 8.15*10^{-4}$; χ^2 test, df = 1).

Although the computation of the OSI index is very popular in studies of neuronal tuning for oriented gratings (see Table 2), this metric, being based on the average responses

of the neurons across repeated stimulus presentations, does not take into account the trial-by-trial variability of firing rates and, as such, does not provide a direct estimate of the information about orientation conveyed by neuronal firing.

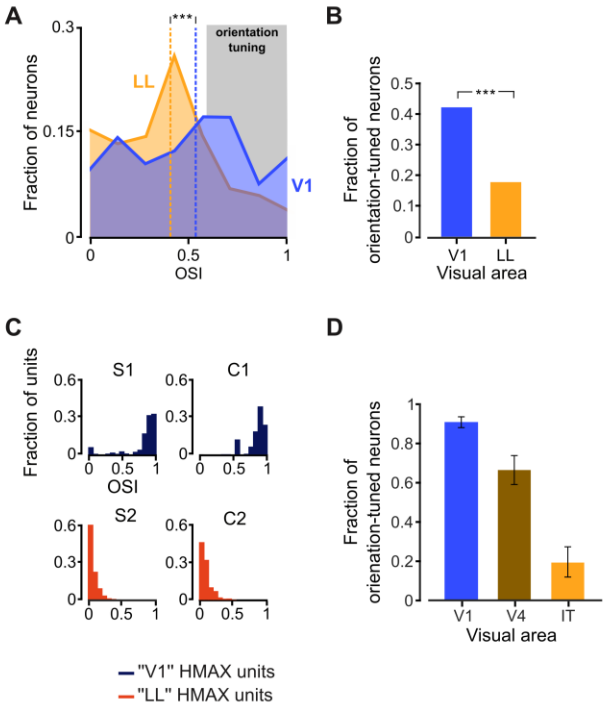

Figure 7. Orientation selectivity decreases from V1 to LL. A) Distributions of the orientation selectivity index (OSI) for the populations of V1 (blue; n = 105) and LL (orange; n = 104) neurons. Medians are shown as dashed lines (**p < 0.01; one-tailed Mann-Whitney U-test). The gray-shaded area highlights 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 (blue) and LL (orange; ***p < 0.001; $\chi 2$ test). C) Evolution of the OSI distribution across the layers of the HMAX model (Fig. 3). D) Fraction of cells that have been reported as being orientation tuned in areas V1 (blue; n = 7), V4 (brown; n = 4) and IT (orange; n = 4) of the monkey ventral stream (mean across n studies in each area ± SE; see Table 2). The drop of orientation tuning across the three areas was highly significant (p < 0.001, one-way ANOVA).

Therefore, we decided to complement the comparison of V1 and LL based on the OSI with an information theoretic analysis. For each neuron, we took its preferred direction (e.g., 60º) and the opposite one (e.g., 240º) and we assigned them to a stimulus category that we labeled as preferred orientation. We then took the two directions orthogonal to the preferred ones (e.g. 150° and 330°) and we assigned them to a category that we labeled as orthogonal orientation. We then pooled all the responses of the neuron to the stimuli within each category and we computed the mutual information I(R;S) between the neuronal response R (discretized into two equi-populated bins) and the stimulus category S (preferred vs. orthogonal). All quantities were computed using the Information Breakdown Toolbox[53] and were corrected for limited sampling bias using the Panzeri-Treves method[54,55]. We found that V1 neurons conveyed, on average, more than twice as much information about orientation than LL units (i.e., 0.054 ± 0.007 bits vs. 0.023 ± 0.003 bits) and this difference was highly significant ($p = 7.82*10^{-5}$; two-tailed, unpaired t-test, df = 207). Moreover, the amount of orientation information coded by the neurons within each population was positively and significantly correlated with OSI (r = 0.57, $p = 2.48*10^{-10}$, df = 103 and r = 0.65, p = 4.52*10-

[14], df = 102 for, respectively, V1 and LL). Overall, these results confirm the validity of the analysis based on the OSI metric, and show that, even when the signal-to-noise ratio of neuronal responses is taken into account, V1 neurons display a much higher degree of orientation selectivity, as compared to LL units.

To check the consistency of this finding with ventral-like processing, we looked at the evolution of orientation tuning in HMAX and we carried out another meta-analysis of the monkey literature. As shown in Fig. 7C, 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 Fig. 3). The shape of the OSI distribution was then maintained in the C2 layer, performing the max-pooling operation over the S2 afferents.

reference	area	number of units	fraction of tuned units	tuned units criterion
Schiller et al.1976	**V1**	654	**0.75**	BW < 90°
De Valois et al. 1982a	**V1**	343	**0.90**	BW < 90°
De Valois et al. 1982b	**V1**	222	**0.92**	BW < 90°
Vogels et al. 1991	**V1**	285	**0.85**	BW < 90°
Zhou et al. 2000	**V1**	63	**0.89**	OSI > 0.6
Ringach et al. 2002	**V1**	308	**1.00**	BW < 90°
David et al. 2006	**V1**	45	**0.90**	BW < 90°
DeSimone et al. 1987	**V4**	322	**070**	BW < 90°
Pasupathy et al. 1999	**V4**	152	**0.52**	orientation effect (ANOVA)
Zhou et al 2000	**V4**	45	**0.82**	OSI > 0.6
David et al. 2006	**V4**	87	**0.50**	BW < 90°
DeSimone et al. 1984	**IT**	66	**0.00**	qualitative
Vogels et al. 1993	**IT**	958	**0.13**	OSI > 0.7
Vogels et al. 1994	**IT**	1885	**0.34**	orientation effect (ANOVA)
Sary et al. 1995	**IT**	330	**0.27**	orientation effect (ANOVA)

Table 2. Full list of research papers included in the meta-analysis shown in Fig. 7D. First three columns: same description as in Table 1. The fourth column shows the fraction of orientation-tuned units reported in each article. The bars shown in Fig. 7D were obtained by averaging the values in this column across studies within each individual cortical area. The fifth column specifies the criterion adopted in each study to define orientation-selective units (BW stands for the bandwidth of the orientation tuning curve, while OSI is the same orientation selectivity index used in our study).

With regard to the monkey ventral stream (Table 2), most authors report a very large fraction of orientation-tuned cells in V1[19,33–35,38,56,57], with a mean across studies close to 89% (blue bar in Fig. 7D). This proportion significantly decreases along downstream areas ($p = 2.31*10^{-6}$, $F_{2,14} = 46.17$, one-way ANOVA), with ~65% and ~19% of orientation-tuned neurons reported, respectively, in V4[19,44,57,58] and IT[59–62] (brown and orange bars in Fig. 7D).

These trends are confirmed by the few studies that applied the same experimental design to directly compare orientation tuning in V1 and either V4[19] or IT[56,62]. 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 (i.e., responding strongly to two non-adjacent orientations) 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[42,43,58,63].

Overall, the trends observed in our HMAX simulations (Fig. 7C) and emerging from the ventral stream literature (Fig. 7D) suggest that 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 (Fig. 7A, B) can be taken as indication that LL neurons are tuned to more complex combinations of visual features, as compared to V1 units.

Support to this conclusion came from comparing the two areas in terms of the tendency of the orientation tuning curves to be bimodal (Fig. 8A). 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 (blue curve). In LL, the BSI distribution also peaked at zero, but featured a heavier tail, with a secondary peak above 0.5 (orange curve). As a result, the two distributions and their medians were significantly different ($p = 0.044$ and $p = 0.029$; 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 ($p = 0.027$; χ^2 test, df = 1; Fig. 8B). Once again, a similar trend was observed in HMAX (Fig. 8C), where the BSI distribution displayed a much heavier tail in the deeper layers (S2 and C2) than in the early ones (S1 and C1).

Quantitatively, the difference in the number of bimodal neurons found in V1 and LL (7/105 vs. 16/104 in LL) may appear small, but, besides being statistically significant, is not dissimilar to the one reported in the monkey[19], when comparing V1 and V4 (see previous paragraph). In fact, while the incidence of bimodal neurons is approximately twice as large in monkey as in rat visual areas (11% vs. 7% in monkey vs. rat V1; 28% vs. 15% in V4 vs. LL), the percent increase from monkey V1 to V4 (~254%) is very similar to the percent increase from rat V1 to LL (~231%). In addition, in both V1 and LL, bimodal neurons came from multiple animals (4/12 in V1 and 4/6 in LL) and multiple recording sessions (4/15 in V1 and 5/9 in LL), which, not surprisingly (given the low incidence of these neurons), accounted for a large fraction of all recorded, visually-driven single units in both areas (56/105 in V1 and 79/104 in LL).

Nevertheless, caution should be taken when comparing the BSI distributions obtained for two neuronal populations with a different level of orientation selectivity, because the BSI and OSI indexes are not constrained to be statistically independent. Intuitively, in presence of

some noise, poorly selective neurons (i.e., cells with low OSI) will tend to have slightly larger BSI values than sharply selective units, even if the underlying tuning curves are unimodal (i.e., even if they are ideal oriented-edge detectors without any true secondary peaks in their orientation tuning curve).

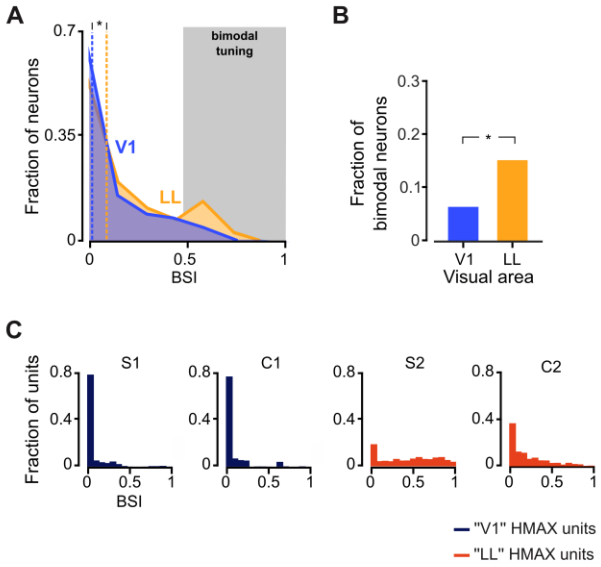

Figure 8. The tendency to be tuned for multiple orientations increases from V1 to LL. A) Distributions of the bimodal selectivity index (BSI) for the populations of V1 (blue; n = 105) and LL (orange; n = 104) neurons. Medians are shown as dashed lines (*p < 0.05; one-tailed Mann-Whitney U-test). The 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 (blue) and LL (orange *p < 0.05; χ^2 test). C) Evolution of the BSI distribution across the layers of the HMAX model (Fig. 3).

This is because, for a very broad tuning curve, the two largest, non-adjacent peaks will tend to be closer, in terms of magnitude, than for a narrow tuning curve. In our data, BSI and OSI were in fact negatively correlated (r = -0.48 and r = -0.4, respectively, in V1 and LL; p = $2.3*10^{-7}$, df = 103 and p = $1.1*10^{-5}$, df = 102; unpaired, two-tailed, t-test). Therefore, we performed an analysis to quantitatively check that the larger BSI values found in LL, as compared to V1 (Fig. 8A), were not simply a byproduct of the lower orientation tuning of the LL neurons (Fig. 7A). In other words, we verified that the larger fraction of units being classified as bimodal in LL, as compared to V1 (Fig. 8B), was not an artifact of the possibly larger proportion of LL neurons with very broad (but unimodal) orientation tuning.

Our analysis was based on simulating two artificial populations of perfectly unimodal units (see examples in Fig. 9A), having: i) the same size of the recorded V1 and LL populations (i.e., 105 and 104 simulated V1 and LL units, respectively); ii) OSI values sampled from the empirical OSI distributions that were observed in the recorded populations (dashed lines in Fig. 9B; same as the distributions shown in Fig. 7A); and iii) peak firing rates (FRs) sampled from log-normal functions (dashed lines in Fig. 9C) that were fitted to the empirical distributions of peak firing rates observed in the recorded populations (black bars in Fig. 9C). As a consequence of point iii) above, the average FRs of the two simulated populations were very close to those of the recorded populations (i.e., 11.4 ± 10.7 spikes/s in V1 and 6.5 ± 4.3 spikes/s in LL; see arrows in Fig. 9C).

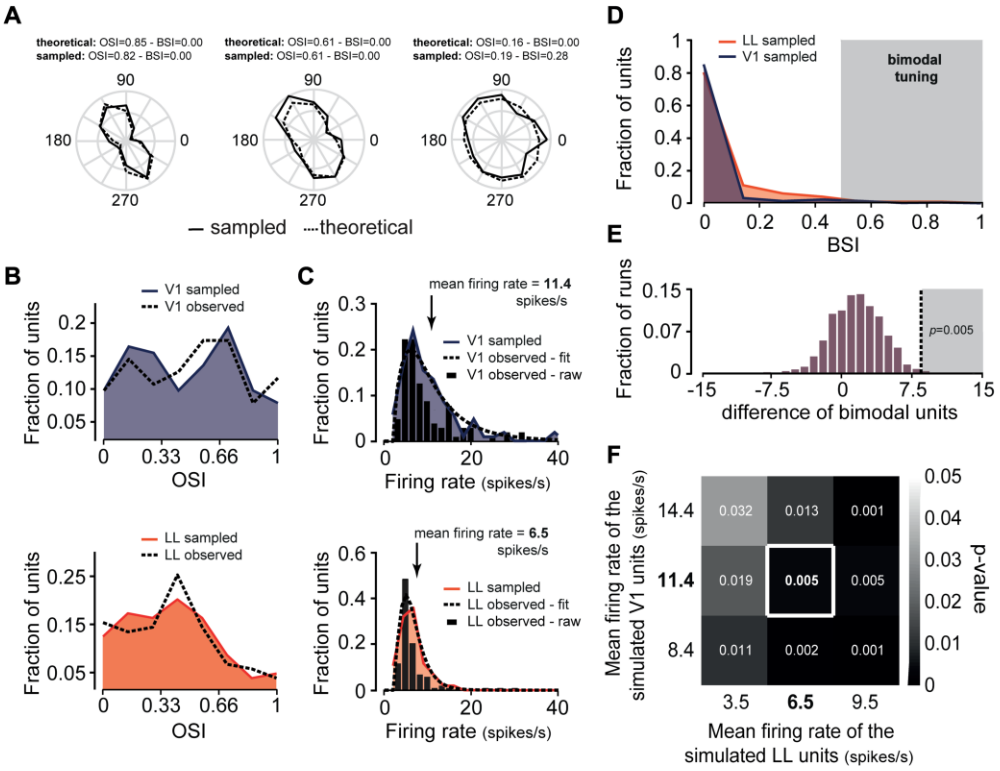

Figure 9. The increase of neurons with bimodal orientation tuning from V1 to LL is not accounted for by the broader orientation selectivity and lower firing rate of LL units. A) Examples of simulated units with unimodal orientation tuning curves, having a different degree of sharpness. Each tuning curve was obtained by summing two Von Mises functions, centered on the two opposite directions that were collinear with a given chosen orientation. The width of the Von Mises functions, which are the circular analogues of Gaussian functions, is controlled by the parameter k, which, in turn, is equivalent to the inverse of the variance of a Gaussian distribution. Starting from these theoretical tuning curves (dashed lines), we sampled 10 repeated Poisson responses for each orientation and we averaged these responses to obtain estimated tuning curves for the simulated units (solid lines). From these estimated tuning curves, we computed the BSI of the simulated units. B) The OSI of each simulated unit was sampled from theoretical distributions that were identical to those obtained for the recorded V1 and LL populations (black dashed lines; same as the distributions shown in Fig. 7A). Obviously, as a result of the Poisson noise in the simulated responses (see A), the sampled OSI distributions obtained in any given run of the simulation (colored curves) were slightly different from the theoretical ones (the colored lines/areas show such sampled distributions for two example simulated populations). C) The peak firing rates (FRs) of the simulated units were sampled from lognormal functions (black dashed lines) that were fitted to the FR distributions obtained for the recorded V1 and LL populations (black bars). Again, because of the Poisson noise, the sampled FR distributions obtained in any given run of the simulation were slightly different from the theoretical ones (the colored lines/areas show such sampled distributions for same simulated populations shown in B). D) Sampled BSI distributions obtained for the same two example populations of simulated V1 (105) and LL (104) units shown in B and C. Despite the tendency of the simulated LL units to have slightly larger BSI values than the simulated V1 units, the fraction of units classified as bimodal (i.e., with BSI > 0.5) was the same (3 units) for these example populations. E) Simulated BSI distributions as those shown in D were obtained in 1000 independent runs, thus yielding a null distribution of differences between the number of neurons classified as bimodal in the two populations (purple bars), under the hypothesis of purely unimodal tuning curves. Such null distribution was used to assess the significance of the measured difference of bimodal neurons between LL and V1 ($p = 0.005$; dashed line). F) The statistical test described in E was repeated for six different combinations of mean peak firing rates in the simulated V1 and LL populations. These six combinations included the mean FR values actually observed in the recorded populations (central square; same analysis shown in E). The additional combinations were obtained by either lowering or increasing the mean FR observed in V1 and LL by 3 spikes/s. For every combination, the difference of bimodal units between the two areas was significantly larger than what expected by chance ($p < 0.05$), under the null hypothesis of purely unimodal tuning curves (p values reported in the cells of the matrix; see also the color bar).

In addition, to account for the effect of estimating orientation-tuning curves from a limited number of repeated trials in noisy neurons, the responses of a simulated unit to a given orientation were drawn from a Poisson distribution with the mean equal to the theoretical value of the simulated tuning curve at that orientation (dashed lines in Fig. 9A).

Ten repeated Poisson responses were simulated for each orientation, and then averaged to obtain sampled tuning curves for the simulated units (solid lines in Fig. 9A). The goal of our simulation was to check if, as a consequence of the Poisson noise, these sampled tuning curves would develop spurious secondary peaks (especially in the case of low orientation tuning) that could lead to artifactual differences of bimodal neurons between LL and V1. It should be noticed that simulating noisy Poisson neurons, along with the fact that a limited number of units were drawn from the theoretical, data-matched OSI and FR distributions, brought the sampled OSI and FR distributions to be slightly different from the theoretical ones. Such differences, however, were minimal, as it can be appreciated by comparing the sampled and theoretical OSI and FR distributions obtained for two example simulated V1 and LL populations in Fig. 9B and C (colored lines/areas vs. black dashed lines). Fig. 9D reports instead the sampled BSI distributions that were obtained for these example simulated populations. As expected, (see previous paragraph), the larger orientation tuning of the simulated LL units did bring their BSI distribution (red curve) to have a slightly heavier tail, as compared to the BSI distribution of the simulated V1 units (blue curve). However, the difference between the two distributions was very minor, as compared to the one observed for the actual populations (compare to Fig. 8A). Most noticeably, the simulated LL distribution lacked the heavy tail, with the secondary peak at BSI > 0.5, that we found for the recorded LL neurons. As a result, the number of units classified as bimodal in these two simulated populations (i.e., with BSI > 0.5) was identical (3 units).

Given the stochastic nature of our analysis, this outcome does not guarantee that, in general, no major differences can be observed in the number of bimodal neurons found in the two simulated populations. Therefore, to test whether we could statistically reject the null hypothesis that the difference of bimodal units observed between the recorded LL and V1 populations was merely due to differences in terms of broadness of orientation tuning and firing rate magnitude (and, as such, response noisiness), we repeated the simulation illustrated above in 1000 independent runs, so as to obtain 1000 comparisons among BSI distributions of simulated LL and V1 units. In each run, we computed the difference between the numbers of neurons classified as bimodal in the two simulated populations. This yielded a null distribution of differences of bimodal units between LL and V1, under the null hypothesis of purely unimodal tuning in both areas (purple bars in Fig. 9E). When we compared the actual, measured difference of bimodal neurons recorded in LL and V1 (dashed line in Fig. 9E) to this null distribution, we found that the probability of getting such a difference (or a larger one) under the null hypothesis was $p = 0.005$. Therefore, we could statistically reject the hypothesis that the larger fraction of bimodal neurons observed in LL was only due to the lower level of orientation tuning observed in this area, as compared to V1, with a 0.005 significance level. Interestingly, this conclusion held true, even when we simulated LL and V1 populations with the same OSI distributions observed in our recordings, but with mean peak firing rates either 3 spikes/s smaller or 3 spikes/s larger than those measured in the two areas (Fig. 9F). In particular, we found that the difference of bimodal

neurons between LL and V1 was significantly larger than expected by chance, also when the statistical comparison was performed against a null distribution obtained from LL and V1 simulated units with peak firing, respectively, lower and larger than that observed in the recorded populations ($p = 0.032$; top-left cell in Fig. 9F).

That is, no matter how noisy the tuning curves of the simulated LL units were made with respect to those of the simulated V1 units, their tendency to display spurious secondary peaks was never so strong as to explain the difference of bimodal units observed in our recordings. Overall, this indicates that the majority of neurons classified as bimodal in LL were truly so, thus confirming that the different proportion of bimodal cells in LL and V1 (Fig. 8B) is a small, but robust effect.

Interestingly, the relatively small incidence of bimodal units in high-order areas is fully consistent with the HMAX simulations, where these units were ~51% in layer S2, but reduced to ~20% in layer C2 (Fig. 8C). This indicates that the max-pooling operation that builds invariance (yielding the C2 layer) partially counterbalances the template-matching computation that builds tuning for multiple features (yielding the S2 layer), when it comes to produce bimodal tuning. Taken together, these considerations support 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.

This conclusion was confirmed by the overall lower effectiveness of the grating stimuli to drive LL neurons, as compared to V1 units (see Fig. 9C). By contrast, only a minor difference was observed in terms of response latencies (102 ± 11 ms in V1 vs 118 ± 9 ms in LL; see Materials and methods). This is in disagreement with previous studies reporting lower latencies in both areas and larger differences between V1 (~40 ms) and LL (~75 ms)[13,14], as expected along a partially feedforward processing hierarchy. Such discrepancy is explained by the fact that full-field drifting gratings, as opposed to the localized, high-contrast visual shapes used in[13,14], are not suitable to properly estimate response latencies. In fact, as shown by the example neurons of Fig. 2, the onset of the neuronal response is determined by time of the drift cycle in which the grating happens to properly align with the RF of the neuron. This, in turn, depends on the initial phase of the grating relative to the neuron's RF (which varied randomly from neuron to neuron in our experiment, being fixed relative to the stimulus display) and on its spatial and temporal frequencies. As a result, response latencies measured with drifting gratings are highly variable – they largely overestimate the actual time at which neurons in a given area start processing visual inputs, and do not allow appreciating differences among areas.

Finally, we also measured the direction selectivity of the two neuronal populations, finding no significant differences in terms of DSI (0.42 in V1 vs. 0.45 in LL; $p = 0.7$; Mann-Whitney U-test), which suggests a lack of specialization of LL for motion processing, as compared to V1.

As an extension of our analysis, we turned to the inclusion in the analysis of the small population of laterointermediate (LI) neurons unwantedly recorded during some of the LL-targeted sessions. The reason why in our original plans we restricted our recordings to LL area is that previous work by our group highlighted such region as bearing the clearest signature of ventral processing within rat lateral extrastriate cortex[13] – i.e., the lowest

sensitivity to low-level image attributes (such as luminance and contrast) and the most invariant coding of object identity. However, a similar advantage, over V1, to support invariant object recognition was also displayed to a lesser extent by area LI, which borders LL medially. As already mentioned, our current study was not designed to obtain a statistical characterization of LI in terms of tuning for drifting gratings. Nevertheless, we recorded a small number of single units also from this region, since, while aiming at LL, a fraction of the recording sites of the silicon probe frequently landed in LI (see Fig. 1). We therefore looked at the sensitivity for the phase of drifting gratings and at the sharpness and shape of the orientation-tuning curves also for the recorded LI neurons (a total of 23 responsive and reproducibly driven units).

As shown in Fig. 10 (leftmost panel), the fraction of phase-modulated, simple-like neurons in LI (brown bar) was intermediate between V1 and LL (blue and orange bars). The same applied to the fraction of orientation-selective units (middle panel), while the proportion of neurons with bimodal tuning (rightmost panel) was similar in LI and LL. We did not quantify statistically these trends, given the low number of cells sampled from LI. Nevertheless, we report them here, because they are consistent with our earlier finding that the processing leading to the specialization of LL for ventral computations takes place gradually, along a processing chain involving multiple areas, of which LI is one of the nodes[13].

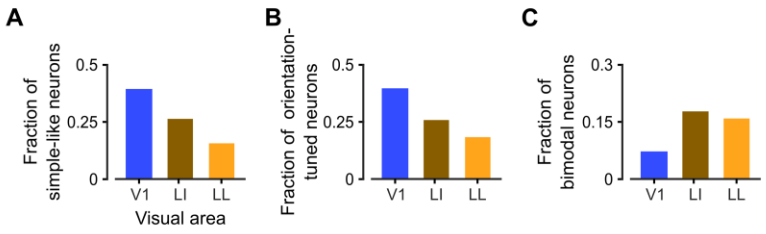

Figure 10. The level of specialization for ventral processing of LI is intermediate between V1 and LL. The fraction of LI neurons (brown bar) with simple-like behavior (left), sharp orientation tuning (middle) and bimodal tuning (right) is shown along the fractions of units displaying such properties in V1 and LL (blue and orange bars, respectively; same bars already shown Fig. 4B, 7B and 8B). No statistical comparison was performed, given the low sample of LI units available (23 responsive and reproducibly driven units).

As shown in the previous sections, most trends of variation observed along rat lateral extrastriate areas, in terms of tuning for drifting gratings, are in qualitative agreement with the prediction of HMAX. Such agreement is not surprising, given the selectivity-building and tolerance-building computations implemented by the units of the model. At the same time, the trends observed in HMAX are not trivial and should still be considered as emergent properties of the model, because, when such computations are implemented separately by thousands of independent units across a cascade of feedforward layers, the outcome of their interaction is not easily predictable. For instance, it is far from obvious that the template-matching computation that increases the complexity of tuning of S2 units will still allow them to show some selectivity for oriented gratings (Fig. 7C), and that such selectivity (also in layer C2) will be as phase-invariant as the one found (by construction) in layer C1 (Fig. 4C). Also, it was interesting to notice how the incidence of bimodal units was boosted (as expected) by the selectivity-building computation in layer S2, but it was dampened by the

tolerance-building pooling performed by C2 units (Fig. 8C), so that, in this layer, the fraction of bimodal units was not far from the one we observed in LL (see previous paragraph).

HMAX, however, is only one of the possible feedforward neuronal networks with a brain-inspired architecture that can serve as a comparison for physiology/behavioral experiments. In particular, HMAX can be considered as one of the ancestors of modern deep convolutional neuronal networks (DCNNs), which, in recent years, have revolutionized the field of machine learning, matching (and even surpassing) human accuracy in a number of object recognition tasks[64]. DCNNs share the same basic architecture of HMAX, with alternating layers of complexification of feature selectivity and tolerance-increasing pooling, but they are much larger, being made of tens of layers, each with hundreds of thousands units (for a total of several millions units). More importantly, while HMAX has a static, hard-wired connectivity (Materials and methods), the units in DCNNs iteratively adjust their synaptic weights, in such a way to learn feature representations that maximize their accuracy in a given image classification task[64]. Therefore, the emergent properties of DCNNs (e.g., in terms of shape selectivity) are less determined by architectural/computational constraints, as compared to HMAX, and more strongly driven by the optimization process underlying the development of powerful representations for object recognition. This has led several neuroscientists to use such networks as benchmarks to study the emergence of selective, yet invariant, representations of visual objects in the ventral stream of both monkeys[65–67] and humans[68–71]. Inspired by these primate studies, we repeated the same analysis carried out in HMAX also in VGG16, a state-of-the-art DCNN that scored second place in object classification and first place in object localization at the ILSVRC-2014 competition[17].

VGG16 is a quite large DCNN, totaling about 138 million parameters across its 16 convolutional and fully-connected layers (Fig. 11A). These layers are grouped in 5 blocks (named B1 to B5 in Fig. 11A), each composed by 2 or 3 convolutional layers (colored boxes) and followed by a max pooling layer (white boxes), with a final stack of 3 fully-connected layers on top, followed by a 1000-way softmax output layer for classification. Thanks to this relatively simple and uniform structure, VGG16 is particularly appealing as a model of how modern DCNNs work.

For our tests, we used VGG16 pre-trained on the ImageNet dataset (a large photographic database very popular as a computer vision benchmark), and, as done with HMAX, we fed to the input layer of the network a set of drifting gratings spanning the same range of directions used in the neurophysiology experiment (Materials and methods). We then computed the activation (over the entire duration of each presented grating) of a pool of 1000 randomly selected units from the first convolutional layer of each block (Fig. 11A). For each unit, we measured the time evolution of its activation, so as to estimate its sensitivity to the phase of the drifting gratings through the MI index, and we obtained orientation-tuning curves, from which we computed the OSI and BSI indexes.

Fig. 11B shows how the distributions of these indexes evolved across the five sampled layers of the network. In the case of the modulation index (top row), we observed a progressive shift from large (> 7.0) to low (< 3.0) MI values, resulting in a monotonic decrease in the mean level of phase modulation along the network (Fig. 11C, leftmost panel). This trend was very similar to the one displayed by HMAX (Fig. 4C), and was consistent with the drop of simple-like cells observed along the primate ventral stream (Fig. 4D) and, in

our experiments, from rat V1 to LL (Fig. 4A-B and Fig. 10, leftmost panel). This confirms that the reduction of phase-modulated units can be taken as a strong marker of the existence of nonlinear, tolerance-building computations along rat lateral extrastriate cortex, similar to the max pooling implemented both in HMAX and VGG16.

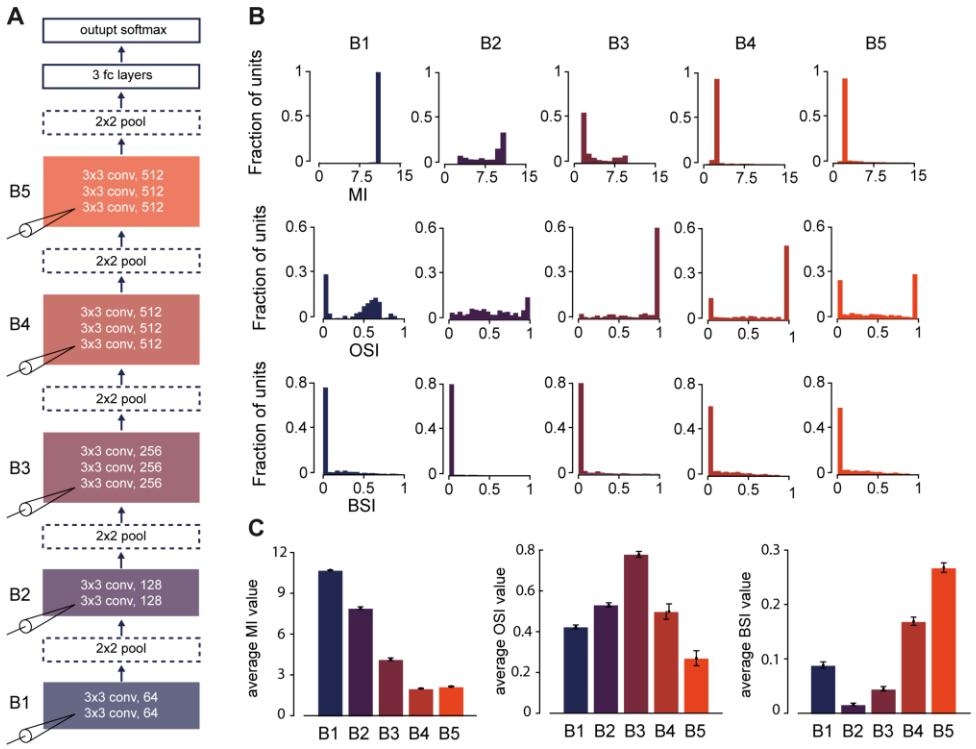

Figure 11. Evolution of the tuning for drifting gratings across the layers of a deep convolutional network. A) Sketch of the architecture of VGG16, a feedforward, deep convolutional network made of five blocks of 2 or 3 convolutional layers, followed by a max pooling layer, with a stack of three additional, fully-connected layers on top, before the softmax output layer for classification. In the drawing, the five blocks of convolutional layers (named B1 to B5) are color coded (from blue to red) to emphasize the progression from lower- (blue) to higher-order (red) representations of visual features implemented by the network. Each block also reports the size and number of the convolutional filters contained in each layer. Similar figures are reported for the pooling layers and for the upper, fully-connected layers (white frames). The weights of the network were those originally obtained by training it to correctly classify the object categories of a large image dataset (ImageNet). In our tests, the input layer of the network was fed with drifting gratings spanning different orientations, spatial frequencies and temporal frequencies. The tuning for key processing properties (MI, OSI and BSI) was then measured for a sample of 1000 randomly selected units from the first convolutional layer of each block (as highlighted by the picture of a probe shown over each sampled layer). B) Distribution of the MI (top), OSI (middle) and BSI (bottom) indexes across the five selected convolutional layers of VGG16 (same color code as in A). C) Average values (mean ± SE) of the MI (left), OSI (middle) and BSI (right) indexes over the pools of 1000 units sampled from of each of the five selected convolutional layers of the network (same color code as in A).

The orientation selectivity index displayed instead a non-monotonic trend (Fig. 11B, middle row and Fig. 11C, middle panel), since it first increased, reaching a peak in B3, the third block of convolutional layers, and then decreased from B3 to B5. This behavior is qualitatively different from that observed in HMAX (Fig. 7C) and may seem inconsistent with the monotonic decrease of OSI reported across the monkey ventral stream (Fig. 7D) and observed, in our study, along rat lateral extrastriate cortex (Fig. 7A-B and Fig. 10, middle panel). Such discrepancy, however, is easily explainable by the fact that in VGG16, differently from HMAX, the first processing layer does not contain sharply-tuned oriented

Gabors (of tens by tens pixels in size), but a first set of very small convolutional filters (of only 3x3 pixels). Having, by construction, such a small receptive field size, these units cannot learn the Gabor-like filters with high aspect ratio that are necessary to achieve sharp orientation tuning. Only the units in the following convolutional layers, as the effective receptive field size increases, can gradually evolve into sharply tuned, oriented-edge detectors. In other words, while HMAX, by design, is meant to simulate the processing that takes place from primary visual cortex onward, VGG16 could be regarded as a model of the entire visual system, with the initial layers, operating on the raw pixel representation of the image, being architecturally and functionally closer to subcortical visual areas (such as retina and LGN) than to the cortex.

In the light of these considerations, it is tempting to interpret the initial increase of OSI in VGG16 as the attempt of the network to first learn a bank of oriented-tuned units. Once this V1-like representation is established (in B3), the network displays a behavior very similar to HMAX and to monkey (rat) temporal (lateral) visual cortical areas, with OSI that gradually and monotonically decreases, as multiple peaks start appearing in the orientation tuning curves – a sign that the units of the upper layers are learning more complex feature representations. This was confirmed by the trend observed for BSI (Fig. 11B, bottom row and Fig. 11C, rightmost panel), which initially decreased (from B1 to B2), but then monotonically increased, with the fraction of units with bimodal tuning reaching ~25% in B5, a value not dissimilar from the one found in HMAX (see above). It is possible that DCNNs based on slightly different architectures, such as Alexnet[72], where the initial layer contains larger convolutional filters, may display a behavior that is even more consistent with HMAX, and show a strictly monotonic decrease of OSI and increase of BSI across consecutive layers (we did not test this possibility, since a systematic comparison of different DCNN architectures is clearly beyond the scope of our study).

Overall, we can conclude that all the trends we observed along rat lateral extrastriate cortex are in qualitative agreement not only with the predictions of the conceptual, hard-wired model of ventral processing implemented in HMAX, but also with the behavior displayed by the upper layers of a state-of-the-art, deep convolutional network trained for image classification. This adds further support to our conclusion that rat lateral visual cortical areas are specialized for processing object information.

Discussion

In this study, we sought to validate and further explore the specialization of rat area LL for ventral computations, as revealed by previous investigation of this region and nearby visual cortical areas by our group[13]. Our work was motivated by the still limited and conflicting evidence about the involvement of these areas in advanced shape processing, which makes their assignment to a ventral-like stream much weaker, at the functional level, than one would expect on anatomical grounds[73–78]. For instance, while our group observed an increase in the ability of rat lateral areas to support transformation-tolerant recognition[13], Vermaercke et al.[14] found that only TO (an area located laterally to LL) was superior to V1 and the other areas in supporting position tolerance, and only in relative terms – i.e., in terms of the stability, rather

than of the magnitude, of the discrimination performance afforded by TO across two nearby positions. Vermaercke et al.[14] also found that orientation tuning increased along the areas' progression – a trend that, as shown in our current study, is in disagreement with both the ventral stream literature and the behavior of ventral-stream models. In a later study by Vinken et al., the same group reported a lack of object categorical representations in rat extrastriate lateral areas, which is unexpected for an object-processing stream[15]. At the same time, they found a decrease, along the lateral progression, in the accuracy of V1-like models to predict neuronal responses, as expected for a ventral-like pathway.

The literature probing the tuning properties of mouse extrastriate areas features a similar variety of findings. In mice, it is unclear whether an equivalent of rat area LL exists – earlier anatomical maps reported such area[79], although more recent studies would place the cortical field lateral to LI either in the anterior part of postrhinal cortex (PORa; also named posterior area 36) or in the posterior part of temporal cortex (TEp)[80–82]. Regardless of the naming convention, this region has been attributed to the ventral stream on anatomical grounds[81,83], along with areas LM (lateromedial) and LI, which have been tested with drifting gratings in several electrophysiological and imaging studies. Van den Bergh et al.[9] measured the phase dependence of neuronal responses in V1 and nearby lateral cortex (named V2 by the authors, which, very likely, corresponds to area LM) and found that the latter had a lower proportion of simple cells (15% vs. 38% in V1). Conversely, the authors reported virtually 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.[8], 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.[11], while other authors found virtually no differences between V1 and some of the putative dorsal- and ventral-stream areas[7,10,84].

The variety and inconsistency of these findings motivated our further search for the signature of ventral processing in LL, where, to be fully complementary with the previous studies from our group, we did not probe neuronal representations using visual objects. With the goal of targeting those core tuning processes that are responsible for building up transformation tolerance and feature selectivity, we deployed drifting gratings and noise movies, and we compared the V1 and LL neuronal populations in terms of: i) the phase-sensitivity of their responses; ii) the goodness of the RF structures recovered through STA; and iii) the sharpness and complexity (i.e., bimodality) of their tuning for orientation. This approach required deriving clear predictions about how these properties should evolve along a ventral-like pathway. This was achieved through a careful survey of the monkey literature and by simulating the basic computational architecture of the ventral stream using a conceptual model of ventral computations (HMAX) and a state-or-the-art deep convolutional network (VGG16). Because of the invariance-building operation, phase-sensitivity sharply decreased across HMAX layers (Fig. 4C). Concomitantly, because of the selectivity-building computation, orientation tuning became smaller (Fig. 7C), with the units acquiring preference for multiple orientations (Fig. 8C). The same drop of phase-sensitivity was observed in VGG16, where orientation tuning also decreased sharply (with a concomitant increase of bimodal tuning) in the last layers of the network (Fig. 11B-C). All these trends were matched by the empirical evidence gathered in our study, when comparing V1 to LL (Figs. 4A-B, 7A-

B and 8A-B), thus suggesting that similar selectivity- and invariance-building computations are at work along the progression of lateral extrastriate areas that culminates with LL. The consistency among these trends and those observed across the primate ventral stream (Figs. 4D and 7D) adds further, critical support to this conclusion.

As such, our results strengthen the previous attribution of LL to a ventral-like stream[13], while shedding new light on the possible origin of the discrepancies with and among earlier rodent studies (see previous paragraphs). For instance, while we presented very large (60ºx110º), full-field gratings, Vermaercke et al.[14] used smaller (33º) circular gratings, shown at the RF center of each recorded neuron. 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[85–89], where surround-suppression can strongly affect shape tuning[90,91]. By contrast, the smaller and more localized gratings used in Vermaercke et al.[14] likely covered an increasingly smaller fraction of both the classical and extra-classical receptive fields while progressing from V1 to LL, given that RF size gradually increases along lateral extrastriate areas, being twice as large in LL as in V1[13]. 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.

This interpretation is only partially supported by the monkey literature. In fact, although surround-modulation has been shown to play a major role in boosting the selectivity of monkey visual neurons[92], most primate studies reviewed in Table 2 and Fig. 7D used circular gratings with a 6-12º diameter to probe orientation tuning, rather than full-field stimuli. Still, the size of those stimuli, relative to the typical RF size of ventral stream areas (i.e., median of 2º and 10º, respectively in V1 and IT), was slightly larger than the size of the circular gratings used by Vermaercke et al.[14], relative the median RF size of rat V1 and LL neurons (15º and 30º respectively), thus possibly engaging stronger extra-classical processing in monkey recordings. Alternatively, it is possible that in rats, even more than in monkeys, the signature of higher-order processing of shape information may clearly emerge only when surround-modulation mechanisms are fully engaged. Support to this conclusion comes from Vinken et al.[15], who, as previously mentioned, found a reduction in the ability of V1-like models (based on combinations of oriented Gabor filters) to predict neuronal responses to natural movies along the lateral progression. Not only this finding is in agreement with the poor estimates of the RF structures that we obtained in LL, as compared to V1, using STA, but strongly suggests that a reduction of orientation tuning has to be expected from V1 to downstream lateral areas (not tested in that study). That is, the conclusions of Vinken et al.[15], obtained with full-field stimuli (size 50º-74º), appear to be in better agreement with the drop of OSI reported in our study than with the increase of OSI previously found by the same group, using smaller, circular gratings[14]. To summarize, the discrepancies among the tuning properties of rat lateral neurons appear to be largely accounted for by differences in terms of visual stimuli, as argued here, as well as experimental design and data analysis, as discussed in Tafazoli et al.[13].

Similar arguments can be applied to interpret the results of mouse experiments, but only for the fraction of studies that used localized, circular gratings[7,84] instead of full-field stimuli[8,9,11]. In the case of mouse studies, however, a more fundamental methodological

difference exists with our and previous rat studies. In mice, many recordings were based on calcium imaging, performed using different indicators (e.g., OGB-1[8] vs. GCaMP3[7]), which diverge from a perfectly linear relationship between fluorescence changes and underlying spiking activity in an-indicator specific way[93]. As pointed out by Niell[94], 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 microelectrodes. In addition, while our recordings targeted cortical layer 5 (Fig. 1C), optical imaging studies typically target superficial layers. Although the previous study from our group found the same proportional increase of ventral-specific processing from V1 to LL in both superficial and deep layers, the magnitude of this difference was smaller in the former[13]. This could make between-area differences harder to detect in superficial layers, especially when precise spike count estimates are not available, as in such imaging studies. This argument does not apply to the experiments of Van den Bergh et al.[9], who performed single-electrode recordings. However, that study targeted only area LM, which has been reported to be functionally very similar to V1, never displaying any sign of ventral-like functional specialization[7,8,13]. Finally, the possibility cannot be excluded that major differences exist, between mouse and rat visual cortex, in terms of high-order processing of shape information. In fact, despite some recent efforts[95,96], mouse visual perception is still largely unexplored, and it remains unknown whether mice are able to perform those complex perceptual tasks, recently demonstrated in rats[97–103], that should specifically engage the ventral stream.

To conclude, we believe that our results nicely complement and extend those of the previous study by our group[13]. In fact, they show how some key neuronal tuning properties, which are revealing 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 – an ability that, in rodents, is likely essential for ecologically relevant tasks such as spatial navigation and foraging[104,105]. In addition, being our experimental approach based on the presentation of widely-used parametric stimuli and the application of intuitive metrics, it provides a standardized and easily interpretable way to carry out a large-scale screening of ventral processing across rodent visual cortical areas. This could help new 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[106]. In this regard, having provided clear predictions about the evolution of neuronal tuning along an object-processing pathway will likely serve as a reference to guide future explorations of ventral functions in both mice and rats.

Materials and methods

Animal preparation and surgery

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). 18 naïve Long-Evans male rats (Charles River Laboratories), with age 3-12 months and weight 300-700 grams, underwent extracellular recordings in either primary visual cortex (V1) or laterolateral (LL) extrastriate visual cortex. Each rat was anesthetized with an intraperitoneal (IP) injection of a solution of 0.3 mg/kg of fentanyl (Fentanest®, Pfizer) and 0.3 mg/kg of medetomidin (Domitor®, Orion Pharma). The level of anesthesia was monitored by checking the absence of tail, ear and hind paw reflex, as well as monitoring blood oxygenation, heart and respiratory rate through a pulse oximeter (Pulsesense-VET, Nonin). A constant flow of oxygen was delivered to the rat throughout the experiment to prevent hypoxia. A constant level of anesthesia was maintained through continuous IP infusion of the same anesthetic solution used for induction, but at a lower concentration (0.1 mg/kg/h Fentanyl and 0.1 g/kg/h Medetomidin), by means of a syringe pump (NE-500; New Era Pump Systems). Temperature was thermostatically kept at 37°C using a heating pad to prevent anesthesia-induced hypothermia.

After induction, the rat was secured to a stereotaxic apparatus (Narishige, SR-5R) in flat-skull orientation (i.e., with the surface of the skull parallel to the base of the stereotax) and, following a scalp incision, a craniotomy was performed over the target area in the left hemisphere (typically, a 2x2 mm^2 window) and the dura was removed to allow the insertion of the electrode array. When targeting V1, the coordinates of penetration were ~6.5 mm posterior from bregma and ~4.5 mm left to the sagittal suture (i.e., AP 6.5, ML 4.5). When targeting LL, the coordinates were ~1 mm anterior from lambda and ~1 mm medial from the cranial ridge (i.e., AP 8, ML 5). Throughout the procedure, the rat eyes were protected from direct light and kept hydrated by repeated application of an ophthalmic ointment (Epigel®, Ceva Vetem).

Once the surgical procedure was completed, and prior to probe insertion, the stereotax was placed on a rotating platform and the rat's left eye was covered with black, opaque tape, while the right eye (placed at 30 cm distance from the monitor) was immobilized using a metal eye-ring anchored to the stereotax. The platform was then rotated, so as to align the right eye with the center of the stimulus display and bring the binocular portion of its visual field to cover the left side of the display. During the recordings, the eye and cortex were periodically irrigated using saline solution to keep them properly hydrated.

Neuronal recordings

Extracellular recordings were performed using single-shank 32-channel silicon probes (NeuroNexus Technologies) with site recording area of 775 μm^2 and 25 μm of inter-site spacing. After grounding (by wiring the probe to the animal's head skin), the electrode was manually lowered into the cortical tissue using an oil hydraulic micromanipulator (Narishige, MO-10; typical insertion speed: ~ 5 μm/s), up to the chosen insertion depth (~800-1000 μm from the cortical surface when targeting V1, and ~2500 μm when targeting LL). To reach LL, the probe was tilted of 30° relative to the vertical to the surface of the skull (i.e., relative to the vertical to the base of the stereotax; Fig. 1A), whereas, for V1 recordings, it was inserted either perpendicularly (about half of the sessions) or with a variable tilt, between 10º and 30º (remaining half of the sessions). Extracellular signals were acquired using a system three

workstation (Tucker-Davis Technologies) with a sampling rate of 25 kHz. Before insertion, the probe was coated with Vybrant® DiI cell-labeling solution (Invitrogen, Oregon, USA), to allow visualizing the probe insertion track post-mortem, through histological procedures. To this aim, at the end of the recording session, an electrolytic lesion was also performed by delivering current (5 μA for 2 seconds) through the 4 deepest channels at the tip of the shank.

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 (i.e., ensure that most RFs fell inside the monitor, by rotating the platform or repositioning the eye through adjustments of the eye-ring). During the RF mapping protocol, the animal was presented with 10°-long drifting bars with four different orientations (0º, 45º, 90º and 135º), tiling a grid of 66 different visual field locations (i.e., 6 rows, spanning vertically 50°, and 11 columns, spanning horizontally 100°). The bars were white over a black background. During presentation of these stimuli, multiunit spiking responses (MUA) were plotted as a function of screen position in real-time, so as to estimate the positions of the RF centers across the recording sites of the probe. This allowed identifying in real-time when area LL was reached during an oblique probe insertion. Specifically, area LL was identified by the third "rightwards" (i.e., from nasal to temporal) reversal of the retinotopy at a recording depth close to 2.5 mm DV[13]. The same analysis was applied off-line to well-isolated single units obtained by spike sorting (see below) to precisely identify the neurons that were recorded from LL, as illustrated in Fig. 1B.

Once the probe was positioned in the final recording location, the main presentation protocol was administered. The rat was presented with: i) 1s-long drifting gratings, made of all possible combinations of 3 spatial frequencies (SF; 0.02, 0.04 and 0.08 cycle/°), 3 temporal frequencies (TF; 2, 4 and 8 Hz), and 12 directions (from 0° to 330°, in 30° increments); and ii) 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 1 s-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 cycle/° 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® 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.

Histology

At the end of the recording session, each animal was deeply anesthetized with an overdose of urethane (1.5 gr/kg) and perfused transcardially with phosphate buffer saline (PBS) 0.1 M, followed by 4% paraformaldehyde (PFA) in PBS 0.1 M, pH 7.4. The brain was then removed from the skull, post-fixed in 4% PFA for 24 h at 4°C, and then immersed in cryoprotectant solution (15% w/v sucrose in PBS 0.1 M, then 30% w/v sucrose in PBS 0.1 M) for 48 h at 4 °C.

The brain was finally sectioned into 30 μm-thick coronal slices using a freezing microtome (Leica SM2000R, Nussloch, Germany). Sections were mounted immediately on Superfrost Plus slides and let dry at room temperature overnight. A brief wash in distilled water was performed, to remove the excess of crystal salt sedimented on the slices, before inspecting them at the microscope. Each slice was then photographed with a digital camera adapted to a Leica microscope (Leica DM6000B-CTR6000, Nussloch, Germany), acquiring both a DiI fluorescence image (700 nm DiI filter) and a bright-field image at 2.5X and 10X magnification.

Following the acquisition of this set of images, the sections displaying the electrode fluorescent track were further stained for Nissl substance using a Cresyl Violet Acetate solution, and new pictures were taken at 2.5X and 10X magnification. By superimposing the fluorescence, bright-field and Nissl-stained images, it was possible to reconstruct the tilt and the anteroposterior (AP) position of the probe during the recording session, as well as the laminar location of the recording sites. Specifically, the boundaries between the cortical layers were identified, based on the difference in size, morphology and density of the Nissl-labeled cells across the cortical thickness.

The position of the probe relative to such boundaries was determined by tracing the outline of the fluorescent track, and taking into account, when available, the location of the electrolytic lesion performed at the end of the recording session. Based on the known geometry of the silicon probe, it was possible to draw the location of each recording site over the shank, thus estimating its laminar location.

Selection of the single units included in the analyses

Responses of single units were isolated offline by applying the spike sorting package KlustaKwik-Phy[107]. 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: i) 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); ii) the shape of the auto- and cross-correlograms (important to decide whether to merge two clusters or not); iii) the variation, over time, of the principal component coefficients of the waveform (important to detect and take into account possible electrode drifts); and iv) 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 for at least 6 trials 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):

$$\mathrm{OSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{ortho}}\right)/\left(R_{\mathrm{pref}} + R_{\mathrm{ortho}}\right) \qquad \mathrm{DSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{opposite}}\right)/\left(R_{\mathrm{pref}} + R_{\mathrm{opposite}}\right)$$

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.[23] 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, 10 ms-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[19,20]:

$$\mathrm{BSI} = \left(R_{\text{2nd peak}} - R_{\text{2nd trough}}\right) / \left(R_{\text{1st peak}} - R_{\text{1st trough}}\right)$$

where $R_{\text{1st peak}}$ and $R_{\text{2nd peak}}$ indicate the two largest peaks in the orientation-tuning curve (with $R_{\text{1st peak}} > R_{\text{2nd peak}}$), and $R_{\text{1st trough}}$ and $R_{\text{2nd trough}}$ indicate the deepest troughs (with $R_{\text{1st trough}} < R_{\text{2nd trough}}$). This index takes values close to one for tuning curves displaying two equally high peaks (regardless of their distance along the orientation tuning curve) and close to 0 for tuning curves with a single peak. Since this index is meant to provide a first-order assessment of the tuning for multiple oriented elements, it does not distinguish between cases where the orientation-tuning curve has two peaks from cases where it has more than two (i.e., the index discriminates curves with only one peak from curves with multiple peaks).

All the above metrics were computed using the responses of each neuron to the gratings shown at the most effective combination of spatial and temporal frequencies. For the vast majority of the recorded units in both areas (i.e., 97% and 86% in, respectively, V1 and LL), the preferred SFs were the lowest we tested (i.e., 0.02 and 0.04 cycles/°). Similarly, the preferred TFs were the lowest we tested (i.e., 2 and 4 Hz) for most neurons in both V1 (97%) and LL (88%).

Analysis of the responses to the noise movies

We used the spike-triggered average (STA) analysis[24,25] 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 Fig. 2). 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[24,25], using Tikhonov regularization to handle covariance matrix inversion. Statistical significance of the STA images was then assessed pixelwise, by applying the following 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 (Fig. 2 and 4). The temporal span of the spatiotemporal linear kernel reconstructed via STA extended from 330 ms before spike generation to the time of spike generation, corresponding to a duration of

10 frames of the noise movie played at 30 Hz (Fig. 2). These procedures were performed on downsampled noise frames (16x32 pixels), with the resulting STA images that were later spline interpolated at higher resolution for better visualization and to obtain predictions of the neuronal responses using LN models (see next paragraph).

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[24,25]. 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 (black vs. colored curves in Fig. 2.i) was quantified by computing the fraction of variance explained by the model prediction. 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 (black vs. colored curves in Fig. 2.ii, bottom), 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 (Fig. 5A, middle plot), we obtained a distribution of contrast intensity values (Fig. 5A, 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 Fig. 2), 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 Fig. 5C). It should be noted that, since the intensity values of the original STA images 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 σ units of the z-scored STA values. The 5.5 value was chosen in such a way to be conservative enough to include only STA images with a good signal-to-noise-ratio, but liberal enough to allow a fraction of LL neurons (whose STA images had typically poor contrast) to be considered for further analysis. This allowed a meaningful statistical comparison between V1 and LL in terms of lobe count (see next paragraph).

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[108] to enhance the high-intensity regions within the RF and reduce noise-driven false lobe detections. We then took the absolute value of the resulting image and we

set a threshold over its z-score values (ranging from 3 to 6 units of standard deviations; Fig. 5B). 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 the 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 (e.g., see the yellow dot in Fig. 5B), especially in the case of RFs with low CI. The total number of accepted centroids at the end of this selection procedure is the lobe count reported in Fig. 5F for the different binarization thresholds.

HMAX simulations

To obtain predictions about the evolution of the tuning properties of visual neurons along an object-processing pathway, we simulated ventral stream functional architecture using the HMAX model[16]. The structure of the model has been described at length in the Results (Fig. 3) and further motivated in the Discussion. Here, we provide some technical details about its specific implementation.

For our application, we have chosen the version of the model described in [27] and downloaded from the website http://maxlab.neuro.georgetown.edu. Briefly, the model is a feedforward neural network with alternating layers of units performing: i) 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 Fig. 3); or ii) 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 = 1000) in each layer, which was enough to obtain smooth distributions for the relevant indexes. The resulting distributions are those shown in Fig. 4C, 7C and 8C.

VGG16 simulations

We further checked whether the processing expected to take place along an object-processing hierarchy is consistent with the trends observed in rat visual cortex, by measuring the tuning for drifting gratins along the layers of VGG16, a state-of-the-art deep convolutional neuronal network (DCNN) for image classification[17]. VGG16 is a large DCNN, totaling about 138 million parameters across its 16 convolutional and fully-connected layers. More specifically, the network consists in 5 blocks (colored boxes in Fig. 11A), made of 2 or 3 convolutional layers, with each block followed by a max-pooling layer (white boxes in Fig. 11A), and a final stack of three additional, fully-connected layers on top, before the softmax output layer for classification.

Detailed explanations of the roles the convolutional and pooling layers can be found elsewhere[64]. Briefly, the convolutional layer is an architectural prior that allows exploiting the structure of natural visual inputs (in particular, the spatial localization of visual objects and the ubiquity of translation as a commonly occurring identity-preserving transformation) to reduce the number of free parameters. Thanks to the structure of the visual world, a feature detector that is useful in one part of the visual field will likely be useful also in another part (because of the above-mentioned translation of visual features, caused by physical movements of objects and/or sensors). This allows "sharing" filter parameters across the visual field, by learning a single spatially localized filter in a given position and then applying it iteratively over the whole span of each input image. This also endows convolutional layers with spatial sparsity of connections: each unit in a convolutional feature map receives information only from a small, localized, subset of units in the previous layer (largely cutting down the number of parameters, as compared to a fully connected network). The pooling layer is another architectural prior that consists in performing a downsampling operation over the input (usually via a "max" computation, equivalent to the one implemented by HMAX), thus hardwiring in the output some amount of translation tolerance. This, again, can be seen as a way of leveraging on the prior knowledge that naturally occurring translation of visual objects preserve their identity, in order to shrink the width of the feature maps (thus reducing the number of parameters needed in the subsequent layers) and, at the same time, helping the network to build robust, invariant representations.

For our tests, we used VGG16 pre-trained as it was for the ILSVRC-2014 competition[17]. The weights were downloaded from http://www.robots.ox.ac.uk/~vgg (Chollet, 2015, GitHub, https://github.com/fchollet/keras) via Keras interface. Those weights were originally obtained by training the network on the ImageNet dataset – a large photographic database very popular as a computer vision benchmark (including over 14 million hand-annotated pictures with labels indicating which objects are pictured in each image). As done with HMAX, we fed to the input layer of the network a set of drifting

gratings, resulting from combining 30 spatial frequencies, 4 temporal frequencies and 24 different directions. We then computed the activation (over the entire duration of each presented grating) of a pool of 1000 randomly selected units from the first convolutional layer of each block, from which we measured the relevant tuning properties of the units, to obtain the statistical characterization shown in Fig. 11B and C.

Experimental design and statistical analysis

To decide how many neurons to record in each of the two areas (V1 and LL) under investigation, we took inspiration from previous studies of rodent and monkey visual cortex (see Tables 1 and 2). In most of these studies, a good statistical assessment of the tuning properties of visual neurons (e.g., orientation tuning and phase sensitivity) was based on a sample size of > 100 units, although a few studies used less (~50 or fewer) and some used more (> 200). We thus set as a target for our study to collect at least 100 very well-isolated single-units (SUs) per area. Since the number of SUs obtained from each recording session was highly unpredictable, we performed multiple recording sessions from the same or different animals until our target number was reached in both areas. This required performing neuronal recordings from V1 of 12 different rats (for a total of 15 recording sessions) and from area LL of 6 different rats (for a total of 9 recording sessions), yielding a total of 105 and 104 responsive and reproducibly driven SUs in, respectively, V1 and LL. In V1, 5 sessions yielded more than 10 SUs, 4 sessions yielded between 5 and 10 SUs, while the remaining 6 sessions yielded less than 5 SUs. In LL, 5 sessions yielded more than 10 SUs, 3 sessions yielded between 5 and 10 SUs, while the last session yielded less than 5 SUs. When considered in terms of neurons yielded by each animal, the mean number of units recorded per rat was 8.75 in V1 and 17.33 in LL, while the minimal and maximal numbers of units per rat were, respectively, 1 and 29 in V1, and 3 and 45 in LL.

Throughout the study, differences between median values of two distributions were quantified using a Mann-Whitney U-test. Differences in the fraction 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 1 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 Fig. 5F). To check whether the distributions of the relevant indexes (i.e., MI, OSI, and BSI) were different between V1 and LL we applied a Kolmogorov-Smirnov test. To compare the average RF size extracted from the STA images, we applied an unpaired, one-tailed t-test with 65 degrees of freedom. To compare the information conveyed by neuronal responses about stimulus orientation in V1 and LL, we applied an unpaired, two-tailed, t-test with 207 degrees of freedom. To assess the significance of the correlation between the information about stimulus orientation and OSI, we used paired, two-tailed, t-tests with 103 (in V1) and 102 (in LL) degrees of freedom after applying Fisher transformation. Similarly, when testing for correlation between BSI and OSI, we adopted paired, two-tailed, t-tests with 103 (in V1) and 102 (in LL) degrees of freedom after applying Fisher transformation. Finally, in the meta-analysis shown in Fig. 4D and 7D, we statistically compared the mean fractions of modulated (Fig. 4D) and orientation-tuned (Fig. 7D) units across primate visual cortical areas V1, V4 and IT using a one-way ANOVA with, respectively, 13 and 14 degrees of freedom.

References

1. Maunsell, J. H., Nealey, T. A. & DePriest, D. D. Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J. Neurosci.* **10**, 3323–3334 (1990).
2. Ferrera, V. P., Nealey, T. A. & Maunsell, J. H. Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *J. Neurosci.* **14**, 2080–2088 (1994).
3. Glickfeld, L. L., Andermann, M. L., Bonin, V. & Reid, R. C. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nat. Neurosci.* **16**, 219–226 (2013).
4. Gao, E., DeAngelis, G. C. & Burkhalter, A. Parallel Input Channels to Mouse Primary Visual Cortex. *J. Neurosci.* **30**, 5912–5926 (2010).
5. Ji, W. *et al.* Modularity in the Organization of Mouse Primary Visual Cortex. *Neuron* **87**, 632–643 (2015).
6. Matsui, T. & Ohki, K. Target dependence of orientation and direction selectivity of corticocortical projection neurons in the mouse V1. *Front. Neural Circuits* **7**, (2013).
7. Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional Specialization of Mouse Higher Visual Cortical Areas. *Neuron* **72**, 1025–1039 (2011).
8. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional Specialization of Seven Mouse Visual Cortical Areas. *Neuron* **72**, 1040–1054 (2011).
9. Van den Bergh, G., Zhang, B., Arckens, L. & Chino, Y. M. Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. *J. Comp. Neurol.* **518**, 2051–2070 (2010).
10. Roth, M. M., Helmchen, F. & Kampa, B. M. Distinct Functional Properties of Primary and Posteromedial Visual Area of Mouse Neocortex. *J. Neurosci.* **32**, 9716–9726 (2012).
11. Tohmi, M., Meguro, R., Tsukano, H., Hishida, R. & Shibuki, K. The Extrageniculate Visual Pathway Generates Distinct Response Properties in the Higher Visual Areas of Mice. *Curr. Biol.* **24**, 587–597 (2014).
12. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How Does the Brain Solve Visual Object Recognition? *Neuron* **73**, 415–434 (2012).
13. Tafazoli, S. *et al.* Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6**, e22794 (2017).
14. Vermaercke, B. *et al.* Functional specialization in rat occipital and temporal visual cortex. *J. Neurophysiol.* jn.00737.2013 (2014) doi:10.1152/jn.00737.2013.
15. Vinken, K., Van den Bergh, G., Vermaercke, B., Beeck, O. de & P, H. Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. *Cereb. Cortex* **26**, 3310–3322 (2016).
16. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat Neurosci* **2**, 1019–25 (1999).
17. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs* (2014).

18. Glickfeld, L. L. & Olsen, S. R. Higher-Order Areas of the Mouse Visual Cortex. *Annu. Rev. Vis. Sci.* **3**, 251–273 (2017).
19. David, S. V., Hayden, B. Y. & Gallant, J. L. Spectral Receptive Field Properties Explain Shape Selectivity in Area V4. *J. Neurophysiol.* **96**, 3492–3505 (2006).
20. Cadieu, C. *et al.* A model of V4 shape selectivity and invariance. *J Neurophysiol* **98**, 1733–50 (2007).
21. Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol.* **283**, 79–99 (1978).
22. Skottun, B. C. *et al.* Classifying simple and complex cells on the basis of response modulation. *Vision Res.* **31**, 1078–1086 (1991).
23. Wypych, M. *et al.* Standardized F1 – A consistent measure of strength of modulation of visual responses to sine-wave drifting gratings. *Vision Res.* **72**, 14–33 (2012).
24. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *J. Vis.* **6**, (2006).
25. Sharpee, T. O. Computational Identification of Receptive Fields. *Annu. Rev. Neurosci.* **36**, 103–120 (2013).
26. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J Physiol* **195**, 215–43 (1968).
27. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. & Poggio, T. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* **29**, 411–26 (2007).
28. Zoccolan, D., Kouh, M., Poggio, T. & Dicarlo, J. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* **27**, 12292–307 (2007).
29. Leibo, J. Z., Liao, Q., Anselmi, F., Freiwald, W. A. & Poggio, T. View-Tolerant Face Recognition and Hebbian Learning Imply Mirror-Symmetric Neural Tuning to Head Orientation. *Curr. Biol.* **27**, 62–67 (2017).
30. Cox, P. H. & Riesenhuber, M. There Is a "U" in Clutter: Evidence for Robust Sparse Codes Underlying Clutter Tolerance in Human Vision. *J. Neurosci.* **35**, 14148–14159 (2015).
31. Girman, S. V., Sauvé, Y. & Lund, R. D. Receptive Field Properties of Single Neurons in Rat Primary Visual Cortex. *J. Neurophysiol.* **82**, 301–311 (1999).
32. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J Neurosci* **28**, 7520–36 (2008).
33. Schiller, P. H., Finlay, B. L. & Volman, S. F. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.* **39**, 1288–1319 (1976).
34. De Valois, R. L., Albrecht, D. G. & Thorell, L. G. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559 (1982).
35. De Valois, R. L., William Yund, E. & Hepler, N. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 531–544 (1982).
36. Foster K H, Gaska J P, Nagler M & Pollen D A. Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J. Physiol.* **365**, 331–363 (1985).

37. O'Keefe, L. P., Levitt, J. B., Kiper, D. C., Shapley, R. M. & Movshon, J. A. Functional Organization of Owl Monkey Lateral Geniculate Nucleus and Visual Cortex. *J. Neurophysiol.* **80**, 594–609 (1998).
38. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. *J. Neurosci.* **22**, 5639–5651 (2002).
39. Kagan, I., Gur, M. & Snodderly, D. M. Spatial Organization of Receptive Fields of V1 Neurons of Alert Monkeys: Comparison With Responses to Gratings. *J. Neurophysiol.* **88**, 2557–2574 (2002).
40. Aronov, D., Reich, D. S., Mechler, F. & Victor, J. D. Neural Coding of Spatial Phase in V1 of the Macaque Monkey. *J. Neurophysiol.* **89**, 3304–3327 (2003).
41. Gur, M., Kagan, I. & Snodderly, D. M. Orientation and Direction Selectivity of Neurons in V1 of Alert Monkeys: Functional Relationships and Laminar Distributions. *Cereb. Cortex* **15**, 1207–1221 (2005).
42. Gallant, J. L., Braun, J. & Essen, D. V. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259**, 100–103 (1993).
43. Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W. & Van Essen, D. C. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* **76**, 2718–2739 (1996).
44. Desimone, R. & Schein, S. J. Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J. Neurophysiol.* **57**, 835–868 (1987).
45. Pollen, D. A., Nagler, M., Daugman, J., Kronauer, R. & Cavanagh, P. Use of gabor elementary functions to probe receptive field substructure of posterior inferotemporal neurons in the owl monkey. *Vision Res.* **24**, 233–241 (1984).
46. Sedigh-Sarvestani, M., Fernandez-Lamo, I., Jaegle, A. & Taylor, M. M. 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 (2014).
47. Fournier, J., Monier, C., Pananceau, M. & Frégnac, Y. Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. *Nat. Neurosci.* **14**, 1053–1060 (2011).
48. Fournier, J. *et al.* Hidden Complexity of Synaptic Receptive Fields in Cat V1. *J. Neurosci.* **34**, 5515–5528 (2014).
49. Montero, V. M. Comparative Studies on the Visual Cortex. in *Multiple Visual Areas* 33–81 (Humana Press, 1981). doi:10.1007/978-1-4612-5814-8_2.
50. Espinoza, S. G. & Thomas, H. C. Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. *Brain Res* **272**, 137–44 (1983).
51. Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local Diversity and Fine-Scale Organization of Receptive Fields in Mouse Visual Cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
52. Hoy, J. L. & Niell, C. M. Layer-Specific Refinement of Visual Cortex Function after Eye Opening in the Awake Mouse. *J. Neurosci.* **35**, 3370–3383 (2015).
53. Magri, C., Whittingstall, K., Singh, V., Logothetis, N. K. & Panzeri, S. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC Neurosci.* **10**, 81 (2009).

54. Panzeri, S. & Treves, A. Analytical estimates of limited sampling biases in different information measures. *Network* **7**, 87–107 (1996).
55. Panzeri, S., Senatore, R., Montemurro, M. A. & Petersen, R. S. Correcting for the Sampling Bias Problem in Spike Train Information Measures. *J. Neurophysiol.* **98**, 1064–1072 (2007).
56. Vogels, R. & Orban, G. A. Quantitative study of striate single unit responses in monkeys performing an orientation discrimination task. *Exp. Brain Res.* **84**, 1–11 (1991).
57. Zhou, H., Friedman, H. S. & Heydt, R. von der. Coding of Border Ownership in Monkey Visual Cortex. *J. Neurosci.* **20**, 6594–6611 (2000).
58. Pasupathy, A. & Connor, C. E. Responses to Contour Features in Macaque Area V4. *J. Neurophysiol.* **82**, 2490–2502 (1999).
59. Sary, G., Vogels, R., Kovacs, G. & Orban, G. A. Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. *J. Neurophysiol.* **73**, 1341–1354 (1995).
60. Vogels, R. & Orban, G. A. Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *J. Neurophysiol.* **71**, 1428–1451 (1994).
61. Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* **4**, 2051–62 (1984).
62. Vogels, R. & Orban, G. A. Responses of monkey infero-temporal units in an orientation discrimination task. in *Progress in Brain Research* (eds. Hicks, T. P., Molotchnikoff, S. & Ono, T.) vol. 95 427–444 (Elsevier, 1993).
63. Pasupathy, A. & Connor, C. E. Population coding of shape in area V4. *Nat. Neurosci.* **5**, 1332–1338 (2002).
64. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
65. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci.* **111**, 8619–8624 (2014).
66. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365 (2016).
67. Hong, H., Yamins, D. L. K., Majaj, N. J. & DiCarlo, J. J. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* **19**, 613–622 (2016).
68. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLOS Comput. Biol.* **10**, e1003915 (2014).
69. Kriegeskorte, N. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annu. Rev. Vis. Sci.* **1**, 417–446 (2015).
70. Kubilius, J., Bracci, S. & Beeck, H. P. O. de. Deep Neural Networks as a Computational Model for Human Shape Sensitivity. *PLOS Comput. Biol.* **12**, e1004896 (2016).
71. Kheradpisheh, S. R., Ghodrati, M., Ganjtabesh, M. & Masquelier, T. Deep Networks Can Resemble Human Feed-forward Vision in Invariant Object Recognition. *Sci. Rep.* **6**, 32672 (2016).

72. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).
73. Gallardo, L. *et al.* Failure by rats to learn a visual conditional discrimination after lateral peristriate cortical lesions. *Physiol. Psychol.* **7**, 173–177 (1979).
74. McDaniel, W. F., Coleman, J. & Lindsay Jr., J. F. A comparison of lateral peristriate and striate neocortical ablations in the rat. *Behav. Brain Res.* **6**, 249–272 (1982).
75. Wörtwein, G., Mogensen, J., Williams, G., Carlos, J. & Divac, I. Cortical area in the rat that mediates visual pattern discrimination. *Acta Neurobiol. Exp. (Warsz.)* **54**, 365–376 (1993).
76. Aggleton, J. P., Keen, S., Warburton, E. C. & Bussey, T. J. 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 (1997).
77. Sánchez, R. F. *et al.* Visuospatial Discrimination Deficit in Rats after Ibotenate Lesions in Anteromedial Visual Cortex. *Physiol. Behav.* **62**, 989–994 (1997).
78. Tees, R. C. The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. *Behav. Brain Res.* **106**, 55–73 (1999).
79. Olavarria, J. & Montero, V. M. Organization of visual cortex in the mouse revealed by correlating callosal and striate-extrastriate connections. *Vis. Neurosci.* **3**, 59–69 (1989).
80. Wang, Q. & Burkhalter, A. Area map of mouse visual cortex. *J. Comp. Neurol.* **502**, 339–357 (2007).
81. Wang, Q., Sporns, O. & Burkhalter, A. Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J. Neurosci.* **32**, 4386–4399 (2012).
82. Gămănuţ, R. *et al.* The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles. *Neuron* **97**, 698-715.e10 (2018).
83. Wang, Q., Gao, E. & Burkhalter, A. Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. *J. Neurosci.* **31**, 1905–1918 (2011).
84. Smith, I. T., Townsend, L. B., Huh, R., Zhu, H. & Smith, S. L. Stream-dependent development of higher visual cortical areas. *Nat. Neurosci.* **20**, 200 (2017).
85. Vaiceliunaite, A., Erisken, S., Franzen, F., Katzner, S. & Busse, L. Spatial integration in mouse primary visual cortex. *J. Neurophysiol.* **110**, 964–972 (2013).
86. Self, M. W. *et al.* Orientation-Tuned Surround Suppression in Mouse Visual Cortex. *J. Neurosci.* **34**, 9290–9304 (2014).
87. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J. & Scanziani, M. A neural circuit for spatial summation in visual cortex. *Nature* **490**, 226–231 (2012).
88. Pecka, M., Han, Y., Sader, E. & Mrsic-Flogel, T. D. Experience-Dependent Specialization of Receptive Field Surround for Selective Coding of Natural Scenes. *Neuron* **84**, 457–469 (2014).
89. Alwis, D. S., Richards, K. L. & Price, N. S. C. Masking reduces orientation selectivity in rat visual cortex. *J. Neurophysiol.* **116**, 2331–2341 (2016).

90. Maunsell, J. H. R. & Newsome, and W. T. Visual Processing in Monkey Extrastriate Cortex. *Annu. Rev. Neurosci.* **10**, 363–401 (1987).
91. Orban, G. A. Higher Order Visual Processing in Macaque Extrastriate Cortex. *Physiol. Rev.* **88**, 59 (2008).
92. Vinje, W. E. & Gallant, J. L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–6 (2000).
93. Akerboom, J. *et al.* Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *J. Neurosci.* **32**, 13819–13840 (2012).
94. Niell, C. M. Exploring the Next Frontier of Mouse Vision. *Neuron* **72**, 889–892 (2011).
95. Aoki, R., Tsubota, T., Goya, Y. & Benucci, A. An automated platform for high-throughput mouse behavior and physiology with voluntary head-fixation. *Nat. Commun.* **8**, 1196 (2017).
96. Yu, Y. *et al.* Mice use robust and common strategies to discriminate natural scenes. *Sci. Rep.* **8**, 1379 (2018).
97. Zoccolan, D., Oertelt, N., DiCarlo, J. J. & Cox, D. D. A rodent model for the study of invariant visual object recognition. *Proc Natl Acad Sci U A* **106**, 8748–53 (2009).
98. Tafazoli, S., Di Filippo, A. & Zoccolan, D. Transformation-Tolerant Object Recognition in Rats Revealed by Visual Priming. *J. Neurosci.* **32**, 21–34 (2012).
99. Alemi-Neissi, A., Rosselli, F. B. & Zoccolan, D. Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition. *J. Neurosci.* **33**, 5939–5956 (2013).
100. Vinken, K., Vermaercke, B. & Op de Beeck, H. P. Visual Categorization of Natural Movies by Rats. *J. Neurosci.* **34**, 10645–10658 (2014).
101. Rosselli, F. B., Alemi, A., Ansuini, A. & Zoccolan, D. Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. *Front. Neural Circuits* **9**, 10 (2015).
102. De Keyser, R., Bossens, C., Kubilius, J. & Beeck, H. P. O. de. Cue-invariant shape recognition in rats as tested with second-order contours. *J. Vis.* **15**, 14–14 (2015).
103. Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H. & Zoccolan, D. Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy. *Curr. Biol.* **28**, 1005-1015.e5 (2018).
104. Cox, D. D. Do we understand high-level vision? *Curr. Opin. Neurobiol.* **25**, 187–193 (2014).
105. Zoccolan, D. Invariant visual object recognition and shape processing in rats. *Behav. Brain Res.* **285**, 10–33 (2015).
106. Hawrylycz, M. *et al.* Inferring cortical function in the mouse visual system through large-scale systems neuroscience. *Proc. Natl. Acad. Sci.* **113**, 7337–7344 (2016).
107. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).
108. Perona, P. & Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 629–639 (1990).

Chapter II

A template-matching algorithm for laminar identification of cortical recording sites from evoked response potentials

Abstract

In recent years, the advent of the so-called silicon probes has made it possible to homogeneously sample spikes and local field potentials (LFPs) from a regular grid of cortical recording sites. In principle, this allows inferring the laminar location of the sites based on the spatiotemporal pattern of LFPs recorded along the probe, as in the well-known current source-density (CSD) analysis. This approach, however, has several limitations, since it usually relies on visual identification of landmark features (i.e., current sinks and sources) in the CSD pattern that can be easily missed, if the probe does not span the whole cortical thickness. To overcome these limitations, we developed an alternative approach, based on finding the optimal match between the LFPs recorded along a probe in a given experiment and a template LFP profile that was computed by merging 18 recording sessions, in which the depth of the recording sites had been recovered through histology. We show that our method can achieve an accuracy of 79 μm in recovering the cortical depth of the recording sites, thus providing an alternative to CSD that, being fully automated, is free of subjective judgments and works reliably also for recordings spanning a limited cortical stretch.

Introduction

Most neuronal circuits in the mammalian brain are characterized by a complex spatial organization that is tightly intertwined with their function. In particular, in the cortex, laminar structure is closely linked to the flow of information among different neuronal populations[1,2]. Therefore, to fully understand the principles of operation of cortical circuits, it is essential to analyze the activity of single neurons in their spatial context.

Historically, extracellular recordings have been the workhorse method for studying cortical functions and they are still largely used in system neuroscience experiments, although neurophysiologists are increasingly replacing the traditional single electrode approach with multielectrode arrays. The latter allow isolating the waveforms of the action potentials (a.k.a. spikes) fired by many single neurons at once with sub-millisecond temporal precision[3] – with the number of recorded units ranging from a few tens to many hundreds, depending on the shape, geometry, technology and materials used to build the array[4,5]. However, such powerful experimental approach has a main limitation: its spatial "blindness". In fact, looking at the spikes detected on a given channel of an extracellular electrode gives no direct information about the spatial (e.g., laminar) location of the source neurons. Thus, it is not surprising that the challenge of coordinating multielectrode recordings with anatomical information has been recognized in the literature as a fundamental one[6,7]. In particular, the most basic anatomical metadata needed to fully make sense of extracellular cortical recordings is the laminar identity of the recorded single units (i.e., the cortical layers in which the recorded neurons sit).

This problem has been partially addressed by the advent of the so-called silicon (or laminar) probes – fork-shaped silicon substrates with several shanks, along which multiple recording sites are placed with a regular spacing[8,9]. These arrays allow spanning homogeneously the whole cortical thickness or a part of it, thus recording simultaneously spiking signals and "local field potentials" (LFPs) from neurons located in multiple cortical layers. This makes it possible, in principle, to infer the laminar location of the recording sites based on the spatio-temporal pattern of LFPs recorded along the probe, without resorting to laborious and time-consuming histological procedures.

Throughout the years, a technique known as "current source-density" (CSD) analysis has been widely exploited to fulfill this goal in different animal models, from monkeys[10–13] to rodents[14–21]. CSD exploits the LFP gradients recorded across geometrically-arranged electrode arrays to estimate the extracellular current flow in the surrounding tissue[22]. This is done by computing the second spatial derivative of the LFPs recorded along the axis perpendicular to the laminar structure of the cortex. In fact, this quantity is directly proportional to the extracellular current, if one assumes the tissue as being composed of a stack of two-dimensional isopotential planes. The resulting pattern of current sinks and sources across the cortical depth, mainly linked to local synaptic activity[6], can thus be used to infer the laminar location of the recording sites by visual inspection. More specifically, a key landmark used in this process is the prominent current sink associated with thalamic afferent inputs impinging into layer 4 (L4) of primary sensory cortices. Such characteristic CSD feature is usually associated with an inversion of the polarity of the stimulus-evoked LFP waveforms – i.e. the evoked response potentials (ERP), a.k.a. visually evoked potentials (VEP) in the case of visual stimulation.

Despite its widespread use, laminar identification through CSD has several drawbacks and limitations. First, the identification of the layers is usually carried out through visual inspection of the CSD pattern. Such reliance on the subjective judgment of the investigator implies a lack of objectivity and standardization, the unknown precision of the inference, and the slowness and laboriousness of the process. Second, although the CSD is meant to enhance the spatial resolution in the localization of the signal's source, as compared to raw LFPs, this

comes at the expense of being more sensitive to imperfections of the electrode array (e.g., unwanted variations of channels impedance along the electrode array that may occur with re-use of the probe or due to fabrication defects). Third, and more importantly, to infer with reasonable confidence the laminar position of the recording sites, it is essential to span a large fraction of the cortical depth, so as to observe the landmark sink in L4. This can be a serious limitation, when silicon probes with tightly packed sites (e.g., 25 μm inter-site spacing) are inserted into the cortex (often with a tilt), so as to densely sample neuronal populations from a specific supragranular or infragranular layer, given that the L4 sink will not be observable in such cases.

The approach presented in this study was developed to overcome these limitations and perform an automated laminar identification of the recording sites along a cortical silicon probe. This was achieved by finding the optimal match between the VEPs recorded across the probe in a given experiment and a template VEP profile, spanning the whole cortical thickness, that was computed by merging several recording sessions, in which the ground-true depth and laminar location of the recording sites were recovered through histology. As a result, our method was able to achieve, without the need of any subjective human judgment, a cross-validated accuracy of 79 μm in recovering the cortical depth of the recording sites and a 72% accuracy in returning their laminar position.

Results

Our automated method for inferring the cortical depth and laminar location of the recording sites of a silicon probe is based on three key steps. First, we had to build a "template VEP profile" along the cortical thickness, by averaging the waveforms of the VEPs recorded at the same cortical depth across multiple, repeated experimental sessions employing the same (or similar) visual stimulation protocol. Second, we had to establish a map between cortical depth and laminar location in the primary visual cortex (V1) of the animal model used to demonstrate our method – the Long-Evans rat. Third, we had to find the optimal match between the template VEP profile and candidate spatial arrangements of the VEPs recorded from a given probe in a given session, so as to infer the cortical depth of the recording sites of the probe and, through the depth-to-layer map, their laminar location.

This required obtaining a rich dataset of VEPs, coupled with the histological depth localization of the electrodes from which they were recorded. This was achieved by merging the V1 recordings collected in [23], for which the histological analysis yielded the most accurate estimates (i.e., 10 recording sessions performed in 8 different rats), with 8 additional V1 recording sessions obtained from 5 rats. Both datasets consisted of extracellular recordings performed in anesthetized, adult, Long Evans male rats, passively exposed to full-contrast, sinewave drifting gratings of different orientations, spatial and temporal frequencies, each lasting 1 s and presented in 20 repeated trials (see Materials and method). The electrode arrays used to perform these recordings were single-shanks Neuronexus® silicon probes, with either 32 or 64 recording sites and 25 μm inter-site spacing. Raw voltage traces were acquired at 24 kHz sampling rate and later downsampled to 610 Hz after lowpass filtering, to obtain LFPs (see Materials and methods).

Let's now delve into the details of the first step required by our method: building the template VEP profile and the depth-to-layer map.

In order to obtain the average waveforms needed to build the template VEP profile, 168 ms-long segments (corresponding to 275 samples) were extracted from the LFP traces that were recorded in response to all the presentations of the drifting gratings, starting from the onset of each stimulus. For any given recording site, these 168 ms-long VEPs were averaged across all presented trials, directions, spatial and temporal frequencies, so as to obtain a very smooth average VEP. The different temporal frequency of the gratings used to collect the two datasets was not an issue, because extracting the first 168 ms of each VEP restricted the analysis to the initial transient of the stimulus-evoked LFP deflection, which is largely independent from the grating's frequency.

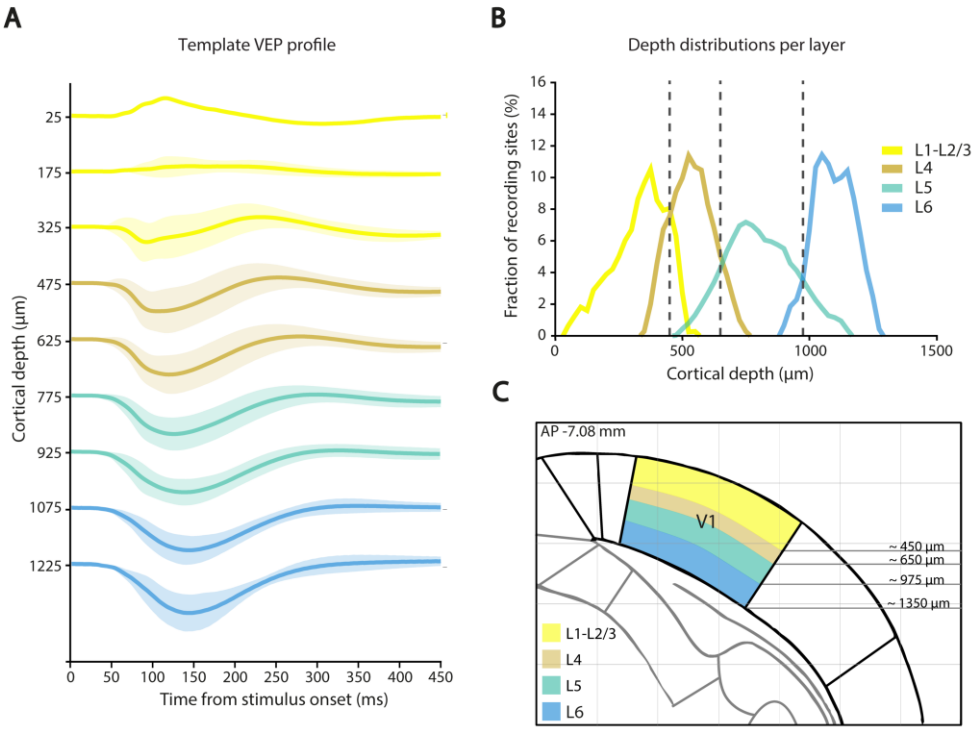

Figure 1. Template VEP profile and depth-to-layer map. A) The average waveforms (thick lines) ± SEM (shaded areas) of the VEPs recorded across all the 18 sessions of our experiment are plotted as a function of the cortical depth. Every waveform is the average of all the VEPs falling within a given 150 µm-wide cortical span. The colors label the cortical layers to which the waveforms belong, according to the map established in B-C (see key in B). B) Each line shows the empirical probability that a recording site within a given cortical layer was located at a given cortical depth (with both the depth and layer attribution recovered from the Nissl-stained brain slices obtained for the 18 recording sessions; see example in Fig. 3C). The dashed lines indicate the optimal boundaries between pairs of adjacent layers, as defined by taking the depths at which the corresponding distributions intersected. C) The span of each cortical layer or group of cortical layers (as defined in B) is color coded (see key) and superimposed to the outline of a coronal section of the rat brain derived from the Paxinos and Watson atlas[26]. AP stands for anteroposterior, and refers to the distance of the coronal section from bregma in mm.

After obtaining the average VEP for every channel of each probe used in our experiments, we further averaged the resulting waveforms, based on the depth (i.e., the distance from the cortical surface) at which they were recorded. To this aim, we discretized the cortical thickness (that, in rat V1, approximately spans 1350 µm[24]) into nine 150 µm-wide bins and we averaged the VEPs recorded from all the sites whose depth, as measured though

histology, fell inside the same bin. This yielded the template VEP profile shown in Fig. 1A, where a small, upward deflection, starting at about 90 ms following stimulus presentation, gradually molds into an increasingly deeper, broader and later downward deflection, while traveling from the surface to the bottom of the cortex – a profile that is qualitatively consistent with the laminar pattern of VEPs reported in previous rodent studies[15,21].

To obtain the depth-to-layer map, we established the laminar location of the recording sites in each recording session through visual inspection of the corresponding Nissl-stained brain slice (see Materials and methods for details and Fig. 3B for an example slice). By combining this information with the depth of the sites (also recovered from the Nissl sections), we computed the probability, for a site within a given cortical lamina, to be found at a given cortical depth. Fig. 1B shows the resulting depth distributions for layers 1-3 (yellow), layer 4 (brown), layer 5 (green) and layer 6 (blue). Given these distributions, the optimal boundary between a pair of adjacent layers can be defined as the depth at which the corresponding distributions intersect, since this choice minimizes the number of incorrect attributions between the two layers. The resulting boundaries (dashed lines) were thus used to build the final depth-to-layer map, shown in Fig. 1C.

Let's now delve into the details of the second step mentioned above: the VEP template-matching algorithm at the core of our method, whose aim is to infer the most likely insertion depth and tilt of each shank of the silicon probe used in a given recording session, relative to the surface of the cortex.

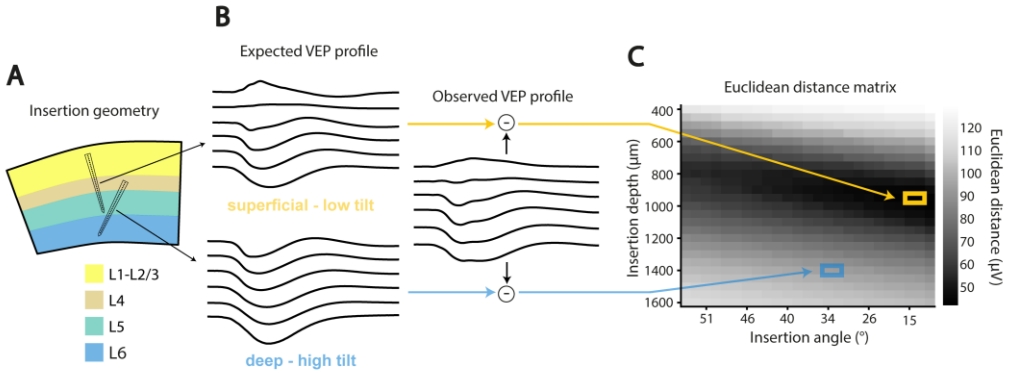

Figure 2. Graphical illustration of the VEP template-matching algorithm. A) The cartoon shows two hypothetical insertions of a single-shank silicon probe, reaching two different cortical depths with two different insertion angles. The cortical layers, which are color coded (see key), are the same as those shown in the depth-to-layer map of Fig. 1C. B) The hypothetical probe insertions shown in a give rise to two different patterns of expected VEPs (left), which are obtained by sampling the waveforms of the template VEP profile (shown in Fig. 1A) according to the expected depths of the recording sites. Each of the expected VEP profile is compared to the pattern of VEPs that was actually measured during the recording session (right), by computing the Euclidean distance. C) The matrix of Euclidean distances obtained by systematically considering all possible combinations of 25 insertion depths (referred to the tip of the probe and ranging from 400 to 1600 µm) and 25 insertion angles (ranging from 0 to 50°). The magnitude of the distance (in µV) is color coded (see color bar). The distances corresponding to the two insertions hypothesized in A-B are highlighted by the yellow and blue frames.

The algorithm consists of two steps. First, we compute the Euclidean distance between the observed input data (i.e., the VEPs recorded across the channels of the shank under exam) and the VEPs expected for any possible combination of depths and tilts of the shank over a 25x25 search grid (spanning a 400-1600 µm range of tip depths and a 0°-50° range of tilts). These expected VEPs are obtained from the template VEP profile (Fig. 1A)

under the hypotheses that the tip of the shank (i.e., the deepest recording site) is positioned at the desired depth and the whole shank is rotated of the desired tilt with respect to the cortical surface. These tilt and depth parameters, combined with the known inter-site spacing, univocally specify the expected depth of each recording site and, therefore, the expected VEP associated to that site, based on the template VEP profile. Fig. 2 graphically illustrates this procedure, by showing how two hypothetical shank insertions, having different tip's depths and tilts, give rise to two different patterns of expected VEPs. These are matched to the pattern of VEPs that was actually observed along the shank. The outcome of this procedure is a 25x25 matrix (Fig. 2C), where each element reports how good the match is between the observed VEPs and the expected VEPs (in terms of Euclidean distance), depending on the hypothesized insertion depth and tilt of the shank.

The second step of the algorithm makes use of this matrix to define the inferred insertion depth and tilt as the weighted average of all tested combinations of depths and tilts, where the weights are the inverse of the Euclidean distances reported in the matrix. Such a final estimate of the insertion depth and tilt of the shank is then combined with its known inter-site spacing to predict the depth of each recording site. This approach was preferred to simply use the depth and tilt values corresponding to the minimum of the matrix of Euclidean distances, because the latter yielded less accurate estimates of the sites' depth – i.e., with a root mean squared error (RMSE) that was more than twice as large as that reported in the next paragraph (see also Fig. 5A). Once inferred the cortical depth of a recording site along a given shank, the depth-to-layer map (shown in Fig. 1C) can be used to get also the putative layer assignment of the site.

Finally, let's turn to the quantitative validation of the depth inference method we propose.

To measure the accuracy of our method at inferring the cortical depth of the recording sites and their laminar location, we used a leave-one-out cross-validation procedure that worked as follows. We took only 17, out of the 18 recording sessions, to build the template VEP profile (see Fig. 1A), and we used it to predict the depth of the recording sites for the remaining left-out session. This procedure was applied exhaustively, so as to obtain the cross-validated accuracy of our predictions for each of the 18 sessions.

Fig. 3 shows the reconstruction of the recording sites of a 32-channel silicon probe, obtained with our template-matching method for an example session. This session was chosen for illustrative purposes, because the error in inferring the depth of the sites was the largest among the 18 sessions and, as such, it allowed a clearer visualization of the difference between prediction and ground-truth. Yet, for each of the 32 sites, the predicted depth (dots) was at less than 212 μm from the actual depth (diamonds), with an overall root mean squared error (RMSE) of 166 μm (the prediction was based on the shank insertion geometry that was inferred from the matrix of Euclidean distances shown in Fig. 2C). The actual depth was established by visual inspection of the Nissl-stained histological section (Fig. 3B), with superimposed the fluorescence image of the insertion track of the probe (in red), which had been coated with the fluorescent dye DiI before starting the recording (see Materials and methods). Visual examination of the section also allowed recovering the boundaries between the cortical layers (white dashed lines). When such ground-truth layer attribution was compared to the one predicted on the basis of the inferred cortical depth and the average

depth-to-layer map of Fig. 1C (also reported in Fig. 3A – see the gray dashed lines), the fraction of sites whose laminar location (color coded in Fig. 3A) was correctly predicted (black-circled dots) was 75%.

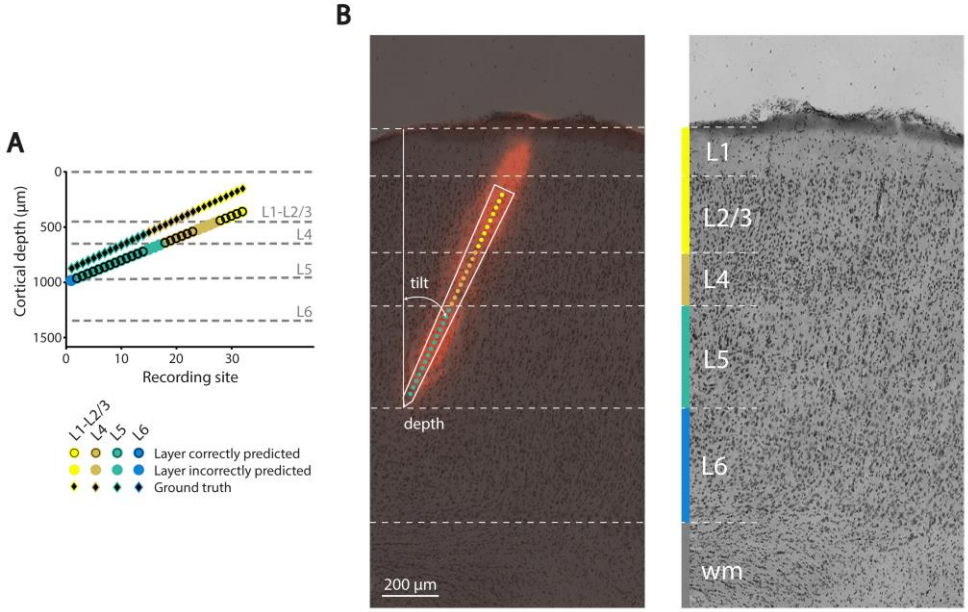

Figure 3. Validation of the VEP template-matching algorithm. (A) The depth of the recording sites of a single-shank silicon probe, as predicted by our VEP template-matching algorithm (colored dots), is compared to the actual depth of the sites, as recovered through histology (black diamonds). The dashed lines show the boundaries of the cortical layers, as established by the depth-to-layer map of Fig. 1C, and the color of the dots indicates the predicted layer of each recording site (see key), according to such map. The circular, black frame around a dot indicates that the layer prediction was correct, when compared to the actual laminar location of the recording site (color coded by the frame placed around each of the black diamond) that was obtained by histological analysis of the corresponding Nissl-stained brain slice. (B) Histological analysis of the brain slice showing the insertion track of the silicon probe analyzed in A. On the left, a bright-field image of the Nissl-stained coronal slice is superimposed with a fluorescence image showing the staining (red) produced by the insertion of the probe, which was coated with the fluorescent dye DiI. From this image, it was possible to estimate the depth and insertion angle (or tilt) of the probe during the session, and, given the known geometry of the probe (white outline), it was possible to recover the depth of each recording site (colored dots). It was also possible to estimate the boundaries between the cortical layers, by inspecting an adjacent, 60 µm-apart Nissl-stained slice that did not bear the mechanical lesion produced by the insertion of the probe (image on the right). The resulting layer boundaries, which were drawn based on the variation of size, morphology and density of the Nissl-labeled cells across the cortical thickness, are marked by the white, dashed lines. These boundaries allowed establishing the laminar location of the sites that is color-coded by the frames around the black diamonds in A. WM stays for white matter. The data reported in this figure refer to the same recording session used to illustrate the VEP template-matching algorithm in Fig. 2.

Overall, this example session allows appreciating how two sources of errors concur to limit the accuracy of the predicted laminar location of the sites: 1) the error on the predicted depth (i.e., the vertical distance between dots and diamonds in Fig. 3A); and 2) the difference between the actual layer boundaries for a specific recording session (as assessed through histology) and those inferred from the depth-to-layer map (i.e., the difference between the white dashed lines of Fig. 3B and the gray dashed lines of Fig. 3A). Nevertheless, despite these potential error sources, we were able to achieve accurate depth and layer predictions for most of the recording sessions. This is illustrated in Fig. 4, which shows the cross-validated accuracy of our method for each of the 18 sessions examined in our study (the gray area highlights the example session we just saw in detail in Fig. 3).

In most cases, the absolute distance between predicted (colored dots) and measured (black dots) depth of the recording sites was lower than 115 µm (75% quantile of the absolute error distribution), with the method yielding, for some sessions, RMSEs as low as a few tens of µm. As a result, the overall distribution of absolute depth errors (across all the sites of all the recording probes) displayed (Fig. 5A) a prominent peak in the 0-50 µm range, with no errors above 220 µm. This yielded a mean RSME ± SEM across sessions of 79 ± 11 µm.

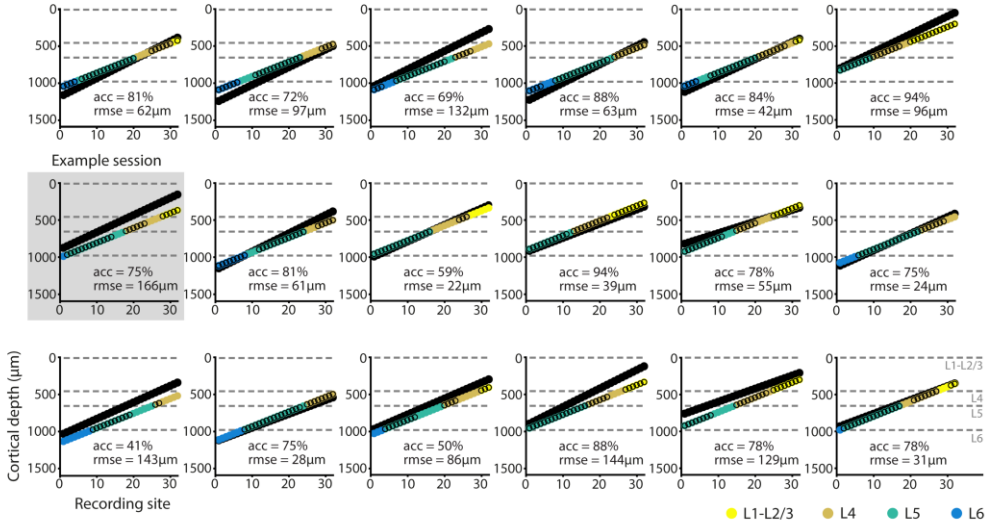

Figure 4. Accuracy attained by the VEP template-matching algorithm in each of the recording sessions. Every panel shows the predicted (colored dots) and measured (black dots) depths of the recording sites in a given session, along with the predicted laminar location (color-coded; same convention as in Fig. 1-3). As in Fig. 3A, a black frame around a dot indicates that the layer prediction was correct. The gray, dashed lines show the boundaries of the cortical layers, as established by the depth-to-layer map of Fig. 1C. For each session, is also reported the RMSE in inferring the depth of the sites and the accuracy in recovering their laminar location (as the percentage of correctly labeled sites). The gray area highlights the example session previously analyzed in Fig. 2-3.

Fig. 4 also reports the predicted layer attribution of the recording sites (color coded, with the layer boundaries marked by dashed lines) and whether such prediction was correct (black-circled dots), according to the histological assessment of the cortical sections. The fraction of correctly labeled sites (i.e. accuracy) ranged from 41% to 94%, with a mean ± SEM of 76 ± 3% across the sessions (first bar in Fig. 5B), when a distinction in four cortical laminae was considered (i.e., same as shown in Fig. 1C). When a coarser grouping of the layers in supragranular (1-3), granular (4) and infragranular (5-6) was considered, the accuracy of the labeling increased to 83 ± 3% (second bar in Fig. 5B), and further grew to 91 ± 2% for a binary partition of the layers into superficial (1-4) and deep (5-6; third bar in Fig. 5B).

The errors in predicting the laminar locations were fairly homogeneously distributed across the cortical thickness, when the accuracy was measured in terms of recall (i.e., the fraction of sites belonging to a given layer that were correctly labeled as such). As shown in Fig. 5C (black bars), recall accuracy peaked in layer 5 (~80% correct), slightly dropped in the adjacent layers (~75%), to become ~60% in the supragranular ones. When measured in terms of precision (i.e., the fraction of sites labeled in a given way, for which the labeling was correct), the accuracy also peaked in layer 5 (~80% correct) and dropped substantially only in layer 6 (~40%).

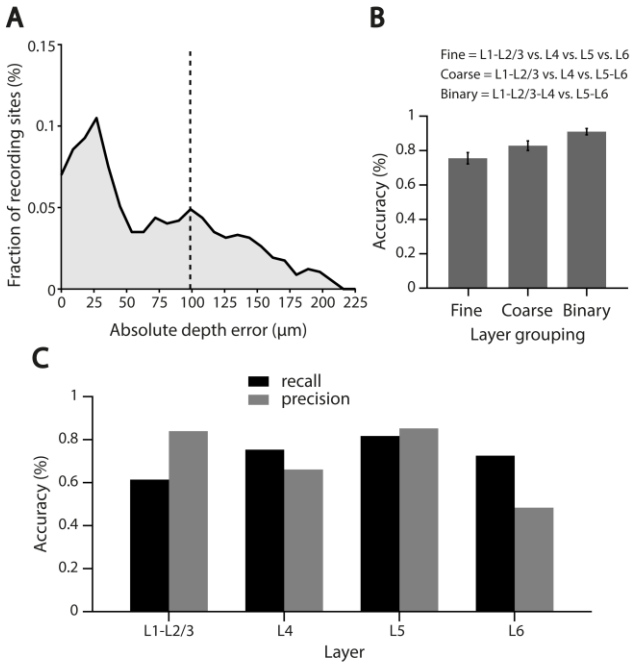

Figure 5. Overall accuracy of the VEP template-matching algorithm. A) Distribution of absolute errors in estimating the depth of the recording sites, as obtained by computing the absolute distance between predicted and measured depth for every site on every probe (the dashed line shows the mean of the distribution). B) Average accuracy (across the 18 recording sessions) in predicting the laminar location of the recording sites ± SEM, as measured for three different levels of distinction between the cortical layers – from fine to binary (see key). C) The overall recall (black) and precision (gray) accuracies in predicting the laminar location of the recording sites are reported as a function of the measured (for the recall) or predicted (for the precision) cortical layer of the sites. See the main text for a definition of recall and precision.

Discussion

The method we developed provides a valuable, automated alternative for inferring cortical depth and laminar location of the recording sites along a silicon probe, as compared to approaches based on visual inspection of the pattern of current sinks and sources that can be derived from the LFPs recorded along the probe, as in CSD analysis[10–21].

One major advantage over these approaches is that, once the template VEP profile (Fig. 1A) and the depth-to-layer map (Fig. 1C) have been established from a number of training sessions (with ground-truth depth and laminar information recovered through histology), our method is fully automatized and does not require any subjective decision about the location of layer boundaries to be taken.

A second, key feature of our method is that its accuracy can be quantitatively and rigorously estimated, given the above-mentioned ground-truth information gathered through histology. Specifically, our cross-validated measurements indicate that our method is very accurate in recovering the depth of the recording sites, with an average RMSE lower than 100 μm (Fig. 5A, dashed vertical line). In addition, the method yields reliable estimates of the cortical laminae, with a 76% accuracy for fine-grained discriminations that increases to 80-90% for coarser groupings of the layers (Fig. 5B), and with both the recall and precision accuracies peaking in layer 5 (~80% correct labeling; Fig. 5C). Unfortunately, a quantitative

comparison with CSD-based methods in terms of accuracy is hard to carry out because, to our knowledge, the accuracy of these approaches, relative to a ground-truth established through histology, typically is not reported.

Finally, another distinctive feature of our method is that it is able to reliably infer the depth and laminar location of the recording sites also in sessions where the probe did not span the whole cortical thickness and, in particular, in sessions where the VEPs were not recorded from the supragranular layers (see, in Fig. 4, the accuracy attained in sessions where the sites were all located below the boundary between layer 3 and 4). As mentioned previously, failure to sample VEPs from layers 2-3 would make it hard to properly identify the current sink associated with thalamic afferent input impinging into layer 4 by visual inspection of the CSD pattern.

In conclusion, we believe that our automated method for depth inference and laminar identification of recording sites represents a valuable alternative over existing, more qualitative approaches. Obviously, since our approach is based on obtaining a reliable estimate of the average VEP profile across the cortical thickness and of the depth and laminar location of the recording sites through histology, its application needs to be fine-tuned as a function of the cortical area under exam. In our study, we implemented and demonstrated the validity of our approach in the primary visual cortex of the rat. Its application to a different cortical area and/or a different species would require building first an appropriate VEP template and obtaining ground-truth depth measurements for that area/species.

Materials and methods

Animal preparation and surgery

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). Extracellular recording data from a total of 13 naïve, Long-Evans male rats (Charles River Laboratories), 8 from V1 sessions of the study[23] reported in chapter I of this thesis and 5 from V1 sessions of the study described in chapter V, were included in the analysis. Their age ranged 3-12 months and their weight ranged from 300 to 600 g. Each rat was anesthetized with an intraperitoneal (IP) injection of a solution of 0.3 mg/kg of fentanyl (Fentanest®, Pfizer) and 0.3 mg/kg of medetomidin (Domitor®, Orion Pharma). The level of anesthesia was monitored by checking the absence of tail, ear and hind paw reflexes, as well as monitoring blood oxygenation, heart and respiratory rate through a pulse oximeter (Pulsesense-VET, Nonin). A constant flow of oxygen was delivered to the rat throughout the experiment to prevent hypoxia. A constant level of anesthesia was maintained through continuous IP infusion of the same aesthetic solution used for induction, but at a lower concentration (0.1 mg/kg/h Fentanyl and 0.1 g/kg/h Medetomidin). This was done using a syringe pump (NE-500; New Era Pump Systems). Internal temperature of the animal was thermostatically kept at 37°C using a heating pad to prevent anesthesia-induced hypothermia.

After induction, the rat was secured to a stereotaxic apparatus (Narishige, SR-5R) in flat-skull orientation (i.e., with the surface of the skull parallel to the base of the stereotax) and, following a scalp incision, a craniotomy was performed over the target area in the left hemisphere (typically, a 2x2 mm window) and the dura was removed to allow the insertion of the electrode array. The coordinates used to target V1 were ~6.5 mm posterior from bregma and ~4.5 mm left to the sagittal suture (i.e., AP 6.5, ML 4.5). Throughout the procedure, the eyes of the animal were protected from direct light and kept hydrated by repeated application of an ophthalmic ointment (Epigel®, Ceva Vetem).

Once the surgery was completed, before probe insertion, the stereotax was placed on a rotating platform and the rat's left eye was covered with black, opaque tape, while the right eye (placed at 30 cm distance from the monitor) was immobilized using a metal eye-ring anchored to the stereotax. The platform was then rotated, so as to align the right eye with the center of the stimulus display and bring the binocular portion of its visual field to cover the left side of the display. For the whole duration of the recordings, eye and cortex were periodically irrigated using saline solution in order to keep them properly hydrated.

Electrophysiological recordings

Extracellular recordings were performed using single-shank, 32-channel silicon probes (NeuroNexus®) with site recording area of 775 μm^2 and 25 μm of inter-site spacing. After grounding (by wiring the probe to the animal's head skin), the electrode was manually lowered into the cortical tissue using an oil hydraulic micromanipulator (Narishige, MO-10; typical insertion speed: ~ 5 μm/s), up to the chosen insertion depth (~800-1200 μm from the cortical surface).

The probes were inserted with a variable tilt, between 0º and 30º, relative to the cortical surface. Extracellular signals were acquired using a system three workstation (Tucker-Davis Technologies) with a sampling rate of 25 kHz. Before insertion, the probe was coated with Vybrant® DiI cell-labelling solution (Invitrogen, Oregon, USA) to allow visualizing the probe insertion track post-mortem through histological procedures. To this aim, at the end of the recording session, an electrolytic lesion was also performed by delivering current (5 μA for 2 seconds) through the 4 deepest channels at the tip of the shank.

Raw voltage traces were acquired at 25 kHz sampling rate and later downsampled to 610 Hz after lowpass filtering to obtain LFPs. Traces were then visually inspected to identify possibly "broken" channels (easily identifiable by having very strongly attenuated voltage variations as compared to the surrounding channels). The traces recorded at such defective sites were replaced by the average of the two surrounding channels (i.e., above and below the broken site), so as to obtain a set of LFP traces without artifactual signal discontinuities across channels.

After this pre-processing step we extracted 168 ms-long (i.e. 275 samples-long) VEP traces from each site/channel, each starting from the onset of stimulus presentation. More specifically, the responses to all the repeated presentations of all the drifting gratings used during a recording session were averaged to obtain a smooth VEP for each channel (see next section and the Results).

Visual stimuli

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

Initially, a 15 min-long receptive field (RF) mapping procedure was used to verify in real-time the identity of the targeted area (based on the assessment of the known retinotopy of rat V1) and to optimize the location of the RF centers for the following, main stimulation protocol (i.e., to ensure that most RFs fell inside the monitor, by rotating the platform or repositioning the eye through adjustments of the eye-ring). Such a brief RF mapping protocol and its use for visual areas identification has been thoroughly described elsewhere[23,25].

Once the probe was positioned at the final recording location, the main presentation protocol was administered. For the 8 animals taken from chapter I study[23], this included 1 s-long drifting gratings, made of all possible combinations of 3 spatial frequencies (SF; 0.02, 0.04 and 0.08 cycle/°), 3 temporal frequencies (TF; 2, 4 and 8 Hz), and 12 directions (from 0° to 330°, in 30° increments). For the 5 additional animals from the study described in chapter V, the main protocol included 1 s-long drifting gratings, made of all possible combinations of 2 spatial frequencies (SF; 0.02, 0.04 cycle/°), 2 temporal frequencies (TF; 2, 6 Hz), and 12 directions (from 0° to 330°, in 30° increments). Each grating stimulus was presented in 20 repeated trials. All stimulus conditions were randomly interleaved, with a 1 s-long inter stimulus interval (ISI), during which the display was set to a uniform, middle-gray luminance level.

Stimuli were generated and controlled in MATLAB® 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.

Histology

At the end of each recording session, the animal was deeply anesthetized with an overdose of urethane (1.5 gr/kg) and perfused transcardially with phosphate buffer saline (PBS) 0.1 M, followed by 4% paraformaldehyde (PFA) in PBS 0.1 M, pH 7.2. The brain was then removed from the skull, post-fixed in 4% PFA for 24 h at 4°C, and then immersed in cryoprotectant solution (30% w/v sucrose in PBS 0.1 M) for at least 48 h at 4 °C. The brain was finally sectioned into 30 μm-thick coronal slices using a freezing microtome (Leica SM2000R, Nussloch, Germany). Sections were mounted immediately on Superfrost Plus slides and let dry at room temperature overnight. A brief wash in distilled water was performed to remove the excess of crystal salt sedimented on the slices, before inspecting them at the epifluorescence microscope. Each slice was then photographed with a digital camera (MBF Bioscience CX9000) adapted to a Leica microscope (Leica DM6000B-CTR6000, Nussloch, Germany), acquiring both a DiI fluorescence image (700 nm DiI filter) and a brightfield image, using a Leica PL Fluorotar 2.5X/0.07 objective. Following the acquisition of this set of images, the sections displaying the electrode fluorescent track were further stained for Nissl substance using a 0.5% Cresyl Violet Acetate solution, and new pictures were taken at 2.5X magnification. By superimposing the fluorescence, bright-field and Nissl-stained

images, it was possible to reconstruct the tilt and the anteroposterior (AP) position of the probe during the recording session, as well as the cortical depth and laminar location of all the recording sites. Specifically, the boundaries between the cortical layers were identified, based on the difference in size, morphology and density of the Nissl-labelled cells across the cortical thickness. The position of the probe relative to such boundaries and to cortical surface was determined by tracing the outline of the fluorescent track, and taking into account, when available, the location of the electrolytic lesion performed at the end of the recording session. Based on the known geometry of the silicon probe, it was possible to infer the location of each recording site along the shank, thus estimating its cortical depth and laminar location. The former was always measured along the line perpendicular to the layers at the site of interest. Such analysis was carried out using Inkscape 0.48.3.1. For illustrative purposes (Fig. 3B), we acquired large, higher magnification images of some slices from a selected representative session. To do so, we used a motorized inverted confocal Nikon Eclipse TI microscope equipped with a digital camera (Hamamatsu C4742-95), with a 20X/ 0.5 (Nikon Plan Fluor) objective. Image acquisition and stitching of a large field of 7x7 mm with 40% overlap was handled by Nikon NIS-Elements AR 4.0 software. Images were cropped and sized using Adobe Photoshop CS6 and the montages were generated in Adobe Illustrator CS6.

References

1. Adesnik, H. & Naka, A. Cracking the function of layers in the sensory cortex. *Neuron* **100**, 1028–1043 (2018).
2. Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51 (2013).
3. Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in neuronal population recordings. *Nat. Neurosci.* **19**, 1165 (2016).
4. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
5. Chung, J. E. *et al.* High-Density, Long-Lasting, and Multi-region Electrophysiological Recordings Using Polymer Electrode Arrays. *Neuron* **101**, 21-31.e5 (2019).
6. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407 (2012).
7. Li, P. H. *et al.* Anatomical identification of extracellularly recorded cells in large-scale multielectrode recordings. *J. Neurosci.* **35**, 4663–4675 (2015).
8. Blanche, T. J., Spacek, M. A., Hetke, J. F. & Swindale, N. V. Polytrodes: High-Density Silicon Electrode Arrays for Large-Scale Multiunit Recording. *J. Neurophysiol.* **93**, 2987–3000 (2005).
9. Steinmetz, N. A., Koch, C., Harris, K. D. & Carandini, M. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr. Opin. Neurobiol.* **50**, 92–100 (2018).
10. Hansen, B. J., Eagleman, S. & Dragoi, V. Examining local network processing using multi-contact laminar electrode recording. *JoVE J. Vis. Exp.* e2806 (2011).

11. Schroeder, C. E. *et al.* Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.* **85**, 1322–1327 (2001).
12. Schroeder, C. E., Mehta, A. D. & Givre, S. J. A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb. Cortex N. Y. NY 1991* **8**, 575–592 (1998).
13. Takeuchi, D., Hirabayashi, T., Tamura, K. & Miyashita, Y. Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* **331**, 1443–1447 (2011).
14. Dadarlat, M. C. & Stryker, M. P. Locomotion enhances neural encoding of visual stimuli in mouse V1. *J. Neurosci.* **37**, 3764–3775 (2017).
15. Deliano, M. *et al.* Dopaminergic neuromodulation of high gamma stimulus phase-locking in gerbil primary auditory cortex mediated by D1/D5-receptors. *Eur. J. Neurosci.* (2018).
16. Happel, M. F., Jeschke, M. & Ohl, F. W. Spectral integration in primary auditory cortex attributable to temporally precise convergence of thalamocortical and intracortical input. *J. Neurosci.* **30**, 11114–11127 (2010).
17. Hoy, J. L. & Niell, C. M. Layer-Specific Refinement of Visual Cortex Function after Eye Opening in the Awake Mouse. *J. Neurosci.* **35**, 3370–3383 (2015).
18. Headley, D. B. & Weinberger, N. M. Relational Associative Learning Induces Cross-Modal Plasticity in Early Visual Cortex. *Cereb. Cortex* **25**, 1306–1318 (2015).
19. Li, F., Jiang, W., Wang, T.-Y., Xie, T. & Yao, H. Phase-specific surround suppression in mouse primary visual cortex correlates with figure detection behavior based on phase discontinuity. *Neuroscience* **379**, 359–374 (2018).
20. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J Neurosci* **28**, 7520–36 (2008).
21. Sieben, K., Röder, B. & Hanganu-Opatz, I. L. Oscillatory Entrainment of Primary Somatosensory Cortex Encodes Visual Control of Tactile Processing. *J. Neurosci.* **33**, 5736–5749 (2013).
22. Nicholson, C. & Freeman, J. A. Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *J. Neurophysiol.* **38**, 356–368 (1975).
23. Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B. & Zoccolan, D. Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J. Neurosci.* **39**, 1649–1670 (2019).
24. Swanson, L. *Brain Maps: Structure of the Rat Brain*. (Gulf Professional Publishing, 2004).
25. Tafazoli, S. *et al.* Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6**, e22794 (2017).
26. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates*. (Academic Press, 2013).

Chapter III

Causal adaptation to visual input dynamics governs the development of complex cells in V1

Abstract

Visual perception relies on cortical representations of visual objects that remain relatively stable with respect to the variation in object appearance typically encountered during natural vision (e.g., because of position changes). Such stability, known as transformation tolerance, is built incrementally along the ventral stream (the cortical hierarchy devoted to shape processing), but early evidence of position tolerance is already found in primary visual cortex (V1) for complex cells. To date, it remains unknown what mechanisms drive the development of this class of neurons, as well as the emergence of tolerance across the ventral stream. Leading theories suggest that tolerance is learned, in an unsupervised manner, either from the temporal continuity of natural visual experience or from the spatial statistics of natural scenes. However, neither learning principle has been empirically proven to be at work in the postnatal developing cortex. Here we show that passive exposure to temporally continuous visual inputs during early postnatal life is essential for normal development of complex cells in rat V1. This was causally demonstrated by rearing newborn rats with frame-scrambled versions of natural movies, resulting in temporally unstructured visual input, but with unaltered, natural spatial statistics. This led to a strong reduction of the fraction of complex cells, which also displayed an abnormally fast response dynamics and a reduced ability to support stable decoding of stimulus orientation over time. Conversely, our manipulation did not prevent the development of simple cells, which showed orientation tuning and multi-lobed, Gabor-like receptive fields as sharp as those found in rats reared with temporally continuous natural movies. Overall, these findings causally implicate unsupervised temporal learning in the postnatal development of transformation tolerance but not of shape tuning, in agreement with theories that place the latter under the control of unsupervised adaptation to spatial, rather than temporal, image statistics.

Introduction

It has long been proposed that the tuning of sensory neurons is determined by adaptation to the statistics of the signals they need to encode[1,2]. In the visual domain, this notion has given rise to two broad families of unsupervised learning algorithms: those relying on the spatial structure of natural images, referred to as "unsupervised spatial learning" (USL) models[1–6]; and those leveraging on the spatiotemporal structure of natural image sequences, referred to as "unsupervised temporal learning" (UTL) models[7–15]. Both kinds of learning have been applied to explain the ability of visual cortical representations to selectively code for the identity of visual objects, a property known as shape tuning, while tolerating variations in their appearance (e.g., because of position changes), a property known as transformation tolerance (or invariance)[16].

These properties are built incrementally along the ventral stream (the cortical hierarchy devoted to shape processing), but the earliest evidence of shape tuning and invariance in the visual system can be traced back to primary visual cortex (V1), where "simple cells" first exhibit tuning for non-trivial geometrical patterns (oriented edges) and "complex cells" first display some degree of position tolerance[17]. In sparse coding theories (arguably the most popular incarnation of USL), maximizing the sparsity of the representation of natural images produces Gabor-like edge-detectors that closely resemble the receptive fields (RFs) of V1 simple cells[5,6]. Other USL models, by optimizing objective functions that depend on the combination of several linear spatial filters, also account for the emergence of position-tolerant edge detectors, such as V1 complex cells[3,4]. The latter, however, have been more commonly modeled as the result of UTL, where the natural tendency of different object views to occur nearby in time is used to factor out object identity from other faster-varying, lower-level visual attributes. Interestingly, while some UTL models presuppose the existence of a bank of simple cells, upon which the complex cells' representation is learned[7,11–15], other models, such as slow feature analysis (SFA), directly evolve complex cells from the pixel (i.e., retinal) representation, thus simultaneously learning shape selectivity and invariance[8,9].

To date, it remains unclear what role these hypothesized learning mechanisms play in the developing visual cortex, if any. In fact, although it is well established that early visual experience can strongly affect the development of visual cortical tuning[18–20], empirical support for the role of sparse coding in determining orientation selectivity is still inconclusive[6,21], while no causal evidence has been gathered yet to demonstrate the involvement of UTL in postnatal development of invariance and/or selectivity – but see[22,23] for a behavioral study of the role of UTL in the development of chicks' object vision.

Results

To address the latter question, we took 18 newborn rats (housed in light-proof cabinets from birth) and, from postnatal day (P)14 (i.e., at eye opening) through P60 (i.e., well beyond the end of the critical period), we subjected them to daily, 4-hours-long exposures inside controlled

visual environments (Fig. 1A). Specifically, 8 animals (the "control group") were exposed to natural movies, while the remaining 10 (the "experimental group") were exposed to their frame-scrambled versions (Fig. 1B). Critically, this manipulation destroyed the temporal continuity of natural visual experience for the experimental rats, while sparing the spatial structure of the individual image frames, which remained the same as for the control animals.

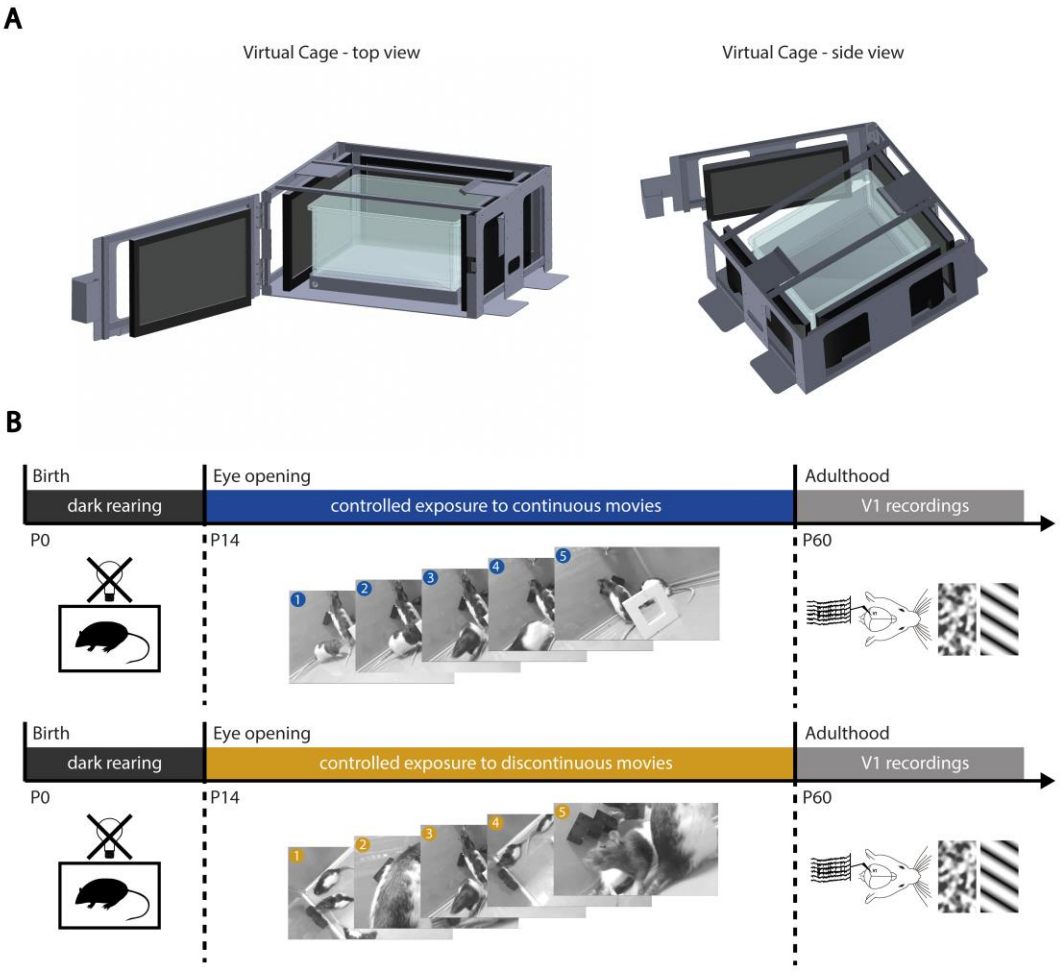

Figure 1. Experimental design. A) CAD drawing of a virtual cage – the custom apparatus used to rear newborn rats in visually controlled environments. The cage consists of a metal, box-like structure (light gray) holding 4 computer monitors (black/dark gray) that fully surround a transparent basin (light blue), where a newborn rat can be placed for immersive exposure to the visual scenes/movies displayed on the monitors. B) Time course of the experiment. Two groups of rats, control (top) and experimental (bottom), were born in dark and housed in lightproof cabinets (i.e. "dark reared") till eye opening (black bars). Afterward, the control rats were subjected to daily, 4-hours long exposures to natural videos inside the virtual cages (blue bar), while the experimental rats were subjected to the frame-scrambled versions of the same movies (orange bar). Starting from P60, neuronal recordings from V1 of both the control and experimental rats were performed under anesthesia (gray bars), while the animals were exposed to drifting gratings and movies of spatially and temporally correlated noise.

This allowed isolating the pure contribution of temporal contiguity to the postnatal development of V1 simple and complex cells.

This was achieved by performing multi-channel extracellular recordings from V1 of each rat under fentanyl/medetomidin anesthesia[24], shortly after the end of the controlled-rearing period. Our recordings mainly targeted layer 5, which is known to contain the largest fraction of complex cells in rodents[25] as well as layer 4, with the distributions of recorded units across the cortical depth and the cortical laminae being statistically the same for the control and experimental groups (Fig. 2).

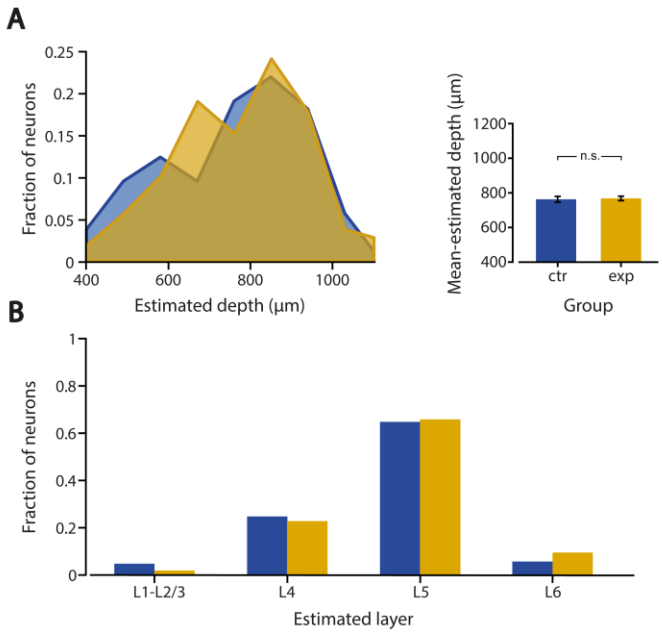

Figure 2. The distributions of cortical depths and laminar locations did not differ between the populations of control and experimental units. A) The distributions of cortical depths (left) and their means (right) did not significantly differ between the populations of control (blue) and experimental (orange) units ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, two-tailed, unpaired t-test). Error bars are SEMs. B) The distributions of laminar locations were also not significantly different between control (blue) and experimental (orange) units ($p > 0.05$, $\chi 2$ test). Both the cortical depths and laminar locations were estimated from the patterns of visually evoked potentials recorded along the silicon probes (see Materials and methods).

During a recording session, each animal was presented with drifting gratings spanning 12 directions (from 0° to 330° in steps of 30°), and with contrast-modulated movies of spatially and temporally correlated noise[24,25]. Responses to the noise movies allowed inferring the linear RF structure of the recorded units using the Spike-Triggered Average (STA) analysis (see Materials and methods). Responses to the drifting gratings were used to estimate the tuning of the neurons with standard orientation and direction selectivity indexes (OSI and DSI; see Materials and methods), as well as to probe their sensitivity to phase-shifts of their preferred gratings, thus measuring their position tolerance[24,25].

This is illustrated in Fig. 3A, which shows an example complex cell from the control group (left: blue lines) an example simple cell from the experimental group (right: orange lines). Those two specific example neurons were chosen for their particularly clear modulation (or lack

of it) of the firing patterns in response to drifting gratings that are the key response property upon which the simple (or complex) classification is carried out.

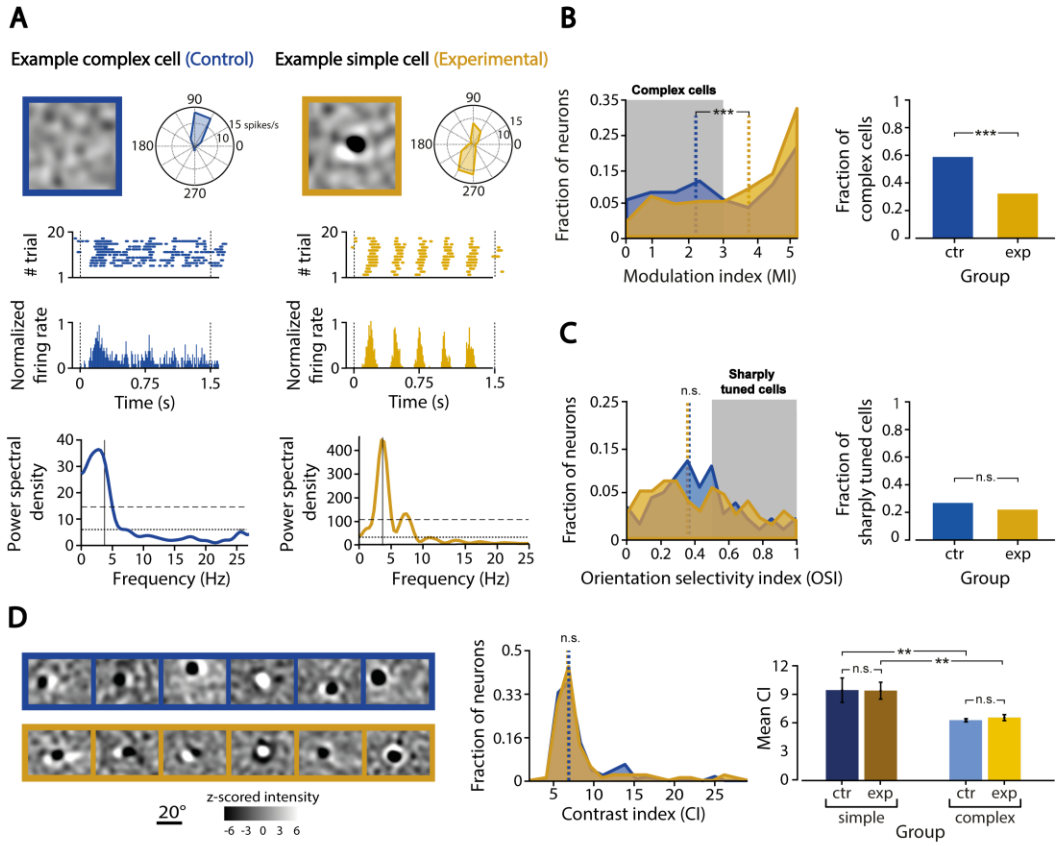

Figure 3. Postnatal rearing in temporally discontinuous visual environments results in an impoverished population of V1 complex cells, but leaves spatial tuning of V1 neurons unaltered. A) A representative V1 complex cell of the control group (left, blue lines) is compared to a representative simple cell of the experimental group (right, orange lines). For each neuron, the graph shows, from top/left to bottom: 1) the linear RF structure inferred through STA; 2) the direction tuning curve; 3) the raster plot, with the number of spikes (dots) fired across repeated presentations of most effective grating stimulus; 4) the corresponding peri-stimulus time histogram (PSTH), computed in 10 ms-wide time bins; and 5) its power spectrum, with indicated its mean (dotted line), its mean plus SD (dashed line), and the 4 HZ frequency of the grating stimulus (vertical line). B) Left: distributions of the modulation index (MI) used to distinguish the poorly phase-modulated complex cells (MI < 3; gray-shaded area) from the strongly modulated simple cells (MI > 3), as obtained for the control (blue) and experimental (orange) V1 populations (only units with OSI > 0.4 included). Both the distributions and their medians (dashed lines) were significantly different ($p < 0.02$, Kolmogorov-Smirnov test; $p < 0.001$, Wilcoxon test). Right: the fraction of units that were classified as complex cells (i.e., with MI < 3) was significantly larger for the control than for the experimental group ($p < 0.001$, Fisher exact-test). C) Left: distributions of the orientation selectivity index (OSI), as obtained for the control (blue) and experimental (orange) V1 populations. No significant difference was found between the two distributions and their medians ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, Wilcoxon test). Right: the fraction of sharply orientation-tuned units (i.e., units with OSI > 0.6) did not differ between the two groups ($p > 0.05$, Fisher exact-test). D) Left: Examples of linear RFs inferred through STA for the control (blue frame) and experimental group (orange frame). Center: distributions of the contrast index (CI) used to measure the sharpness of the STA images, as obtained for the control (blue) and experimental (orange) V1 populations. No significant difference was found between the two distributions and their medians ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, Wilcoxon test). Right: Mean values (± SEM) of the CIs, computed separately for the simple (dark bars) and complex (light bars) cells of the two groups (only units with OSI > 0.4 included). Within each group, the mean CI was significantly larger for the simple than for the complex cells (** $p < 0.01$, two-tailed, unpaired t-test).

Both units displayed sharp orientation tuning (polar plots), but the STA method successfully recovered a sharp, Gabor-like RF only for the simple cell - as expected, given the nonlinear stimulus-response relationship of complex cells[24]. Consistently, the response of the complex cell was only weakly modulated at the temporal frequency (4 Hz) of its preferred grating (middle plots), with the highest power spectral density concentrated at frequencies < 4 Hz (bottom plot). By contrast, the response of the simple cell was strongly phase-modulated, with a power spectrum narrowly peaked at the grating frequency. Thus, by z-scoring the power spectral density of the response at the preferred grating frequency, it was possible to define a modulation index (MI) that distinguished between complex (MI < 3) and simple (MI > 3) cells[24].

We applied this criterion to the neuronal populations of 105 and 158 well-isolated single units recorded from, respectively, the control and experimental group, and we found a significantly lower fraction of complex cells in the latter (39%, 47/158) with respect to the former (55%, 58/105; $p < 0.01$, Fisher exact-test). Consistently, the median MI for the control population (2.69 ± 0.29) was significantly smaller than for the experimental one (3.52 ± 0.25; $p < 0.05$, Wilcoxon test). Such a difference became very sharp after restricting the comparison to the neurons that, in both populations, were at least moderately orientation tuned (i.e., 50 control and 75 experimental units with OSI > 0.4). The resulting MI distribution for the control group had a typical double-peak shape[24], featuring two maxima, at MI~2 and MI~5, corresponding to the two classes of the complex and simple cells (Fig. 3B, blue curve). Instead, for the experimental group, the peak at low MI was flattened out, leaving a single, prominent peak at MI~5 (orange curve). This resulted in a large, significant difference between the two distributions and their medians (dashed lines), with the fraction of complex cells being almost half in the experimental (35%, orange bar) than in the control group (60%, blue bar). Conversely, no difference was observed between the two groups in terms of orientation tuning (Fig. 3C), with the OSI distributions (blue and orange curves) and their medians (dashed lines) being statistically undistinguishable, as well as in terms of the fraction of sharply orientation-tuned units (i.e., neurons with OSI > 0.6; blue vs. orange bar). A similar result was found for direction tuning (Fig. 4).

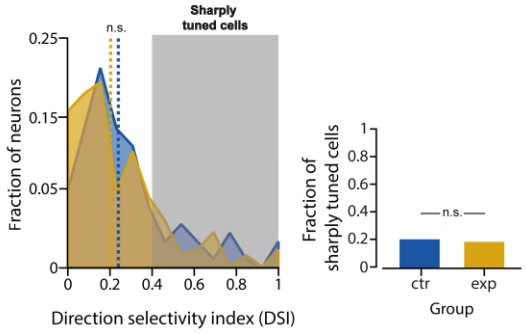

Figure 4. Postnatal rearing in temporally discontinuous visual environments leaves direction tuning of V1 neurons unaltered. Left: distributions of the direction selectivity index (DSI) used to measure the sharpness of orientation tuning, as obtained for the control (blue) and experimental (orange) V1 populations. No significant difference was found between the two distributions and their medians ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, Wilcoxon test). Right: the fraction of sharply orientation-tuned units (i.e., units with DSI > 0.6) did not differ between the two groups ($p > 0.05$, Fisher exact-test).

Taken together, these findings suggest that our experimental manipulation substantially impaired the development of complex cells, but not the emergence of orientation and motion sensitivity, and the development of simple cells. This was confirmed by the fact that STA was as successful at yielding sharp, linear RFs (often similar to Gabor filters) for the experimental units as for the control ones (see examples in Fig. 3D, left). The sharpness of the STA images, as assessed through an expressly devised contrast index[24] (CI; see Materials and methods), was similar for the two groups, with the CI distributions and their medians being statistically undistinguishable (Fig. 3D, blue vs. orange curve/line). As expected, for both groups, the mean CI was significantly larger for the simple than for the complex cells (dark vs. light bars), reflecting the better success of STA at inferring the linear RFs of the former, but no difference was found between the mean CIs of the simple cells of the two groups (Fig. 3D, dark blue vs. brown bar) and the mean CIs of the complex cells (Fig. 3D, light blue vs. yellow bar).

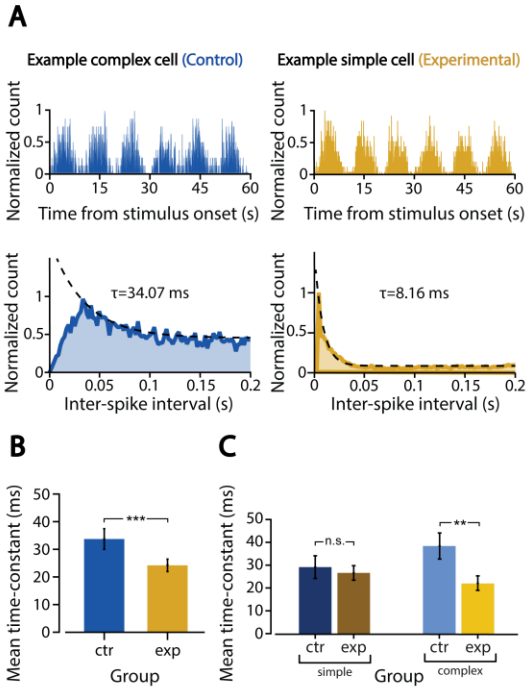

Figure 5. Postnatal rearing in temporally discontinuous visual environments leads to the development of complex cells with abnormally fast response dynamics. A) Top: PSTHs showing the average responses of the two example neurons of Fig. 3A to the contrast-modulated noise movies (see Materials and methods). For both neurons, the firing rate was strongly modulated at the frequency of variation of the contrast of the movies (i.e., 0.1 Hz). Bottom: distributions of inter-spike intervals (ISIs) of the spike trains evoked by the noise movies for the two example neurons. The resulting autocorrelograms were fitted with exponential decaying functions (dashed lines) to measure the slowness (i.e., the time constant τ of the exponential fit) of the responses. In this example, the complex cell (blue curve) displays slower dynamics (i.e., larger τ) than the simple cell (orange curve). The two units also differ in the number of counts at low ISIs, which is much larger for the simple cell, as expected for a unit firing tightly packed trains of spikes (see Fig. 3A). B) Mean values (± SEM) of the time constants τ, computed for the control (blue) and experimental (orange) populations (*** $p < 0.001$, two-tailed, unpaired t-test). C) Mean values (± SEM) of the time constants τ, computed separately for the simple (dark bars) and complex (light bars) cells of the two groups. While the simple cells had equally fast dynamics, the complex cells were significantly slower in the control than in the experimental group (** $p < 0.01$, two-tailed, unpaired t-test).

Next, we tested the extent to which the experimental units that had been classified as complex cells fully retained the functional properties of this class of neurons. Based on intuitive considerations (i.e., the local invariance of complex cells) and the predictions of SFA[8,9], complex cells should fire more persistently than simple cells, in response to a continuous, spatiotemporally correlated visual input. To measure the persistence (i.e. "slowness") of neuronal responses, we computed the time-constants of the exponential fits to the autocorrelograms of the spike trains evoked by the noise movies (see examples in Fig. 5A).

As expected, the average time constant was larger for the control than for the experimental units (Fig. 5B). Such difference, however, was not merely driven by the larger fraction of complex cells in the control group (Fig. 3B). In fact, while the average time constants did not significantly differ between the simple cells of the two groups (Fig. 5C, dark blue vs. brown bar), the firing of complex cells was faster for the experimental than for the control units (yellow vs. light blue bar).

To understand the functional implication of such abnormally fast dynamics, we assessed the ability of the four distinct populations of simple and complex cells of the two groups to support stable decoding of stimulus orientation over time. To this aim, we randomly sampled 300 neurons from each population (after having first matched the populations in terms of OSI and orientation preference distributions; see Materials and methods), so as to obtain four, equally sized and similarly tuned pseudo-populations, whose units homogenously covered the orientation axis.

We then trained binary logistic classifiers to discriminate between 0°- and 90°-oriented gratings (drifting at 4 Hz), based on the activity of each pseudo-population. Each classifier was trained using neuronal responses (spike counts) in a 33 ms-wide time bin that was randomly chosen within the presentation epoch of the gratings. We then tested the ability of each classifier to generalize the discrimination to test bins at increasingly larger time lags (TLs) from the training bin (see Fig. 6A and Materials and methods for details).

As expected, given the strong phase dependence of their responses (see cartoon in Fig. 6A, top), the simple cells from both groups yielded generalization curves that were strongly modulated over time and virtually identical (Fig. 6B, dark blue and brown curves). The performance was high (≥ 80% correct) at test bins where the phase of the grating was close to that of the training bin (i.e., at TLs that were multiple of the 250 ms grating period), but it dropped to less than 30% correct (i.e., well below chance; dashed line) at test bins where the grating was in opposition of phase with respect to the training bin (e.g., at TL ~ 125 ms). By comparison, the complex cells of the control group, by virtue of their weaker phase dependence (see cartoon in Fig. 6A, bottom), afforded a way more phase-tolerant decoding of grating orientation, with the performance curve never dropping below chance level at any TL (Fig. 6B, light blue curve). However, for the complex cells of the experimental group, the performance curve (in orange) was not as stable – at most TLs, it was 5-10 percentage points smaller than the performance yielded by the control complex cells, dropping significantly below chance at test bins where the grating was in opposition of phase with respect to the training bin. That is, the

ability of the experimental complex cells to support phase-tolerant orientation decoding was somewhat in between that of properly developed complex cells and that of simple cells.

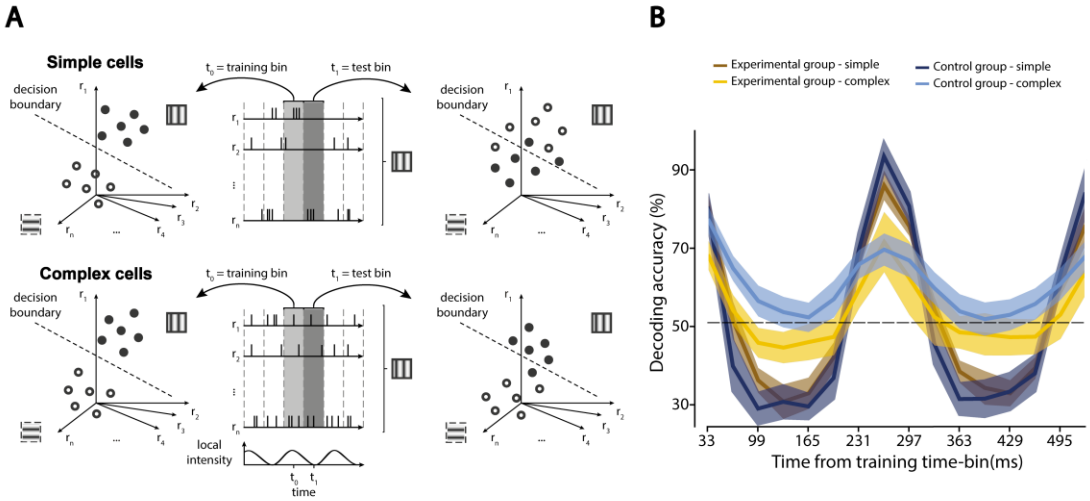

Figure 6. Postnatal rearing in temporally discontinuous visual environments reduces the ability of complex cells to support phase-tolerant discrimination of grating orientation. A) The cartoon illustrates the expected outcome of the decoding analysis to test the ability of simple and complex cells to support phase-tolerant discrimination of grating orientation. In the case of simple cells (top), a linear classifier built at time t_0 (middle, light gray shading) to successfully discriminate a vertical from a horizontal drifting grating (left: the filled and empty dots are well separated, within the neuronal representational space, by the linear decision boundary) will generalize poorly, when tested at a later time t_1 (middle, dark gray shading), with the accuracy dropping even below chance (right: the filled and empty dots swap sides of the linear decision boundary), because of the strong phase dependency of the responses r_i (middle: some neurons firing at t_0 will stop firing at t_1, while some other units that are silent at t_0 will respond at t_1). By contrast, for a population of complex cells (bottom), given the larger stability of the responses r_i (middle), the decision boundary resulting from the training at t_0 (left) will generalize better at t_1 (right: the filled and empty dots are still mostly on the original side of the decision boundary). B) Decoding accuracy yielded by the four populations of simple control (dark blue), complex control (light blue), simple experimental (brown) and complex experimental (orange) cells, in the vertical (i.e., 90º) vs. horizontal (i.e., 0º) grating discrimination task, in 33 ms-wide test bins located at increasingly larger time lags from the training bin (i.e., bin with lag = 0). The solid curves are the averages of many resampling loops of the neuronal population vectors and the training bins (see Materials and methods). The shaded regions are the bootstrap-estimated 95% confidence intervals of the solid lines (see Materials and methods).

Discussion

Overall, these findings show that destroying the temporal continuity of early visual experience severely interferes with the typical development of complex cells in V1, leading to a sizable reduction of their number (Fig. 3B) and an impairment of their functional properties (Fig. 5C and 6B). This implies that experience with the temporal structure of natural image sequences plays a critical role in the postnatal development of the earliest form of invariance found along the ventral stream. Such an instructive role of temporal continuity of visual stimuli, so far, has been empirically demonstrated only in adult monkeys, at the very last stage of this pathway: the inferotemporal cortex[26,27]. At the same time, our experiments show that development of

orientating tuning is unaffected by the lack of experience with temporal continuity (Fig. 3D), with simple cells exhibiting unaltered spatial (Fig. 3C), temporal (Fig. 5C) and functional (Fig. 6B) properties.

Even if quite pronounced, all the above mentioned alterations of complex cells reported in our experimental animals didn't amount to a complete abolishment of their existence following the temporal discontinuous rearing. At a first glance, this may seem partially at odd with the idea that temporal continuity is strictly necessary for the development of transformation tolerance in V1. However, it should be noticed that, in practice, input temporal structure disruption our animas were subject to, was not full. In fact, even if the rearing videos were generated to lack any temporal structure, due to self-motion of the animals themselves and to features of the interior of the rearing cages, such as monitor border frames, the rats still could experience some examples of continuous visual transformations (even if less than control animals – see Materials and methods for further discussion). The presence of such residual temporal structure in the visual input to the developing cortex of experimental rats could easily motivate the incomplete disruption of complex cells observed in our animals.

Another reason that could likely explain such observation, is the possibility that complex cell properties are − to some extent − the result of genetically-encoded developmental programs defining their wiring (and therefore their initial functional properties), in an experience independent way. Hints supporting such "hardwiring" hypothesis (thoroughly laid out in a recent perspective[28]) exist, e.g. for what concerns orientation selectivity in the mouse[29,30]. No data is however available at the moment regarding the state of complex cells transformation tolerance before the onset of visual experience in rodents. It should also be noticed how the innate "hardwiring" scenario doesn't exclude the possibility that an experience-dependent refinement of innate functional properties may occur, through the exposure to the statistics of the sensory environment, at later developmental stages.

Furthermore, another way to make sense of the partial preservation of V1 transformation tolerance in experimental animals while, at the same time, reconciling the unsupervised learning and the innatistic stances, is "innate learning"[31].

Key to this concept is the idea that, during development, neural circuits could, by virtue of their genetically-determined structure, self-generate neural activity patterns able to act as "training examples", thus sculpting the wiring of the same or other circuits. Such activity dependent structuring may be driven by the same unsupervised plasticity rules that will, after the onset of sensory experience, act on evoked activity. Specifically, in the context of primary visual cortex development, "retinal waves" have been proposed to play the role of spontaneously generated training input. Spatiotemporally correlated activity patterns evoked by retinal waves in V1, before eye opening, have been implicated in playing a pivotal role in driving the development of columnar organization in cats, ferrets and monkeys[32,33]. Crucially, and more importantly to our discussion, theoretical work demonstrated that retinal waves, sharing many relevant temporal and spatial properties with the natural visual input, are sufficient for UTL (instantiated as SFA) to reproduce a rich set of complex-cell features[34].

In conclusion, in light of what explained above, it appears quite likely that the altered visual exposure undergone by our experimental animals didn't act on visual cortical circuits in "blank" state, but rather on the resulting state induced by early, innate, developmental programs, combined with the effect of internally generated activity such as retinal waves.

Summing up, from a theoretical standpoint, our results causally validate the family of UTL models[7–15] at the neural level, albeit limiting their scope to the development of invariance. More in general, since slowness has been related to predictability[35–37], our results are also consistent with normative approaches to sensory processing that are based on temporal prediction[38]. On the other hand, our findings, by showing that exposure to the spatial structure of natural images alone is not enough to enable proper development of complex cells, reject computational accounts of invariance based exclusively on USL[3,4], while leaving open the possibility that the latter may govern the experience-dependent development of shape tuning[1,2,5,6]. Consequently, our study tightly constrains unsupervised models of visual cortical development, supporting theoretical frameworks where the objectives of sparseness and slowness maximization coexist, to yield, respectively, shape selectivity and transformation tolerance[13,14,39].

Materials and Methods

Animal procedures and controlled rearing

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 DGSAF 22791-A, submitted on September 7, 2015 and approved on December 10, 2015, approval 1254/2015-PR).

Data were obtained from 18 Long–Evans male rats that were born and reared in our facility for visually controlled rearing. The facility consists of a small vestibule, where the investigators can wear the infrared goggles that are necessary to operate in total darkness, and a larger, lightproof room, containing a lightproof housing cabinet (Tecniplast) and four custom cabinets (Tecniplast) for exposure of the rats to controlled visual environments.

Pregnant mothers (Charles River Laboratories) where brought into the housing cabinet about one week before delivery. Pups were born inside the cabinet and spent the first two weeks of their life in total darkness with their mothers. Starting from P14 (i.e., at eye opening) through P60 (i.e., well beyond the end of the critical period), each rat, while still housed in full darkness (i.e., inside the housing cabinet) for most of the day, was also subjected to daily, 4-hours-long exposures inside an immersive visual environment (referred to as the "virtual cage"), consisting of a transparent basin (480x365x210 mm, Tecniplast 1500U), fully surrounded by 4 computer-controlled LCD monitors (one per wall; 20" HP P202va; see Fig. 1A), placed on the shelf of one of the custom cabinets (each cabinet had 4 shelves, for a total of 16 rats that could be simultaneously placed in the visually controlled environments). These controlled-rearing

environments (which are reminiscent of those used to study the development of object vision in chicks[22,23,40]) were custom designed in collaboration with Videosystem, which took care of building and installing them inside the custom cabinets.

Different visual stimuli were played on the monitors, depending on whether an animal was assigned to the experimental or the control group. Rats in the control group (n = 8) were exposed to natural movies, including both indoor and outdoor scenes, camera self-motion and moving objects. Overall, the rearing playlist included 16 videos of different duration, lasting from a few minutes to half an hour. The playlist was played in random order and looped for the whole duration of the exposure. Rats from the experimental group (n = 10) were exposed to a time-shuffled version of the same movies, where the order of the frames within each video was randomly permuted, so as to destroy the temporal continuity of the movie, while leaving unaltered the natural spatial statistics of the individual image frames. All movies were played at 15 Hz, which is approximately half of the critical flicker fusion frequency (~30-40 Hz) that has been measured for the rat[41]. This ensured that, while the temporal correlation of the input was substantially broken, no fusion occurred between consecutive frames of the movies, thus allowing the rats to experience the spatial content of the individual image frames. On the other hand, choosing a frame rate lower than the flicker fusion frequency allowed the rats, at least in principle, to still experience some residual amount of temporal continuity in the visual input, resulting from scanning the image frames trough head or eye movements. This, together with the presence of some stable visual features in the environment (e.g., the dark edges of the monitors) and the possibility for the rats to see their own body, may account for the residual fraction of complex cells observed in experimental group (Fig. 2B).

Animal care, handling and transfer operations were always executed in absolute darkness, using night vision goggles (Armasight NXY7), in such a way to prevent any unwanted exposure of the animals to visual inputs different from those chosen for the rearing.

Surgery and recordings

Acute extracellular recordings were performed between P60 and P90 (last recording). During this 30-days period, the animals waiting to undergo the recording procedure were maintained on a reduced visual exposure regime (i.e., 2-hours-long visual exposure sessions every second day; see previous section).

The surgery and recording procedure was very similar to the one described for V1 recording sessions of the study[24] described in chapter I of this thesis. Briefly, the day of the experiment, the rat was taken from the rearing facility and immediately (within 5-10 minutes) anesthetized with an intraperitoneal injection of a solution of 0.3 mg/kg of fentanyl (Fentanest, Pfizer) and 0.3 mg/kg of medetomidin (Domitor, Orion Pharma). A constant level of anesthesia was then maintained through continuous intraperitoneal infusion of the same anesthetic solution used for induction, but at a lower concentration (0.1 mg/kg/h fentanyl and 0.1 g/kg/h medetomidine), by means of a syringe pump (NE-1000; New Era Pump Systems). After induction, the rat was secured to a stereotaxic apparatus (Narishige, SR-5R) in flat-skull

orientation (i.e., with the surface of the skull parallel to the base of the stereotax) and, following a scalp incision, a craniotomy was performed over the target area in the left hemisphere (typically, a 2x2 mm^2 window). The dura was then removed to allow the insertion of the electrode array. The coordinates of penetration used to target V1 were ~6.5 mm posterior from bregma and ~4.5 mm left to the sagittal suture (i.e., anteroposterior 6.5, mediolateral 4.5). Once the surgical procedure was completed, and before probe insertion, the stereotax was placed on a rotating platform and the rat's left eye was covered with black, opaque tape, while the right eye (placed at 30 cm distance from the monitor) was immobilized using a metal eye-ring anchored to the stereotax. The platform was then rotated in such a way to bring the binocular visual field of the right eye to cover the left side of the display.

Extracellular recordings were performed using either single- (or double-) shank 32- (or 64-) channel silicon probes (NeuroNexus®) with site recording area of 775 μm^2 and 25 µm of intersite spacing. After grounding (by wiring the probe to the animal's head skin), the electrode was manually lowered into the cortical tissue using an oil hydraulic micromanipulator (Narishige, MO-10; typical insertion speed: 5 m/s), up to the chosen insertion depth (800 –1000 µm from the cortical surface), either perpendicularly or with a variable tilt, between 10° and 30° relative to the vertical to the surface of the skull. Extracellular signals were acquired using a system 3 workstation (Tucker Davis Technologies) with a sampling rate of 25 kHz.

Since, in rodents, the largest fraction of complex cells in found in layer 5 of V1[25], our recordings aimed at sampling more densely that layer. This was verified a posteriori (Fig. 2), by estimating the cortical depth and laminar location of the recorded units, based on the patterns of visually evoked potentials (VEPs) recorded across the silicon probes used in our recording sessions. More specifically, we used the template-matching algorithm for laminar identification of cortical recording sites the development and validation[42] of which was described in chapter II of this thesis.

Briefly, the method finds the optimal match between the pattern of VEPs recorded, in a given experiment, across a silicon probe and a template VEP profile, spanning the whole cortical thickness, that has been computed by merging an independent pool of 18 recording sessions, in which the ground-true depth and laminar location of the recording sites had been recovered through histology. The method achieves a cross-validated accuracy of 79 µm in recovering the cortical depth of the recording sites and a 72% accuracy in returning their laminar position, with the latter increasing to 83% for a coarser grouping of the layers into supragranular (L1-3), granular (L4) and infragranular (L5-6).

Visual stimuli

During a recording session, each animal was presented with: i) 20 repetitions (trials) of 1.5-s-long drifting gratings, made of all possible combinations of two spatial frequencies (0.02 and 0.04 cycle/°), two temporal frequencies (2 and 4 Hz), and 12 directions (from 0° to 330°, in 30° increments); and ii) 20 different 60-s-long spatially and temporally correlated, contrast modulated, noise movies[24,25]. All stimuli were randomly interleaved, with a 1-s-long inter-

stimulus interval (ISI), during which the display was set to a uniform, middle-gray luminance level. To generate the movies, random white noise movies were spatially correlated by convolving them with a Gaussian kernel having full width at half maximum (FWHM) corresponding to a spatial frequency of 0.04 cycle/° SF. Temporal correlation was achieved by convolving the movies with a causal exponential kernel with a 33 ms decay time-constant. To prevent adaptation, each movie was also contrast modulated using a rectified sine wave with a 10 s period from full contrast to full contrast[25].

Stimuli were generated and controlled in MATLAB® (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.

Single unit isolation

Single units were isolated offline using the spike sorting package KlustaKwik-Phy[43]. Automated spike detection, feature extraction, and expectation maximization clustering were followed by manual refinement of the sorting using a customized version of the Phy interface. Specifically, we took into consideration many features of the candidate clusters: a) 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); b) the shape of the auto- and cross correlograms (important to decide whether to merge two clusters or not); c) the variation, over time, of the principal component coefficients of the waveform (important to detect and take into account possible electrode drifts); and d) 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. After the manual refinement step, we included in our analyses only units that were: i) well-isolated, i.e. with less than 0.5% of "rogue" spikes within 2 ms in their autocorrelogram; and ii) grating-responsive, i.e., with the response to the most effective grating condition being larger than 2 spikes per second (baseline-subtracted) and being larger than 6 z-scored points relative to baseline activity. 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 for the control group and 158 units for experimental group.

Quantification of selectivity

The response of a neuron to a given drifting grating was computed by counting the number of spikes during the whole duration of the stimulus, averaging across trials and then subtracting the spontaneous firing rate (computed over ISI periods). 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 in the following way:

$$\mathrm{OSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{ortho}}\right)/\left(R_{\mathrm{pref}}\right) \qquad \mathrm{DSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{opposite}}\right)/\left(R_{\mathrm{pref}}\right)$$

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.

Quantification of phase modulation (i.e. position tolerance)

Since phase shifts of a grating are equivalent to positional shifts of the whole, 2D sinusoidal pattern, a classical way to assess position tolerance of V1 neurons (thus discriminating between simple and complex cells) is to probe the phase sensitivity of their responses to optimally oriented gratings. Quantitatively, the phase-dependent modulation of the spiking response at the temporal frequency f_1 of a drifting grating was quantified by a modulation index (MI) adapted from[44] and already used in[24], 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, 10 ms-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 cells), whereas values smaller than 3 indicate poor modulation (typical of complex cells). On this ground, we adopted MI = 3 as a threshold for classifying neurons as simple or complex.

Linear receptive field estimation

We used the Spike-Triggered Average (STA) method[45,46] to estimate the linear RF structure of each recorded neuron. The method was applied to the spike trains fired by neurons in response to the spatiotemporally correlated and contrast modulated noise movies described above. To account for the correlation structure of our stimulus ensemble and prevent artifactual blurring of the reconstructed filters, we "decorrelated" the raw STA images by dividing them by the covariance matrix of the whole stimulus ensemble[45,46]. We used Tikhonov regularization to handle covariance matrix inversion. Statistical significance of the STA images was then assessed pixel-wise, by applying the following 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. The temporal span of the spatiotemporal linear kernel we reconstructed via STA extended from 330 ms before to the moment of spike generation (corresponding to 10 frames of noise at 30 Hz frame rate). The STA analysis was performed on downsampled noise frames (16x32 pixels) and the resulting filters were later spline-interpolated at higher resolution for better visualization.

To estimate the amount of signal contained in a given STA image, we used the contrast index (CI) metric that we have introduced in a previous study[24], as described in chapter I of this thesis. The CI is a robust measure of maximal local contrast in a z-scored STA image. Since the intensity values of the original STA images were expressed as z-scores (see above), a given CI value can be interpreted in terms of peak-to-peak (i.e. white-to-black) distance in σ units of the z-scored STA values. For the analysis shown in Fig. 2C, the STA image with the highest CI value was selected for each neuron.

Quantification of response slowness

For each neuron, we quantified the slowness of its response to the same noise movies used to estimate its RF, by computing the time constant of the autocorrelogram of the evoked spike trains (i.e., the probability density function of inter-spike intervals). Being the noise movies composed of richer visual patterns than drifting gratings (i.e., richer orientation and spatial frequency content), this was a way to assess the response properties of the recorded population in a slightly more naturalistic stimulation regime. The time constants τ were computed by fitting the autocorrelogram with the following exponential function:

$$f(\Delta t) = Ae^{-\frac{\Delta t}{\tau}} + C$$

where Δt is the inter-spike time interval (see Fig. 5A, bottom).

Only neurons that were strongly modulated at the frequency of variation of the contrast in the movies were included in this analysis. The level of modulation was quantified by an index similar to the one used to assess the phase-sensitivity of responses to the gratings. To this aim, we built PSTHs for the noise movies, by considering each of the 20 different movies we presented as a different trial of the same pseudo-stimulus, so as to highlight the effect of contrast modulation (see examples of highly contrast modulated neurons in Fig. 5A, top).

Orientation decoding analysis

The goal of this analysis was to build four pseudo-populations of neurons – i.e., control simple (CS), control complex (CC), experimental simple (ES) and experimental complex (EC) cells – with similar distributions of orientation tuning and orientation preference, and then compare their ability to support stable decoding of the orientation of the gratings over time.

The pseudo-populations were built as follows. We first matched control and experimental populations in terms of sharpness of their orientation tuning. To this aim, we took the OSI distributions of the two populations (i.e., the blue and orange curves in Fig. 3C) and, for each bin b in which the OSI axis had been divided (i.e., 10 equi-spaced bins of size = 0.1), we took as a reference the population with the lowest number of units N_b in that bin. For this population, all the N_b units were considered, while, for the other population, N_b units were randomly sampled (without replacement) from those with OSI falling in the bin b. Repeating this procedure for all the 10 bins, we obtained two downsampled populations of control and experimental units, having all the same OSI distribution and the same number of units (n = 92). When considering separately the pools of simple and complex cells within such downsampled populations, the resulting mean OSI were very similar (CS: 0.44 ± 0.04, n = 43; CC: 0.42 ± 0.03, n = 49; ES: 0.46 ± 0.03, n = 57; EC: 0.38 ± 0.04, n = 35) and not statistically different pairwise (two-tailed, unpaired t-test $p > 0.05$). Matching the four populations in terms of the OSI was essential, but not sufficient, to make sure that they had approximately the same power to support discrimination of the oriented gratings. In fact, the populations could still differ in terms of the distributions of orientation preference. To also equate them in this sense, and make sure that all possible orientations were equally discriminable, we replicated each unit 11 times, by circularly shifting its tuning curve of 11 incremental steps of 30º. This yielded four final pseudo-populations of 473 (CS), 539 (CC), 627 (ES) and 385 (EC) units, with matched orientation tuning and homogeneous orientation preference, to be used for the decoding analysis.

The latter worked as follows. From each pseudo-population, we sampled (without replacement) 300 units (referred to as decoding pool in what follows) and we built 300-dimensional population vectors having as components the responses (i.e., spike counts) of the sampled units in randomly selected presentations (i.e., trials) of either the 0°- or the 90°-oriented grating (drifting at 4 Hz), with each response computed in the same, randomly chosen 33 ms-wide time bin within the presentation epoch of the grating. More specifically, this time bin was chosen under the constraint of being between 561 ms and 957 ms from the onset of stimulus presentation, so that the drifting grating continued for at least 2 full cycles (i.e., 561 ms) after the selected bin. The random sampling of the trial to be used in a given population vector was performed independently for each neuron (and without replacement), so as to get rid of any noise correlation among the units that were recorded in the same session. Given that 20 repeated trials were recorded per neuron and stimulus condition, a set of 20 population vectors was built for the 0°-oriented grating and another set for the 90°-oriented gratings. These vectors were used to train a binary logistic classifier to discriminate between the two stimuli. The resulting classifier was then tested for its ability to discriminate the gratings in 33 ms-wide test bins that were increasingly distant (in time) from the training bin, covering two full cycles of the drifting gratings (i.e., from 33 to 561 ms following the training bin; see abscissa in Fig. 6B). This analysis was repeated for 50 random samplings (without replacement) of the decoding pools and, given a decoding pool, for 10 independent random draws (without replacement) of the training

time bin. The resulting 500 accuracy curves were then averaged to yield the final estimate of the stability of the classification over time (solid curves in Fig. 6B).

To obtain 95% confidence intervals (shaded regions in Fig. 6B) for such average classification curves, we ran a bootstrap analysis that worked as follows. For each of the four pseudo-populations, we sampled (with replacement) 50 surrogate populations and we used those to re-run the whole decoding analysis described in the previous paragraph. This yielded 50 bootstrap classification curves that were used to compute standard errors for the actual accuracy curve. The standard errors were then converted into confidence intervals by multiplying them by the appropriate critical value 1.96 as described in [47].

References

1. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
2. Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).
3. Karklin, Y. & Lewicki, M. S. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* **457**, 83–86 (2009).
4. Hyvärinen, A. & Hoyer, P. Emergence of Phase- and Shift-Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces. *Neural Comput.* **12**, 1705–1720 (2000).
5. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–9 (1996).
6. Hunt, J. J., Dayan, P. & Goodhill, G. J. Sparse Coding Can Predict Primary Visual Cortex Receptive Field Changes Induced by Abnormal Visual Input. *PLOS Comput. Biol.* **9**, e1003005 (2013).
7. Foldiak, P. Learning invariance from transformation sequnces. *Neural Comput.* **3**, 194–200 (1991).
8. Berkes, P. & Wiskott, L. Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis* **5**, 579–602 (2005).
9. Wiskott, L. & Sejnowski, T. J. Slow feature analysis: unsupervised learning of invariances. *Neural Comput* **14**, 715–70 (2002).
10. Cadieu, C. F. & Olshausen, B. A. Learning Intermediate-Level Representations of Form and Motion from Natural Movies. *Neural Comput.* **24**, 827–866 (2011).
11. Wallis, G. Using Spatio-temporal Correlations to Learn Invariant Object Recognition. *Neural Netw.* **9**, 1513–1519 (1996).
12. Wallis, G. & Rolls, E. T. Invariant face and object recognition in the visual system. *Prog Neurobiol* **51**, 167–94 (1997).
13. Einhäuser, W., Kayser, C., König, P. & Körding, K. P. Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur. J. Neurosci.* **15**, 475–486 (2002).

14. Körding, K. P., Kayser, C., Einhäuser, W. & König, P. How Are Complex Cell Properties Adapted to the Statistics of Natural Stimuli? *J. Neurophysiol.* **91**, 206–212 (2004).
15. Wyss, R., König, P. & Verschure, P. F. M. J. A Model of the Ventral Visual System Based on Temporal Stability and Local Memory. *PLOS Biol.* **4**, e120 (2006).
16. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How Does the Brain Solve Visual Object Recognition? *Neuron* **73**, 415–434 (2012).
17. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J Physiol* **195**, 215–43 (1968).
18. Berardi, N., Pizzorusso, T. & Maffei, L. Critical periods during sensory development. *Curr. Opin. Neurobiol.* **10**, 138–145 (2000).
19. Espinosa, J. S. & Stryker, M. P. Development and plasticity of the primary visual cortex. *Neuron* **75**, 230–249 (2012).
20. White, L. E. & Fitzpatrick, D. Vision and Cortical Map Development. *Neuron* **56**, 327–338 (2007).
21. Ohshiro, T., Hussain, S. & Weliky, M. Development of cortical orientation selectivity in the absence of visual experience with contour. *J. Neurophysiol.* **106**, 1923–1932 (2011).
22. Wood, J. N., Prasad, A., Goldman, J. G. & Wood, S. M. W. Enhanced learning of natural visual sequences in newborn chicks. *Anim. Cogn.* **19**, 835–845 (2016).
23. Wood, J. N. & Wood, S. M. W. The Development of Invariant Object Recognition Requires Visual Experience With Temporally Smooth Objects. *Cogn. Sci.* **42**, 1391–1406 (2018).
24. Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B. & Zoccolan, D. Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J. Neurosci.* **39**, 1649–1670 (2019).
25. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J Neurosci* **28**, 7520–36 (2008).
26. Li, N. & DiCarlo, J. J. Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex. *Science* **321**, 1502–1507 (2008).
27. Li, N. & DiCarlo, J. J. Unsupervised Natural Visual Experience Rapidly Reshapes Size-Invariant Object Representation in Inferior Temporal Cortex. *Neuron* **67**, 1062–1075 (2010).
28. Zador, A. M. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **10**, 3770 (2019).
29. Li, Y. *et al.* Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature* **486**, 118–121 (2012).
30. Hagihara, K. M., Murakami, T., Yoshida, T., Tagawa, Y. & Ohki, K. Neuronal activity is not required for the initial formation and maturation of visual selectivity. *Nat. Neurosci.* **18**, 1780–1788 (2015).
31. Albert, M. V., Schnabel, A. & Field, D. J. Innate Visual Learning through Spontaneous Activity Patterns. *PLoS Comput. Biol.* **4**, e1000137 (2008).
32. Wong, R. O. L. RETINAL WAVES AND VISUAL SYSTEM DEVELOPMENT. *Annu. Rev. Neurosci.* **22**, 29–47 (1999).

33. Thompson, I. Cortical development: A role for spontaneous activity? *Curr. Biol.* **7**, R324–R326 (1997).
34. Dähne, S., Wilbert, N. & Wiskott, L. Slow Feature Analysis on Retinal Waves Leads to V1 Complex Cells. *PLoS Comput Biol* **10**, e1003564 (2014).
35. Creutzig, F. & Sprekeler, H. Predictive Coding and the Slowness Principle: An Information-Theoretic Approach. *Neural Comput.* **20**, 1026–1041 (2007).
36. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci.* **115**, 186–191 (2018).
37. Weghenkel, B. & Wiskott, L. Slowness as a Proxy for Temporal Predictability: An Empirical Comparison. *Neural Comput.* **30**, 1151–1179 (2018).
38. Bialek, W., Steveninck, R. R. D. R. V. & Tishby, N. Efficient representation as a design principle for neural coding and computation. in *2006 IEEE International Symposium on Information Theory* 659–663 (2006). doi:10.1109/ISIT.2006.261867.
39. Lies, J.-P., Häfner, R. M. & Bethge, M. Slowness and Sparseness Have Diverging Effects on Complex Cell Learning. *PLOS Comput. Biol.* **10**, e1003468 (2014).
40. Wood, J. N. Newborn chickens generate invariant object representations at the onset of visual object experience. *Proc. Natl. Acad. Sci.* **110**, 14000–14005 (2013).
41. Williams, R. A., Pollitz, C. H., Smith, J. C. & Williams, T. P. Flicker detection in the albino rat following light-induced retinal damage. *Physiol. Behav.* **34**, 259–266 (1985).
42. Matteucci, G., Riggi, M. & Zoccolan, D. A template-matching algorithm for laminar identification of cortical recording sites from evoked response potentials. *bioRxiv* 749069 (2019) doi:10.1101/749069.
43. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).
44. Wypych, M. *et al.* Standardized F1 – A consistent measure of strength of modulation of visual responses to sine-wave drifting gratings. *Vision Res.* **72**, 14–33 (2012).
45. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *J. Vis.* **6**, (2006).
46. Sharpee, T. O. Computational Identification of Receptive Fields. *Annu. Rev. Neurosci.* **36**, 103–120 (2013).
47. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap*. (Chapman & Hall/CRC, 1998).

Chapter IV

Behavioral evidence of spontaneous motion integration in rats

Abstract

Computing the global motion direction of an extended visual object made of multiple oriented elements is an higher-order visual function extensively studied in human and non-human primates[1]. However, dissecting the underlying neuronal circuits will likely require the molecular tools that are now largely used in rodent studies. Recent physiological work[2,3] reported the presence of neurons selective for global motion in mouse visual cortex. Direct psychophysical evidence that the visual system of rodents is capable of combining multiple motion signals into global, integrated percepts is, however, still lacking. To address this question, we trained two groups of rats to discriminate the motion direction of either gratings or plaids (i.e., superpositions of gratings with different directions) and we tested whether these visual patterns, when used as prime stimuli, were able to bias rat perception in a way that is consistent with an integrated representation of the plaids. More specifically, our experiments allowed inferring whether, and under which conditions, the representation of these two types of stimuli is shared (i.e., coded by overlapping neuronal populations) and, if so, whether its organizing principle is the local or the global direction of the stimulus. We found that, depending on the identity of the training stimuli, rats displayed either a shared representation based on global motion, or a non-shared one. These results highlight the complexity of rat motion perception and provide the first direct behavioral evidence about the ability of a rodent species to spontaneously extract global motion information from complex visual patterns.

Introduction

A crucial step in processing visual motion information is the computation of the direction of motion of visual objects. Such computation has been widely studied in the dorsal stream of primates, and, in particular, in the middle temporal area (MT) of the macaque[1]. This area

receives direct input from neurons in primary visual cortex (V1), signaling the "local" direction of moving elements (e.g., edges) in small, localized, regions of the visual field – i.e., the receptive fields (RFs). MT units appear to combine the afferent V1 inputs in such a way to acquire selectivity for the "global" direction of motion of visual objects (or patterns) made of multiple local edges[4].

The leading hypothesis emerging from the primate literature is that this is achieved by integrating the local motion signals carried by V1 afferents over different spatial positions and directions[1,4,5]. Notably, such computation is necessary to extract the direction of complex objects, made of multiple oriented features. The local output of any V1-like edge detector is, in fact, intrinsically ambiguous on its own, being compatible with infinite combinations of global object directions and speeds. This ambiguity is at the core of what is known, in the psychophysics literature, as the "aperture problem"[6–8]: only by integrating information from multiple moving-edge detectors, it is possible to infer global motion direction. Understanding the precise circuit and synaptic mechanisms underlying such computation remains a key question in system neuroscience.

Over the past 10 years, rodents have emerged as interesting model systems to study visual cortical processing, given the wide array of experimental tools for the functional dissection of neural circuits that are available in mice[9–13], combined with the demonstration of sophisticated visual behaviors in rats[14]. This led some investigators to look for the signature of motion integration in the rodent brain[2,3,15]. Inspired by monkey studies[16–18], they recorded the responses of mouse visual cortical neurons to drifting gratings and coherent plaids − i.e., complex visual patterns made of two overlapping gratings with the same contrast and speed, but moving along independent directions.

These stimuli have been widely used to investigate motion integration in human and non-human primates since they enable to clearly distinguish local- from global-based motion responses. A plaid, in fact, has two well-defined local motion components and a global direction that is different from the local ones[5], while, for a grating, the local and global directions coincide[5]. By employing these stimuli, mouse visual neurons could be classified as "pattern" (i.e., responsive to global motion) or "component" (i.e., responsive to local motion). The majority of neurons amenable to such classification fell into the component category, but a small fraction of pattern units was reported in V1[3,15] and in two extrastriate areas, the lateromedial (LM) and rostrolateral (RL) areas[2]. However, the relevance of such tiny population of pattern cells in determining mouse motion perception was left untested (but see further discussion of [3] below).

More in general, even if a few studies tested the ability of rats and mice to discriminate the dominant motion direction of random dot kinematograms[19–22] (RDKs; a task linked to motion integration, since it requires spatial integration of local motion cues), no study ever probed the ability of rodents to integrate into a global motion percept multiple oriented elements moving along different directions. In addition, all the studies employing RDKs explicitly trained the animals in the direction discrimination task.

Critically, this prevented assessing the spontaneous, purely perceptual motion integration abilities of naïve rodents. In fact, if feedback (e.g., reward) is provided to an animal about the outcome of his perceptual choices, it may learn to associate a given stimulus to a given response category without necessarily perceiving the sensory quality that the

investigator intends to associate to that category[14,23,24]. For instance, an animal could learn to classify in the same category a rightward drifting grating and a rightward drifting plaid, even if it perceives the latter in a non-integrated way (i.e., as the superimposition of two independently drifting gratings), simply because of the reinforcement received during training. It would thus be impossible to infer from such experiments whether the choices of the animal are driven by a general-purpose, stimulus-independent representation of global motion direction at the level of visual cortex (as the one found in monkey MT), or by the ability to learn abstract representations of arbitrary stimulus categories in higher-order association areas involved in perceptual decision making, such as posterior parietal cortex (PPC)[25–28].

Only in [3] this confound was avoided, by monitoring the direction of the optokinetic nystagmus (OKN) while presenting plaids to naïve, head-fixed mice. The resulting bimodal distribution of OKN movements tracking both the local-components and the overall global directions of the stimulus is suggestive of mouse ability to spontaneously perceive global motion. However, this conclusion is affected by the intrinsic limitation of OKN-based experiments to probe cortical processing. In fact, OKN is a reflexive phenomenon that is known to be largely controlled by subcortical structures[29–31], while motion perception in direction discrimination tasks has been shown to rely heavily on visual cortex[22,32,33]. As a result, it remains untested whether rodents are capable of combining independent but coherent motion signals (as those found in plaids) into integrated percepts of global motion direction.

Our study was designed to address this question and provide a thorough psychophysical assessment of rat ability to spontaneously perceive the global direction of drifting plaids. To this aim, we relied on a visual priming paradigm[23] that allowed measuring the perceptual similarity between gratings and plaids with the same global direction, without the need of explicitly training the animals to associate these stimuli to the same response category.

Results

In the first phase of the study, a group of 11 male Long-Evans rats (referred to as the G group in what follows) was trained to discriminate leftward- from rightward-drifting gratings (Fig. 1A, left), using the high-throughput behavioral rig described in[14] and previously employed in several investigations of rat object recognition[23,34–37]. Another group of 10 rats (referred to as the P group in what follows) was trained to discriminate leftward- from rightward-drifting plaids, made of the superposition of two gratings, whose directions were 120º apart (Fig. 1B, left). Rats that maintained at least a criterion performance of 75% correct choices over a four consecutive day period were moved to the second phase of the study to be tested with the priming paradigm (all rats starting the training phase managed to reach such criterion). 10 out of 11 G rats and 9 out of 10 P rats were included in the subsequent analysis displaying a strong identity prime effect (see Materials and Methods).

In priming phase, the “target” stimulus remained the same as in training phase (i.e., G and P rats still had to report the direction of drifting gratings and plaids, respectively), but

now the presentation of the target was preceded by the brief presentation (75 ms) of either a grating or a plaid drifting along one out of 19 possible directions (from 0° = rightward to 180° = leftward, in steps of 10°; Fig. 1 A-B, right). The identity (i.e., either grating or plaid) and the motion direction of such "prime" stimulus was randomly selected in each trial.

We refer to trials in which the identities of prime and target coincide as "identity-priming condition" (e.g., a grating prime stimulus followed by a grating target stimulus), whereas we refer to the opposite case as "cross-priming condition" (e.g., a plaid prime stimulus followed by a grating target stimulus).

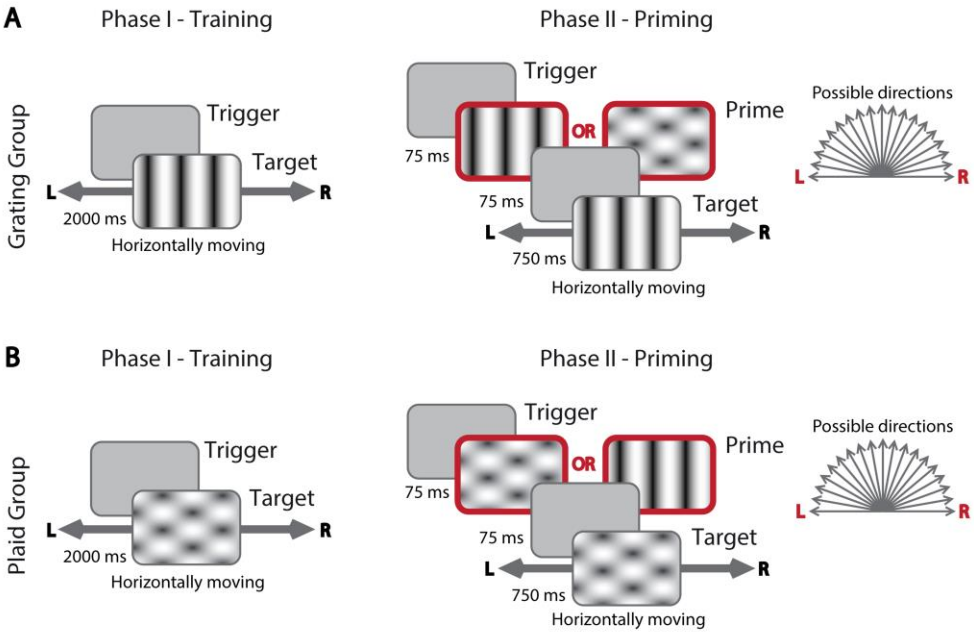

Figure 1. Task structure. A) Grating group. During both training and priming phase, trials are started when animals trigger the central port of a 3-way licking sensor. During the training phase (left), the target grating (either leftwards or rightwards moving) is presented immediately at the start of each trial. In order to receive reward, animals have to reach the lateral port, corresponding to the direction of motion of the presented target. During the priming phase (right), the structure of the task remains similar, but at the start of each trial, before the appearance of the target, a moving prime stimulus is shortly presented (75 ms duration vs. 750 ms of the target). The prime stimulus can move in 19 directions (from 0° = right to 180° = left) and can be either a grating or a plaid. B) Plaid group. The structure of the task is the same described for the Grating group in both phases. The key difference is that the target stimulus is always a plaid instead of a grating. Reward is delivered when the animal reaches the lateral port consistent with the global direction of the target.

Critically, the rats kept receiving feedback (i.e., either reward or a time-out period, in case of correct and incorrect choices, respectively) only about the correctness of their responses to the target stimuli. The identity and direction of the prime stimuli were never paired to either the leftward or rightward response categories. As such, the extent to which the prime stimuli were able to affect the choices of the rats was purely due to the spontaneous, perceived similarity between primes and targets. Assessing such similarity was the ultimate goal of our experiments, since it allowed understanding whether, in cross-prime condition, the perceived motion direction of plaids was their global one.

The rationale of this design rests on well-established findings in the human psychophysics literature[38–41], where presenting an "adapter" before a "test" stimulus has been

shown to bias the perception of the test, depending on two key factors: the duration of the adapter and the inter stimulus interval (ISI) between adapter and target presentation.

When the bias induced by the adapter attracts the perceptual choices towards the identity of the adapter itself, this effect is called "priming". Vice versa, when it repels the perceptual choices away from the identity of the adapter, it is named "adaptation after-effect*"*. Our group has previously shown that brief presentation of a static shape (~50 ms), followed by a short ISI (66 ms), is able to induce a strong and robust priming effect on rat response to a target shape[23]. Here we extended this paradigm to moving stimuli, and we relied on previous motion adaptation studies in humans[40] to select the timing parameters of the task, in the attempt of inducing a strong priming effect**.**

In the following we will refer to the performance in the target direction discrimination as a function of the direction coherence of a prime stimulus (i.e. how close the direction of the prime is to the target one) as "priming curve". In other words, by "direction coherence" we mean the degree to which the direction of prime and target is the same. Consistently, we will refer to the extremes of such curves as "coherent" or "incoherent" priming, respectively. We will call "neutral" priming the mid-point of such curves, consisting in the performance when the prime carried no left/right-directional information because it moved vertically. Such terminology will be used both for identity- and cross-priming conditions, for both P and G groups.

The first step in our analysis was to verify whether the prime stimuli were indeed capable of biasing rat choices consistently with their direction coherence in the identity-prime condition.

For the G group, performances in the identity-priming condition (see Fig. 2A, left, black curve), were strongly modulated by the direction coherence of the prime grating. In the coherent priming case, rat classification of target direction was strongly facilitated, as compared to the reference neutral priming condition. Conversely, in the incoherent priming condition, rat classification accuracy was substantially lower than with the reference neutral prime. Overall, the identity-priming curve was approximately sigmoidal with the accuracy depending smoothly from the direction coherence of the identity-prime stimulus.

Also for the P group, in the identity-priming condition, a strong priming was observed (see Fig. 2A, right; black curve). In this case too, the magnitude and sign of the priming effect depended smoothly on the similarity between the direction of the prime and that of the target, resulting in an approximately monotonic drop of rat classification accuracy from the fully coherent priming condition towards the fully incoherent one.

The observed strong modulation of rat classification accuracy in identity-priming condition is important for two reasons. First, it verified the effectiveness of the priming paradigm. Second, and more importantly, it served as a reference against which to compare the modulation of the accuracy observed in the cross-priming condition. In fact, the degree of similarity between the priming curves measured in identity- and cross-priming condition (i.e., between the black and gray curves in Fig. 2A) allowed answering two key questions about the neuronal representations of stimuli (i.e., local or global direction-based).

The first question that will be addressed focusing on the cross-priming curves is whether these representations are "shared" (i.e., the same neural population encodes both gratings and plaids) or "non-shared" (i.e., distinct, non-overlapping neural populations

represent the two types of stimuli). In the former case, we would expect a substantial modulation of rat classification accuracy also in the cross-priming condition, possibly as large as in identity-priming condition. In the latter case, no cross-priming effect would be possible (because of no interaction between the populations representing the two stimuli).

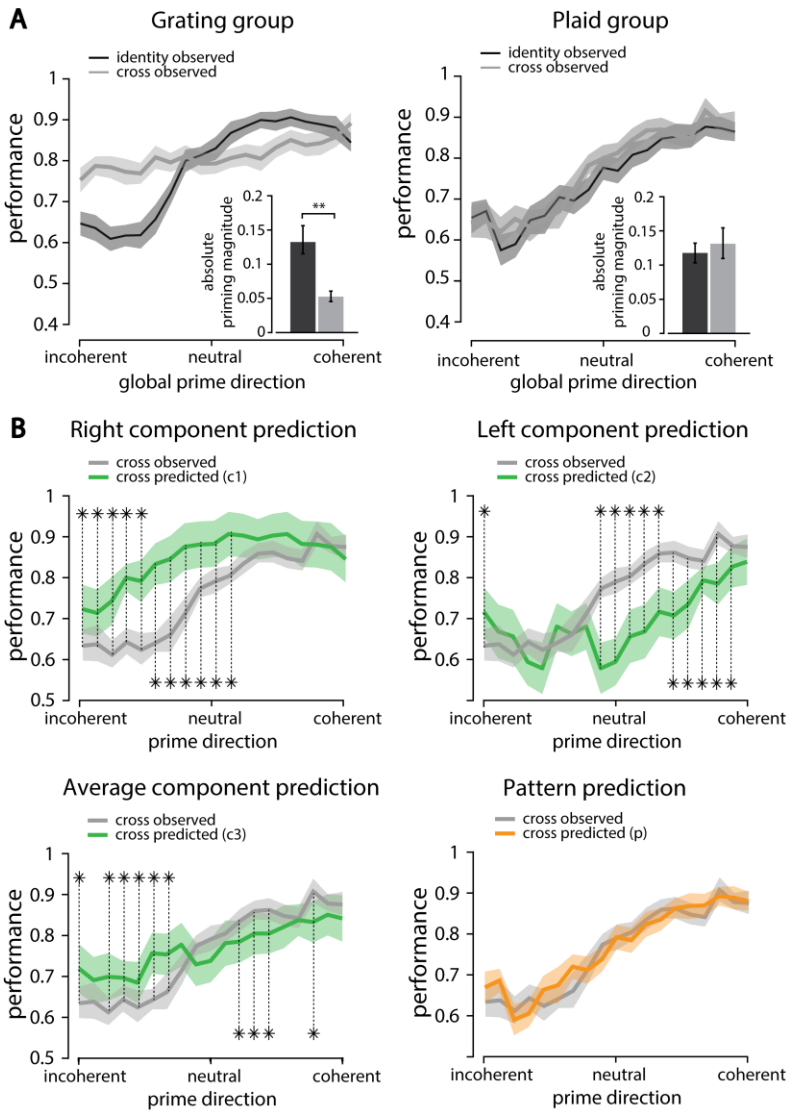

Figure 2. Observed and predicted priming curves. A) Observed identity- (black) and cross- (gray) priming curves for the G (left) and P (right) groups. Insets show absolute prime effect magnitudes in each group. Notice in the G group case (left) a significant difference between identity and cross conditions (** p<0.01, t-test). B) Observed (gray) and predicted (colored) cross-priming curves. Top and bottom-left subpanels compare predictions obtained assuming a local-direction-based representation of prime stimuli (in green) with the observed data (in gray). Bottom-right subpanel compares prediction obtained assuming a global-direction-based representation of prime stimuli (in orange) with the observed data (in gray). Asterisks and dashed lines highlight directions for which the difference between observed and predicted cross-priming curves is not compatible with 0 (* p<0.05, bootstrap-estimated). In all panels, shaded areas represent bootstrap-estimated 95% confidence intervals whereas solid lines are group average priming curves.

In case the representations of drifting gratings and plaids were found to be shared, then a second, deeper question could be answered by comparing the shapes of the priming curves obtained in identity- and cross-priming conditions. If rats spontaneously represented stimuli based on their global direction, then the priming curves produced by gratings and

plaids would be largely overlapping. If the contrary were true (stimuli represented on the basis of local direction), priming curves produced by gratings and plaids would be different. The latter case is the one to be expected if rats, when processing plaid stimuli, were "blind" to their global motion direction and extracted instead only the local directions of the constituent gratings.

This qualitative intuition can be translated into quantitative hypotheses about the possible shapes of the cross-priming curve, based on the most plausible ways (i.e., "combination rules") by which, in presence of a non-integrated (i.e. local-direction based) percept, the prime stimuli could exert their priming influence on the discrimination of the target. All the hypotheses formulated below are grounded on the assumption that, in each trial, one component of a plaid would prevail in biasing rat perception of the target towards its direction, consistently with a form of winner-take-all competition among overlapping components[42] (the standard alternative to integrative strategies).

In the first and simplest scenario, a rat, when presented with a plaid, always perceives (or is influenced by) the same component of the stimulus – either the more leftward- or the more rightward-drifting component grating (where with leftward or rightward we mean with respect to the global direction of the plaid). As a result, the cross-priming curve can be obtained from the identity-priming curve, by simply shifting the latter either leftward ("left component prediction") or rightward ("right component prediction") by 60º, i.e., half the angle between the two components of the plaid (see Materials and methods and Fig. 6).

An alternative (and more likely) scenario is that a rat perceives (or is influenced by) just one of the two components of a plaid, but this component randomly changes across the trials in which the plaid is used as a prime. Such combination rule would produce a cross-priming curve (referred to as "average prediction" in what follows) obtainable as the average of the left and right component predictions.

Finally, a quantitative prediction about the shape of the cross-priming curve in case of integration of the local components into a global motion percept is simply given by the identity-priming curve itself.

To find out which of the above-mentioned hypotheses better matched rat perceptual behavior, we compared the corresponding predictions with the cross-priming curves that were observed experimentally (computed by averaging over all the rats within a group). We estimated 95% confidence intervals for the group average priming curves by bootstrapping over sessions (i.e. ensemble of all trials performed in a given day), independently for each animal (see Materials and methods). Observed cross-priming curves are shown in Fig. 2A.

In the G group (Fig. 2A, left), the cross-priming curve (gray) is almost flat and significantly different from the identity-priming curve (black), as shown by the non-overlapping 95% confidence intervals (shaded areas). This means that the plaids, when used as prime stimuli, were virtually ineffective at biasing rat discrimination of the target gratings. Magnitude of priming effects were quantified by measuring the mean absolute difference between the value of the curve at the neutral prime direction and the values of the curve at the four leftmost and four rightmost prime directions (i.e., close to the most incoherent and most coherent priming conditions).

The priming magnitude was found to be significantly smaller than in the identity-priming case (Fig. 2A, inset, left, gray vs. black bar; ** $p < 0.01$; t-test). This suggests that

the neuronal population which G rats relied upon to represent the direction of the (isolated) gratings was largely inactive during presentation of the plaids, and thus unable to encode the global direction of these stimuli, as well as the local directions of their components.

By contrast, in the P group, the cross-priming and identity-priming curves largely overlapped (Fig. 2A, right: gray vs. black curve). As a result, the priming magnitude observed in the two conditions was equally large (Fig. 2A, inset, right: gray vs. black bar; $p > 0.05$; t-test). This can be interpreted as the signature of a shared representation, where the same neural population codes for both kinds of stimuli (see Discussion).

As explained above, the existence of such a shared representation allowed further investigating its organizing principle (local or global direction). To this aim, in Fig. 2B we quantified the overlap between the observed cross-priming curve (in gray) and each of the predicted cross-priming curves resulting from assuming either a local representation of the individual components of the plaids (including all combination rules discussed above, in green) or an integrated representation of its global motion (in orange). The right and left component predictions (top subpanels, green) diverged substantially from the observed cross-priming curve (gray), with 60% of points being significantly different from what expected by chance ($p < 0.05$; test based on bootstrapping the confidence interval of the difference between predicted and observed curves; see Materials and methods and Fig. 6). The average component prediction (bottom left subpanel, green) yielded a cross-priming curve that is closer to the observed one (gray), yet 55% of the points are significantly different between the two curves. By contrast, the prediction based on global motion integration yielded a cross-priming curve (bottom right subpanel, orange) that largely overlapped with the observed one (gray), with no points being significantly different between the two curves.

Based on these results, we can infer that the representation of plaids in rats trained to discriminate such stimuli (P group) is not only shared with the representation of gratings, but is structured in such a way to encode global motion direction.

Discussion

The results presented in Fig. 2 show that rats are capable of spontaneously combining local motion cues into integrated percepts of global motion direction of a complex visual pattern. This establishes a candidate perceptual correlate of the neuronal selectivity for global motion, previously reported across rodent visual cortical areas[2,3,15].

Critically, such motion integration ability only emerged in rats trained to discriminate the plaid stimuli (P group). The animals trained in the grating discrimination task (G group) were instead virtually insensitive to the plaids, indicating that the neuronal population they relied upon to encode the gratings failed to effectively represent the plaids. This suggests two intriguing hypotheses about the representational and perceptual decision processes underlying motion discrimination in rats.

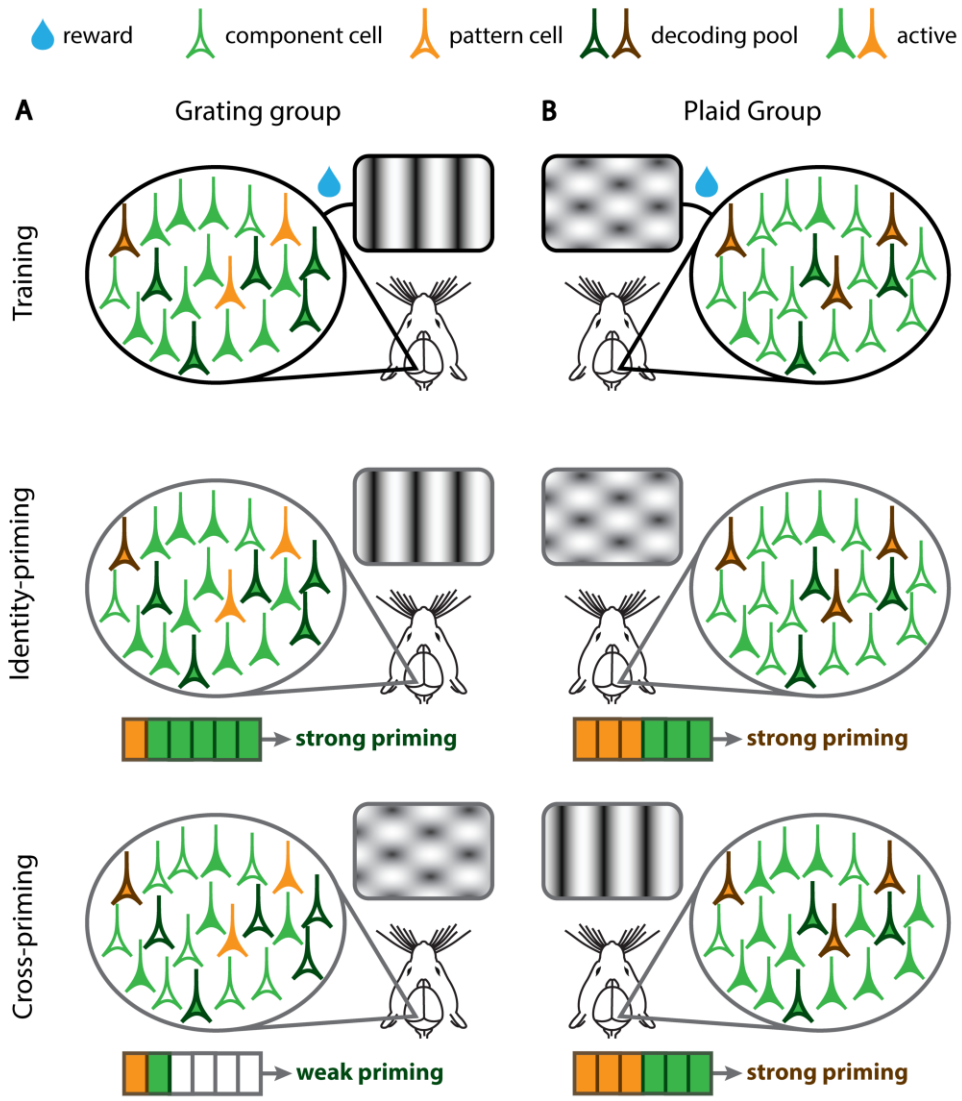

Figure 3. Mechanism proposed to explain the inter-group difference in absolute cross-priming effect magnitudes. A) Grating group. During training, the neuronal ensemble representing the target stimulus (color filled neurons) is coupled with a reward following correct trials. We posit that reward-gated Hebbian plasticity mechanisms recruit an unbiased subset of those neurons in the decoding pool (dark-bordered neurons) supporting the perceptual discrimination in well-trained animals. In presence of a cross-orientation-suppression effect affecting mainly component cells, the latter category of neurons will respond less to the cross-prime plaid stimulus than to the identity-prime grating one (notice more empty green neurons on the bottom subpanel). A weak activation of the decoding pool neurons (as symbolized by the stack with many empty rectangles) in the cross-priming case will likely result in reduced priming efficacy in this condition (as observed in our data). B) Plaid group. During training, due to the asymmetric responsivity to the plaid target of pattern and component cells (induced by cross-orientation suppression) reward-gated plasticity will likely recruit a pattern-biased subset of neurons in the decoding pool (dark-bordered neurons). This time in the cross-priming condition (as well as in the identity-priming one, since the decoding pool neurons were already selected among the most plaid responsive), the decoding pool will not be weakly activated (as symbolized by the stacks full of color-filled rectangles). This will result in both a strong identity- and cross- priming effect consistent with a global-direction based representation (as observed in our data).

The first is that the specific discrimination being reinforced, through reward delivery, during training, biases the recruitment of the neuronal pool within visual cortex that is read out by downstream decision neurons towards either a more component-enriched or a more pattern-enriched cell population (referred to as the “decoding pool” in what follows).

The second hypothesis is that component cells not only fail to represent the global direction of the plaids, but signal the local directions of their constituent gratings less reliably than when such gratings are presented alone (as it would happen if having a lower response firing rate in presence of Poisson variability of firing).

We checked whether these hypotheses could be supported by the known response properties of component and pattern cells, focusing on two specific questions: 1) the relative proportion of these functional cell classes within rodent visual cortex and 2) how vigorously these two classes of neurons respond to gratings and plaids.

Although the number of studies investigating global motion integration in rodent visual cortical areas is still limited[2,3,15], a finding that is consistent among studies is that component cells by far outnumber pattern cells, with the former being ~25% of direction-tuned units and the latter being about ~5% (with the rest being unclassified).

With respect to the second question, a well-known phenomenon likely affecting the strength of grating and plaid responses is cross-orientation suppression[43,44], which has been recently documented also in rodents[45–47]. It consists in the reduction of the response to a preferred oriented stimulus, when this is presented along with other stimuli with different orientations. In principle, cross-orientation suppression could lead to diminished responses to plaids, as compared to gratings. However, studies documenting this phenomenon mainly targeted V1, without distinguishing between pattern and component cells. When focusing on the motion integration literature, it turned out to be difficult to draw any useful data concerning the presence of such effect in pattern and component cells from rodent studies[2,3,15], because of methodological issues (e.g., tuning curves normalization, and calcium imaging lacking spike-count-level information). On the other hand, in the primate literature, we found hints of cross-orientation suppression being stronger in component cells than in pattern cells[16,17]. If one assumes that the same cross-orientation suppression also affects rodent component cells, this phenomenon, by itself, could support the second hypothesis formulated above – a decoding pool largely made of component cells would respond less (and likely less reliably) to plaids than to gratings. Crucially, this would explain the reduced priming efficacy of plaid cross-prime stimuli over target gratings that we observed in the G group. It remains to be explained, however, how training with gratings, rather than with plaids, can lead to a component-rich, rather than to a pattern-enriched, decoding pool in the first place (i.e., the first hypothesis formulated above). Intuitively, this question can be addressed by taking into account the effect of both the scarcity of pattern cells and their lower susceptibility to cross-orientation suppression, as compared to component cells. This is graphically illustrated in Fig. 3, where the pattern and component cells are represented as stylized pyramidal neurons with different colors (orange for the pattern and green for the component).

In the case of training with gratings (Fig. 3A, top), these stimuli will strongly activate both pattern and component cells (filled neurons). As such, the selection of the neuronal decoding pool by a downstream decision neuron will not be biased towards either class of neurons. However, since, in the overall population, the component cells substantially outnumber the pattern cells, the units contributing to the decoding pool (dark-boundary neurons) will mostly be component cells. As already mentioned, all the neurons in the decoding pool will be active during presentation of the grating primes (Fig. 3A, middle; filled

neurons). By contrast, most component cells in the pool will be strongly depressed by cross-orientation suppression during presentation of the plaid primes (Fig. 3A, bottom; empty dark-boundary neurons), and the few pattern cells in the pool (orange-filled, black-boundary neurons), although active, will not be enough to support the representation (and the percept) of the pattern stimuli. Hence the lack of cross-priming in the group of rats trained with gratings (Fig. 2A, left).

In the case of training with plaids (Fig. 3B, top), these stimuli will strongly activate the population of pattern cells (orange-filled neurons) but weakly excite component cells, activating only a small fraction of them (green-filled neurons) in each trial. This asymmetry would lead to a strong selection bias favoring the recruitment of pattern cells in the decoding pool (dark-boundary neurons) by a downstream decision neuron. This is consistent with the idea of such process being mediated by some form of plasticity similar to the reward-gated Hebbian one described in [48] (since stronger firing neurons at time of reward delivery will be more likely to be included in the decoding pool). When tested with prime stimuli, such pattern-enriched pool would respond with a similar intensity to both kinds of primes (grating and plaids; Fig. 3B, middle and bottom panels respectively), leading to similar priming curves in both identity- and cross-priming conditions (Fig. 2A, right). This would account for the shared representation of global motion direction observed for P group rats.

To quantitatively test whether the mechanistic account hypothesized in the previous paragraphs could explain our findings, we built a computational model where a decision neuron has to select his afferent units from a mixed population of component and pattern cells in the visual cortex, in order to discriminate the direction of either drifting gratings or plaids (Fig. 4A). The relative proportion of simulated component (80%) and pattern (20%) cells roughly matched that found, on average, across mouse visual areas V1, LM and RL[2,3,15]. Each unit was simulated as a Poisson spiking neuron, whose average firing rate as a function of direction was defined as a von Mises function[49,50] having a peak that was randomly sampled across the direction axis (see Materials and methods for details). In addition, to simulate cross-orientation suppression in component cells, their peak responses to the plaids were set to half their peak responses to the gratings, while pattern cells were assumed to be unaffected by cross-orientation suppression. The decision neuron consisted in a logistic regression classifier that was trained to discriminate between leftward and rightward stimuli. Such classifier performs a weighted sum of the inputs provided by all the simulated component and pattern cells. With training, the weights are adjusted in such a way to maximize its classification accuracy (eventually attaining 100%).

When trained in the grating discrimination task (reproducing G group training condition), the classifier learned to rely more heavily on those motion detectors (no matter whether component or pattern cells) preferring directions closer to that of the training stimuli. This is shown in Fig. 4C (top), where the weights of the simulated units as a function of their preferred direction are reported. For both the component (green dots) and pattern (orange dots) cell populations, the most informative neurons were those with preferred direction close to 0º/360º (rightward-drifting gratings) and 180º (leftward-drifting gratings), while the units tuned around the vertical drift direction (i.e., 90º and 270º) were assigned close-to-zero weights.

Given these weight distributions, we simulated different degrees of sparsity constraints on the connectivity of the decision neuron. This was done by pruning input connections with progressively larger weight magnitude, thus gradually reducing the effective size of the decoding pool.

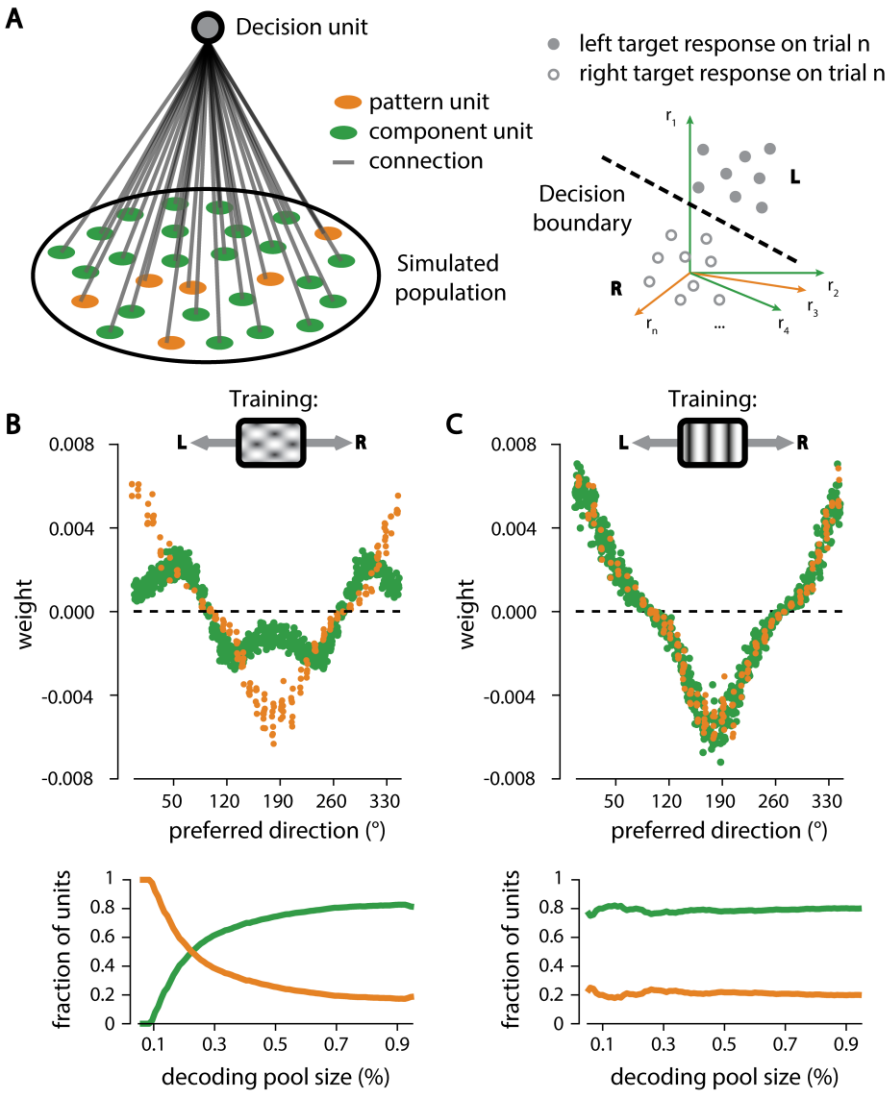

Figure 4. Computational proof-of-principle of the proposed decoding pool-selection mechanism. A) Cartoon illustrating the structure of the simulation. A "decision unit" (left subpanel, gray circle) consisting in a logistic regression classifier is trained to discriminate leftward from rightward moving target on a trial-by-trial basis (right subpanel). This is done by reading out the representation of the stimulus provided by the activity of a simulated population of Poisson-firing ideal pattern and component units (left subpanel, colored circles). Upper subpanels of B) and C) display the distribution of classifier weights after training. Weights are shown organizing neurons according to their preferred direction. In the grating training case (C) weight values are similar for pattern (orange) and component units (green). In the plaid training case (B), as expected due to their different tuning, pattern and component units show maximal weight values at different preferred directions. Crucially, maximal weight values themselves differ (i.e. patterns reach twice the maximal components weight value). Lower subpanels of B) and C) show the fraction of pattern and component units in the decoding pool at different pruning threshold levels (resulting in different final decoding pool size). In the grating training case (C), for any pruning level, the fraction of pattern and component units in the decoding pool is very similar to the original population (i.e. 20% and 80%). In the plaid training case (B), the different magnitudes of peak weights for patterns and components cause a reversal of their proportion in the decoding pool at high pruning levels (i.e. at low decoding pool size).

However, given the closely matching weight magnitudes of the component and pattern cells, the level of pruning did not alter their relative proportion in the decoding pool – this proportion remained the same as in the original population (i.e., 80% vs. 20%; Fig. 4C, bottom).

This confirms the intuition (see Fig. 3A) that training with gratings, even in presence of asymmetric (i.e. acting on component but not pattern cells) cross-orientation suppression, leads to no bias in resulting decoding pool composition (that is therefore component-rich).

A different scenario emerged when we simulated the training in the plaid discrimination task (reproducing P group training condition, see Fig. 4B). The weight distribution learned by the classifier for the pattern cell population (left, orange dots) was the same as in the grating discrimination task. This is because simulated pattern cells are, by hypothesis, immune to cross-orientation suppression. On the other hand, the weight distribution for the component cell population (green dots) was very different from that obtained in the grating discrimination task. First, since these cells responded to the plaid components (i.e., the constituent gratings), the classifier learned to assign larger weights to those units with preferred direction at ±60º (i.e. half plaid cross-angle) with respect to the global directions of the leftward- and rightward-drifting plaids. In addition, because the components cells are affected by cross-orientation suppression, the classifier learned to rely less on component than on pattern cells, thus assigning much lower weights to the former than to the latter (compare the absolute height of the peaks in the distributions of green and orange dots in Fig. 4B, top). As a result, when the sparsity of the connections to the decision neuron was increased, by leaving in the decoding pool only inputs with progressively larger weight magnitude (i.e. performing connection pruning), the proportions of component and pattern cells in the pool followed opposite, mirror trends (Fig. 4B, bottom). While the former decreased (orange curve), the latter increased (green curve), eventually leading to a reversal of the fraction of component and pattern cells in the decoding pool, when the decision neuron was allowed to keep only ~10% of its potential inputs. This confirms the intuition (see Fig. 3B) that cross-orientation suppression can, at least in principle, strongly bias the recruitment of pattern cells in the decoding pool, under the constraint of sparse connectivity to the decision neurons (leading to a pattern-enriched decoding pool).

Furthermore, to demonstrate that cross-orientation suppression acting over component cells (i.e. asymmetric) was actually necessary and sufficient to produce the above discussed decoding pool selection bias effect, we repeated the simulation described above "knocking-out" the cross-orientation suppression effect itself. As can be seen in Fig. 5, the result of such computational control experiment was, as expected, completely identical to the previous case for what concerns the grating training (Fig. 5B).

On the other hand for what concerns the plaid training (Fig. 5A) the weights magnitude difference and decoding pool selection effects obtained in presence of asymmetric cross-orientation suppression were now completely abolished.

Overall, these results not only provide solid behavioral evidence of motion integration in rodents, but also allow making explicit, testable predictions about the neurophysiological mechanisms underlying such ability – i.e., the existence of cross-orientation suppression in component but not in pattern cells. More in general, our study provides solid grounding for future investigations aimed at dissecting the neuronal circuits underlying integration of local

motion cues into global motion percepts. In fact, as highlighted in a recent perspective[51], a behavioral task involving direction discrimination of gratings and plaids is a necessary ingredient of future experiments aimed at establishing a correlational and/or causal link between pattern-like responses observed in rodent visual cortex and integrated motion perception.

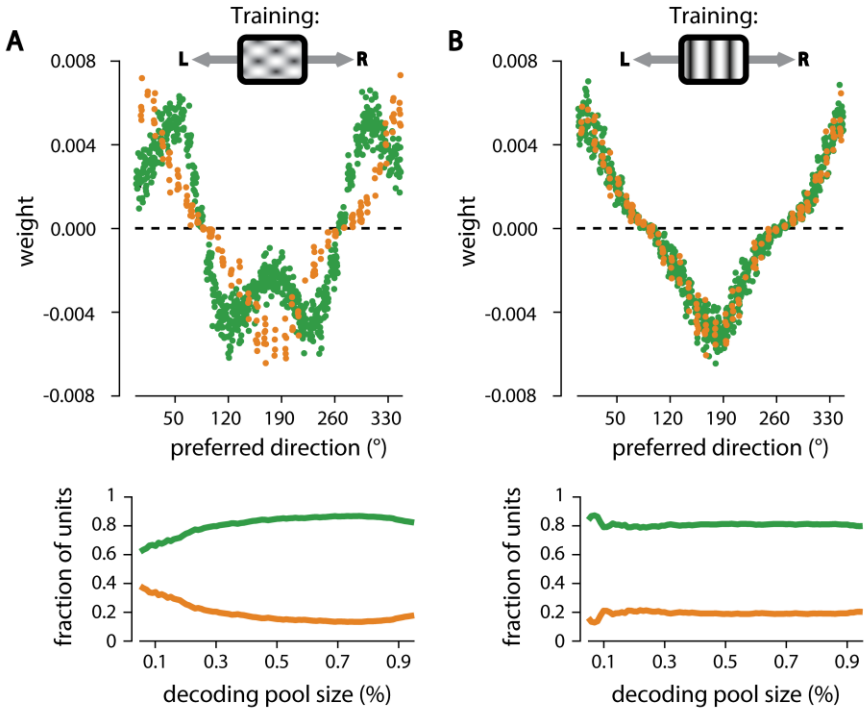

Figure 5. Computational control experiment showing the need for asymmetric cross-orientation suppression to observe a biased decoding pool-selection. Structure and parameters of the simulation are the same as in Fig. 4 except for the absence of cross-orientation suppression in component cells. A) Results for the plaid training case. Upper subpanel displays distribution of classifier weights after training. Weights are shown organizing neurons according to their preferred direction. Notice how the peak values of pattern and component weight distributions are now the same. This is reflected, as shown in the lower subpanel, in the absence of the previously observed reversal in the proportion of patterns and components included in the decoding pool for any pruning regime. B) Results for the grating training case. Both the weight distributions (upper subpanel) and pattern and component fractions (lower subpanel) closely match what shown in Fig. 4.

Materials and methods

Animal procedures

We trained 21 male Long Evans rats (Charles River Laboratory) in a motion direction discrimination task. Upon arrival in the lab, rats weighed ~250 g, and they grew up to ~500 g. Animals started training during ~ 7th postnatal week. During the experimental period, they had free access to food, while their access to water was restricted. Their daily liquid intake included 5 – 15 ml of a 1:4 juice-water solution, plus an ad libitum water access for one hour after training. Training sessions lasted 50 – 70 minutes and took place 5 days per week. All animal procedures were conducted in accordance with the international and institutional standards for the care and use of animals in research and were approved by the Italian

Ministry of Health and after consulting with a veterinarian (Project DGSAF 22884, submitted on November 24, 2014 and approved on September 4, 2015, approval 940/2015-PR).

Behavioral apparatus and stimuli

The behavioral setup comprised an operant box, with a 3 cm-diameter hole to allow animals to extend their head outside the box. An LCD monitor (ASUS® Ve228) was placed in front of the hole (30 cm distance). The box was equipped with three stainless-steel feeding needles (Cadence Science®), each connected to a led-photodiode pair to detect animal's interaction with the port. Feeding needles positioned at the right and at the left of the animal's head starting position were connected to the reward delivery system (lateral response ports). The central feeding needle was not connected to the reward delivery system (trigger port). This three-way needle-detector system is referred as "licking sensor". When an animal reached the feeding needles of different ports, photodiodes detected the response side (left or right port) or the triggering of a trial (trigger port). Stimulus presentation, response collection and reward delivery were controlled via workstations running open source MWorks software (https://mworks.github.io/). Liquid reward solution delivery was actuated by two computer-controlled syringe pumps (New Era Pump Systems NE-500®). The pumps were connected to the right and the left feeding needles in order to deliver the reward in case of correct response. Target grating stimuli consisted in full-field, full-contrast, sine wave drifting gratings, moving horizontally toward left (180°) or right (0°), with temporal frequency of 2 Hz and spatial frequency of 0.04 cycles/°. Target plaid stimuli were constructed by superimposition of two half-contrast gratings (again 0.04 cycles/° and 2 Hz) with a motion direction difference of 120° (i.e. 120° plaid cross-angle) with the same global motion direction of the grating targets (i.e. left or right). Prime stimuli drift direction (from 0° to 180° in steps of 10°) and identity, during the priming phase, were randomly selected on each trial (see Fig. 1).

Experimental design - Training phase

Rats were able to trigger a trial by reaching the central feeding needle. As soon as the animal reached the trigger, the target stimulus appeared (either a grating or a plaid depending on the animal's experimental group). Responses with reaction times <300 ms from the trigger time were considered "too fast": in this case the trial was aborted and excluded from the analysis. Similarly, responses occurring >2000 ms after the end of the stimulus (lasting 2000 ms) were considered "ignored" and excluded from the analysis. No reward was delivered in these invalid trials. In order to accomplish the task and receive the reward, the rat had to reach the feeding needle matching the global motion direction of the target stimulus. Correct execution of such task was rewarded equally for both directions during each experimental session. In correct trials, the target stimulus remained on the screen for the whole reward delivery period, as to strengthen operant association between stimulus response and reward. On the other hand, when the animal provided an incorrect response, besides not delivering any reward, an aversive 5-20 cycles of black and grey flashes was shown together with the playing of an auditory aversive tone ("time-out").

Experimental design - Priming phase

In the priming phase, a moving prime stimulus was shown to the animal before the moving target stimulus. The prime duration was 75 ms, and the ISI (i.e. time separating its offset from the target onset) was 75 ms. As in the training phase, “too fast” or “too long” trials with were excluded from the analysis. The definition of “too long” in this second phase of the experiment was changed to 1000 ms after target onset (so response window spanned times from 300 to 1000 ms from target onset). In order to obtain the reward, the animal, as in the training condition, had to reach the feeding needle corresponding to the global-direction of the target (i.e. right or left). In this phase, the target was not kept on screen during the reward, because there was no longer any need to favor the association between response, stimulus and reward. Only animals maintaining a performance in the training task greater that 75% for 4 consecutive days were transferred to this second phase.

Rat selection criterion

Only rats displaying a strong identity-priming effect where included in the analysis described in the main text. Quantitatively this corresponded to setting a threshold on the absolute priming magnitude in the identity priming condition amounting to 5%. The enforcement of such criterion led to the rejection of one rat per group bringing the number of rats included in the analysis 10 animals for the G group and to 9 animals for the P group out of the original 11 animals trained in grating direction discrimination and 10 trained in plaid direction discrimination.

Cross-priming predictions

Having observed a shared representation of grating and plaid primes in the P group, we were interested in obtaining predictions in deriving cross-priming predictions starting from the identity-priming curve (i.e. predicting the priming effect of the grating primes from that of plaid primes, over plaid targets). For what concerns the left and right predictions, to do so, we assumed as predicted cross-priming effect, for each direction, the one observed for the identity prime containing that specific grating as left or right component respectively (Fig. 6). For what concerns the average prediction, we averaged the left and right predictions, considering that this would be the effect of randomly selecting one of the two components over many trials. To make the analysis more sensitive to effects linked to the direction of prime stimuli (that are the ones enabling to infer the organizing principle of the perceptual representation), and less sensitive to intrinsic differences in the difficulty of the task between the identity and cross condition the offset needed to match the neutral prime level in the cross-priming condition (i.e. 90° priming performance) with the one observed in the identity-priming one was applied to the original identity priming-performances when generating either the left or right component prediction.

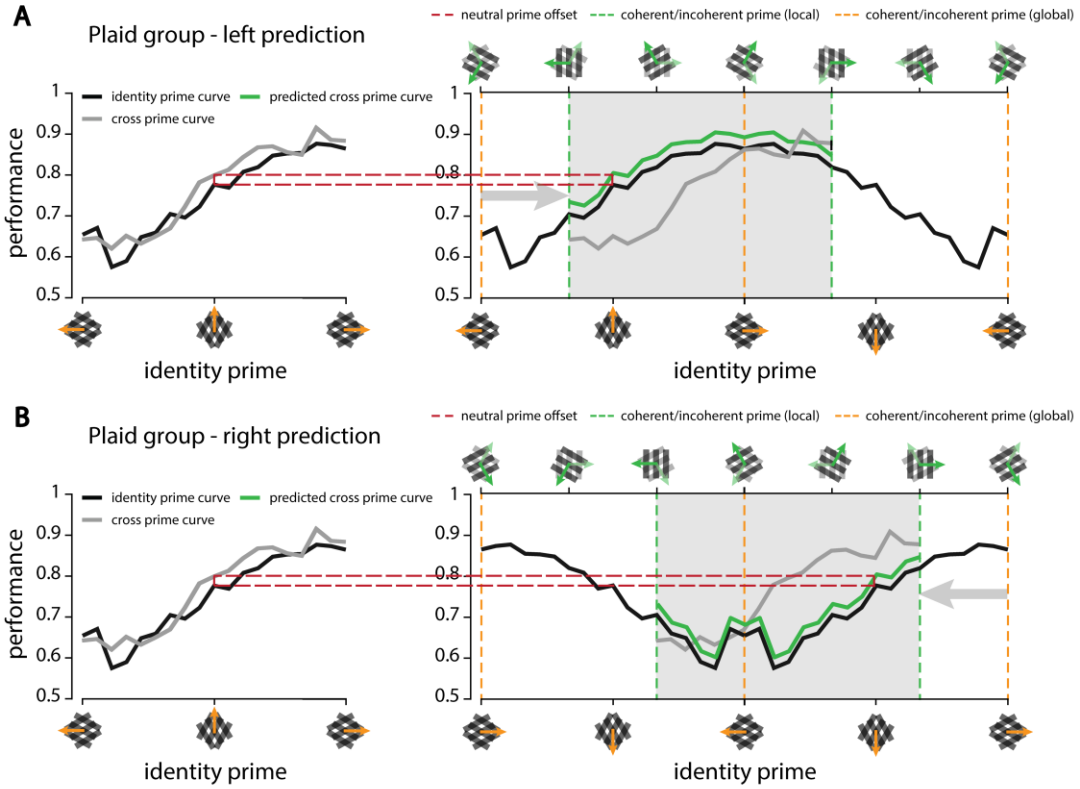

Figure 6. Graphical illustration of how left and right component predictions are obtained. Only P group case is depicted in figure. The point of the whole procedure is to choose, among the different directions of the identity-prime stimulus (black, left subpanel of A and B), the best suited one to be used to predict the effect of the cross-prime (light gray, left subpanel of A and B). Assuming a component representation of plaids, for each grating cross-priming direction, two plaid identity-prime directions are plausible candidate to this aim: the two in which the plaid contains the grating of interest as one its components. Considering always the candidate identity-prime containing the grating of interest as left component (wrt. to the global direction) we obtain the "left" prediction (A), whereas considering the candidate identity-prime containing the grating of interest as right component we obtain the "right" prediction (B). Matching the chosen candidate identity-prime with the corresponding cross-prime grating means shifting the observed identity-priming curve (after wrapping it around the unit circle) either 60° rightward (A - left prediction) or 60° leftward (B - right prediction). This is illustrated in right subpanels of A and B and highlighted by gray transparent arrows and shaded regions. As final step in such data-driven prediction generation procedure, a (small) offset (highlighted by red lines) is applied to the shifted curve (black) in order to match the original neutral prime level (i.e. 90° priming performance) in the identity and cross conditions, thus obtaining the green curve. This is meant to make the analysis more sensitive to effects linked to the direction of prime stimuli (that are the ones enabling to infer the organizing principle of the perceptual representation), and less sensitive to intrinsic differences in the difficulty of the task between the identity and cross condition.

Decoding pool selection simulation

To provide computational support to the mechanism we proposed, we trained a logistic regression classifier to perform the same task our animals were trained to perform (i.e. left/right direction discrimination of either gratings or plaids). As described in the main text, the stimuli representation assumed in the simulation is the one originating from a population of Poisson neurons containing more component than pattern cells (80% vs. 20%), having 10 Hz of peak firing rate, 0.9 of mean OSI (gaussianly distributed with standard deviation of

0.05), 0.7 of mean DSI (gaussianly distributed with standard deviation of 0.05) with randomly centered von Mises tuning curves (uniformly distributed across the circle).
To test whether cross-orientation suppression is necessary and sufficient to produce the bias in the selection of the decoding pool we hypothesized, we ran both a simulation with it (Fig. 4) and a control simulation without it (Fig. 5). A hypothetical "decision neuron" located upstream the sensory area(s) containing the decoded population would plausibly form synaptic connections with only a very small (i.e. sparse) subset of the neurons in that area. For this reason, in order to obtain a more biologically realistic estimate of the magnitude of the selection effect when taking into account such sparsity of connections, we quantified the proportion of pattern and component units in the decoding pool as a function of the quantile used to threshold the classifier weights (i.e. putting to zero any weight below the chosen quantile of the distribution − Fig. 4B and Fig. 5B bottom plots). This is referred elsewhere as "connection pruning". The results we obtained are consistent with what we hypothesized from the physiological literature. In fact, in the presence of cross-orientation suppression, the plaid-group classifier (Fig. 4B, left) ended up assigning much more weight to pattern cells, whereas this was not the case for the grating-group classifier (Fig. 4B, right).
Being pattern cells much rarer than component, the decoding pool resulted always (i.e. independently of the threshold chosen) unbalanced towards component cells in the grating-group case (Fig. 4B bottom right). Crucially, on the other hand, in the plaid training case when thresholding at plausibly high quantiles (i.e. >0.75), in the cross-orientation suppressed case, the opposite was true. Specifically, in this case, we observed decoding pools composed mainly by pattern cells (up to 100%), demonstrating a strong selection effect (Fig. 4B bottom left). Even without cross-orientation suppression, however, a very mild effect favoring the inclusion of pattern cells in the decoding pool was apparent, but the proportion of the pattern and component cells were never reversed (Fig. 5B bottom left). This is most likely due to the difference in signal-to-noise ratio between pattern and component cells originating from intrinsically different broadness of plaid tuning curves characterizing the two cell classes.
In conclusion this suggests that, in presence of cross-orientation suppression, in physiologically plausible conditions, a plaid-trained decision neuron would likely end up recruiting a pattern-enriched population whereas a grating-trained one would recruit a mainly component one.

References

1. Born, R. T. & Bradley, D. C. Structure and function of visual area MT. *Annu Rev Neurosci* **28**, 157–189 (2005).
2. Juavinett, A. L. & Callaway, E. M. Pattern and Component Motion Responses in Mouse Visual Cortical Areas. *Curr. Biol.* **25**, 1759–1764 (2015).
3. Palagina, G., Meyer, J. F. & Smirnakis, S. M. Complex Visual Motion Representation in Mouse Area V1. *J. Neurosci.* **37**, 164–183 (2017).
4. Orban, G. A. Higher Order Visual Processing in Macaque Extrastriate Cortex. *Physiol. Rev.* **88**, 59 (2008).

5. Adelson, E. H. & Movshon, J. A. Phenomenal coherence of moving visual patterns. *Nature* **300**, 523–525 (1982).
6. Fennema, C. L. & Thompson, W. B. Velocity determination in scenes containing several moving objects. *Comput. Graph. Image Process.* **9**, 301–315 (1979).
7. Wuerger, S., Shapley, R. & Rubin, N. "On the Visually Perceived Direction of Motion" by Hans Wallach: 60 Years Later. *Perception* **25**, 1317–1367 (1996).
8. Marr, D. *Vision: a computational investigation into the human representation and processing of visual information*. (MIT Press, 2010).
9. Fenno, L., Yizhar, O. & Deisseroth, K. The Development and Application of Optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
10. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295 (2013).
11. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
12. Luo, L., Callaway, E. M. & Svoboda, K. Genetic Dissection of Neural Circuits. *Neuron* **57**, 634–660 (2008).
13. Luo, L., Callaway, E. M. & Svoboda, K. Genetic Dissection of Neural Circuits: A Decade of Progress. *Neuron* **98**, 256–281 (2018).
14. Zoccolan, D. Invariant visual object recognition and shape processing in rats. *Behav. Brain Res.* **285**, 10–33 (2015).
15. Muir, D. R., Roth, M. M., Helmchen, F. & Kampa, B. M. Model-based analysis of pattern motion processing in mouse primary visual cortex. *Front. Neural Circuits* **9**, (2015).
16. Rust, N. C., Mante, V., Simoncelli, E. P. & Movshon, J. A. How MT cells analyze the motion of visual patterns. *Nat Neurosci* **9**, 1421–31 (2006).
17. Smith, M. A., Majaj, N. J. & Movshon, J. A. Dynamics of motion signaling by neurons in macaque area MT. *Nat. Neurosci.* **8**, 220–228 (2005).
18. Solomon, S. S. *et al.* Visual motion integration by neurons in the middle temporal area of a New World monkey, the marmoset: Pattern motion sensitivity in middle temporal area of marmosets. *J. Physiol.* **589**, 5741–5758 (2011).
19. Douglas, R. M., Neve, A., Quittenbaum, J. P., Alam, N. M. & Prusky, G. T. Perception of visual motion coherence by rats and mice. *Vision Res.* **46**, 2842–2847 (2006).
20. Stirman, J. N., Townsend, L. B. & Smith, S. L. A touchscreen based global motion perception task for mice. *Vision Res.* **127**, 74–83 (2016).
21. Samonds, J. M., Lieberman, S. & Priebe, N. J. Motion Discrimination and the Motion Aftereffect in Mouse Vision. *eNeuro* **5**, (2018).
22. Marques, T. *et al.* A Role for Mouse Primary Visual Cortex in Motion Perception. *Curr. Biol. CB* **28**, 1703-1713.e6 (2018).
23. Tafazoli, S., Di Filippo, A. & Zoccolan, D. Transformation-Tolerant Object Recognition in Rats Revealed by Visual Priming. *J. Neurosci.* **32**, 21–34 (2012).
24. Zoccolan, D. & Di Filippo, A. Chapter 5 - Methodological Approaches to the Behavioural Investigation of Visual Perception in Rodents. in *Handbook of Behavioral Neuroscience* (eds. Ennaceur, A. & de Souza Silva, M. A.) vol. 27 69–101 (Elsevier, 2018).

25. Nikbakht, N., Tafreshiha, A., Zoccolan, D. & Diamond, M. E. Supralinear and Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and Neuronal Mechanisms. *Neuron* **97**, 626-639.e8 (2018).
26. Licata, A. M. *et al.* Posterior Parietal Cortex Guides Visual Decisions in Rats. *J. Neurosci.* **37**, 4954–4966 (2017).
27. Odoemene, O., Pisupati, S., Nguyen, H. & Churchland, A. K. Visual Evidence Accumulation Guides Decision-Making in Unrestrained Mice. *J. Neurosci.* **38**, 10143–10155 (2018).
28. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792 (2014).
29. Stahl, J. S. Using eye movements to assess brain function in mice. *Vision Res.* **44**, 3401–3410 (2004).
30. Smith, A. T. & Harris, L. R. Use of plaid patterns to distinguish the corticofugal and direct retinal inputs to the brainstem optokinetic nystagmus generator. *Exp. Brain Res.* **86**, (1991).
31. Harvey, R. J., De'sperati, C. & Strata, P. The Early Phase of Horizontal Optokinetic Responses in the Pigmented Rat and the Effects of Lesions of the Visual Cortex. *Vision Res.* **37**, 1615–1625 (1997).
32. Newsome, W. & Pare, E. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211 (1988).
33. Salzman, C. D., Britten, K. H. & Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**, 174–177 (1990).
34. Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H. & Zoccolan, D. Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy. *Curr. Biol.* **28**, 1005-1015.e5 (2018).
35. Alemi-Neissi, A., Rosselli, F. B. & Zoccolan, D. Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition. *J. Neurosci.* **33**, 5939–5956 (2013).
36. Rosselli, F. B., Alemi, A., Ansuini, A. & Zoccolan, D. Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. *Front. Neural Circuits* **9**, 10 (2015).
37. Zoccolan, D., Oertelt, N., DiCarlo, J. J. & Cox, D. D. A rodent model for the study of invariant visual object recognition. *Proc Natl Acad Sci U A* **106**, 8748–53 (2009).
38. Bar, M. & Biederman, I. Subliminal Visual Priming. *Psychol. Sci.* **9**, 464–468 (1998).
39. Leopold, D. A., O'Toole, A. J., Vetter, T. & Blanz, V. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat Neurosci* **4**, 89–94 (2001).
40. Kanai, R. & Verstraten, F. A. J. Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Res.* **45**, 3109–3116 (2005).
41. Kravitz, D. J., Kriegeskorte, N. & Baker, C. I. High-Level Visual Object Representations Are Constrained by Position. *Cereb. Cortex* **20**, 2916–2925 (2010).
42. Groh, J. M. Reading neural representations. *Neuron* **21**, 661–664 (1998).

43. DeAngelis, G. C., Robson, J. G., Ohzawa, I. & Freeman, R. D. Organization of suppression in receptive fields of neurons in cat visual cortex. *J. Neurophysiol.* **68**, 144–163 (1992).
44. Busse, L., Wade, A. R. & Carandini, M. Representation of Concurrent Stimuli by Population Activity in Visual Cortex. *Neuron* **64**, 931–942 (2009).
45. Self, M. W. *et al.* Orientation-Tuned Surround Suppression in Mouse Visual Cortex. *J. Neurosci.* **34**, 9290–9304 (2014).
46. Alwis, D. S., Richards, K. L. & Price, N. S. C. Masking reduces orientation selectivity in rat visual cortex. *J. Neurophysiol.* **116**, 2331–2341 (2016).
47. Samonds, J. M., Feese, B. D., Lee, T. S. & Kuhlman, S. J. Nonuniform surround suppression of visual responses in mouse V1. *J. Neurophysiol.* **118**, 3282–3292 (2017).
48. Roelfsema, P. R. & Holtmaat, A. Control of synaptic plasticity in deep cortical networks. *Nat. Rev. Neurosci.* **19**, 166–180 (2018).
49. Kohn, A. & Movshon, J. A. Adaptation changes the direction tuning of macaque MT neurons. *Nat Neurosci* **7**, 764–772 (2004).
50. Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B. & Zoccolan, D. Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J. Neurosci.* **39**, 1649–1670 (2019).
51. Rasmussen, R. & Yonehara, K. Circuit Mechanisms Governing Local vs. Global Motion Processing in Mouse Visual Cortex. *Front. Neural Circuits* **11**, 109 (2017).

Chapter V

Linear receptive field structure does not account for pattern motion responses in rat visual cortex

Abstract

A key processing stage of motion perception is solving the aperture problem – i.e. extracting the global direction of visual objects by integrating many local direction signals. Decades of studies on primates documented how primary visual cortex (V1) contains almost exclusively neurons that are selective for local direction (i.e., "component cells"), whereas higher-order dorsal stream areas, such as the middle-temporal (MT) and the medial superior temporal (MST) areas, are strongly enriched in global-direction selective units (i.e., "pattern cells"). In rodents, the investigation of the neuronal substrates of motion integration started only recently, with contrasting findings about the presence of pattern cells in mouse V1, and one study reporting non-negligible fractions of such cell type in mouse lateromedial (LM) and rostrolateral (RL) areas. None of these studies, however, mapped the linear receptive field (RF) structure of the recorded neurons, leaving open the possibility that the integrative properties of these putative pattern cells might trivially originate from the specific geometry (e.g., the low aspect ratio) of linear RF of non-integrative (i.e., component-like) neurons. To investigate the extent to which rodent visual cortex contains true (i.e. integrative) pattern cells, we performed extracellular recordings in V1, LM and RL of anesthetized rats, during the presentation of drifting gratings, plaids (i.e., superpositions of gratings with different directions) and noise movies. This revealed the presence of pattern cells in rat V1 and LM, but not in RL, corroborating some previous results while contradicting others. Moreover, by carrying out reverse correlation analysis, we were able to reconstruct the linear RFs of some pattern and component cells. Using them as filters to predict grating and plaid direction tuning curves showed that the linear RF structure does not account for pattern motion responses in rat visual cortex. This strongly suggests that rodent pattern cells truly perform a nonlinear integration of local motion signals to extract the global motion direction of complex visual patterns.

Introduction

Perceiving the velocity (i.e. motion direction and speed) of visual objects is critical to interact effectively with the environment. Primary visual cortex can be thought as a bank of local moving-edge detectors providing a representation of visual input from which object motion direction can be extracted. Carrying out such computation, however, is not completely trivial. The output of a single localized edge-detector is, in fact, intrinsically ambiguous, since it reflects the projection of the global object velocity vector onto the local edge-orthogonal direction – in this way, any information regarding the edge-parallel component of object motion is lost in the computation. This makes the output of such kind of filter compatible with infinite combinations of global directions and speeds, and therefore insufficient to fully specify the velocity of the object causing it. Only by combining (i.e., integrating) multiple local direction signals of this kind, global object direction and speed can be fully determined. This ambiguity is known in the neuroscientific literature as the "aperture problem"[1–3]. Psychophysically, it can be appreciated by the fact that observers looking at a drifting object through a small aperture will perceive the edge seen through the aperture as always drifting in the perpendicular direction to the edge itself, irrespectively of the global direction of the object behind the aperture[3]. If not handled properly by the visual system, the aperture problem would lead to illusory and inaccurate motion measurements.

In the mammalian brain, motion integration is known to be achieved by pattern cells, which are abundant in monkey dorsal stream areas such as MT[4–11] and MST[9,10]. The complementary cell class is that of component cells, neurons that − behaving like localized edge detector filters − are sensitive to the aperture problem. Such kind of neurons have been reported to be predominant in V1[4,6,9,10] and widespread across multiple areas of the monkey visual cortex.

In rodents, only a handful of studies investigated the distribution of pattern and component cells in V1[12,13] and bordering high-order visual areas[14]. Furthermore, such studies yielded contrasting results about the presence of pattern cells in V1: two of them reported a small but consistent fraction of pattern units in this area[12,13], while another did not find any[14]. Despite the lack of a clear picture on the electrophysiological front, recent work provided compelling evidence about cortex-dependent motion processing capabilities of rodents[15]. Furthermore, as described in the previous chapter of this thesis, we recently provided behavioral evidence for the ability of rats to spontaneously perceive global motion.

Interestingly, the computational model included in that study was able to explain our behavioural observations assuming two key properties for the neural representation of motion signals: an asymmetry in the fraction of component and pattern cells (with the former being much more abundant than the latter) and an asymmetry in the effect of cross-orientation suppression on these two classes of neurons (with component cells being much more strongly suppressed by the presence of multiple oriented gratings/edges than pattern cells). Identifying such properties in some visual area of the rodent brain would convincingly candidate that region to be the area supporting our behavioral task. More generally, the rise of rodents as interesting models for motion integration calls for: i) the collection of new data capable of resolving the ongoing debate about the proportion of pattern units across their visual areas;

and ii) the investigation of the computational mechanisms underlying their tuning properties[16].

To achieve the first aim (i), we undertook an electrophysiological investigation of V1 and two putative dorsal extrastriate areas, LM and RL. LM was chosen by virtue of being positioned at the interface between ventral and dorsal streams on functional[17–19] and anatomical[20] grounds, consistently with a proposed homology to monkey V2. Furthermore, LM was reported by Juavinett et al.[14] as highly enriched in pattern cells, making it a very interesting target for our investigation. RL, on the other hand, has been proposed as a rodent homologue of monkey MT by the same authors[14], who found the highest proportion of pattern cells in such area across all mouse visual cortex.

To address the second issue (ii), we considered the possibility of explaining the observed pattern responses as trivially emerging from the responses of broadly tuned linear units due to the "blobbyness" (i.e., low aspect ratio) of their receptive fields (RFs). In fact, even if Palagina et al.[13] reported no significant difference of average tuning broadness between pattern and component cells, they found that pattern responses in mouse V1 are not cross-angle invariant (i.e. they change their pattern/component behaviour depending on the angle between the two component grating forming the plaid). Such sensitivity to the cross-angle between the grating components of the plaid led them to suggests that the observed pattern-selectivity may trivially rely on the geometry of the RF[13]. As originally proposed by Tinsley et al.[21], apparent pattern responses could emerge from linear units when local contrasts "blobs" of the plaid tightly overlap the excitatory/inhibitory subfields of the neuron's RF. Such explanation looks particularly plausible in rodents, given that, compared to higher acuity mammals[22–25], rodent visual neurons have broader tuning curves, lower spatial frequency preferences[18,26,27], and RFs with a lower aspect ratio[26,28]. This is an important factor since, as confirmed by a simple simulation (described below; see Fig. 1), for an ideal Gabor filter, blobbyness of the receptive field (i.e., low aspect ratio), broadness of the direction tuning curve and the ability to produce pattern-like responses (when probed with plaid stimuli) are tightly interconnected.

In the light of these considerations, when investigating rodent visual cortex, the risk of misclassifying linear, non-integrative units (that should be properly considered as broadly tuned component cells) as pattern cells cannot be overlooked. This could trivially explain pattern selectivity observed in rodents without invoking nonlinear hierarchical mechanisms similar to those emerging from the primate literature[8,29]. As suggested by Palagina et al.[13], the only direct way to test such blobbyness hypothesis is to reconstruct the linear receptive fields of putative pattern neurons and try to predict their plaid responses on that basis. If the blobbyness hypothesis holds true, the responses predicted by linear RFs should still be pattern-like. On the other hand, if nonlinear, truly integrative mechanisms are at work, linear RFs should fail to produce pattern-like responses.

Results

To provide a quantitative proof-of-principle of the blobbyness effect described above, we ran a simple simulation with a Gabor-filter-based linear-nonlinear (LN) neuron model (Fig. 1).

We fed to the model gratings and plaids (made of the superposition of two gratings, whose directions were 60º apart – this small cross-angle value was chosen to make the effect appear already at moderate broadness levels, for illustrative purposes), with 12 equi-spaced drift directions for each stimulus type (from 0° to 330°). We used the simulated responses of the model neuron to these stimuli to compute tuning curves. We considered 3 different aspect ratios of the Gabor filter. In the "high" aspect ratio condition (see the elongated shape of the filter shown in Fig. 1A), the tuning curves obtained with the grating (solid line) and plaid (dotted line) stimuli both displayed narrow peaks (i.e., sharp tuning).

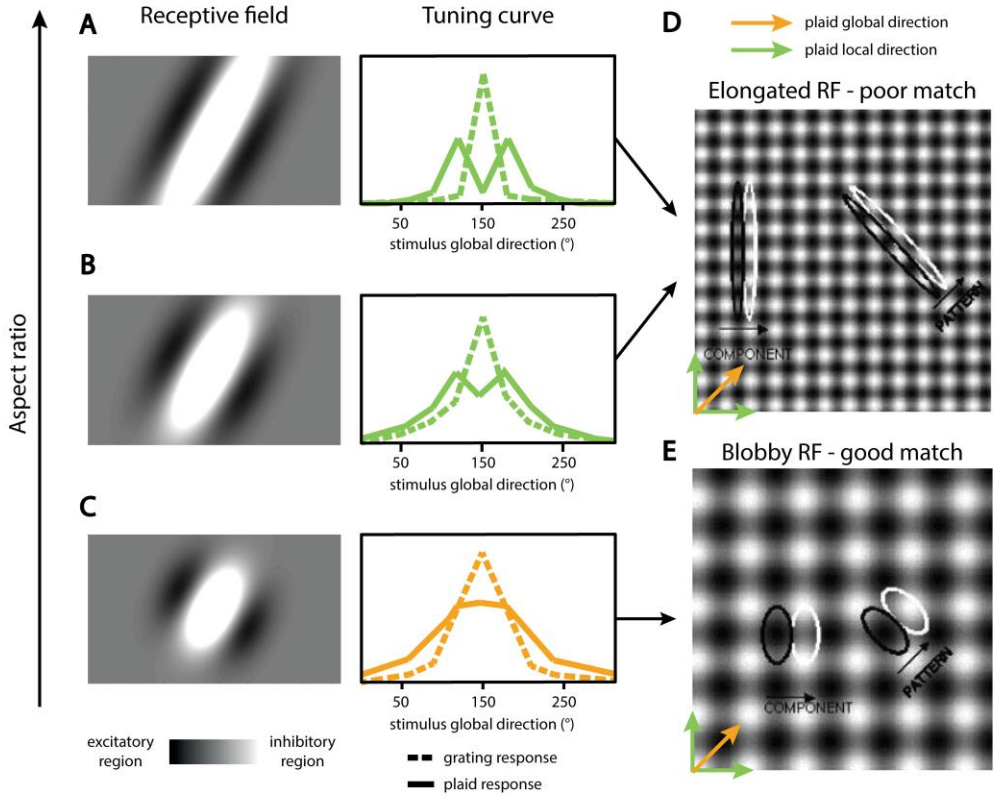

Figure 1. IIllustration of the blobbyness effect. A) On the left, spatial snapshot of the linear filter (i.e. linear RF) underlying the selectivity of a simulated LN neuron. On the right, grating (dashed line) and plaid (continuous line) direction tuning curves produced by the LN neuron are shown. Notice how the high aspect ratio of the filter (i.e. "elongated") cause the single peak of grating tuning curve being narrow (i.e. sharp tuning). In this condition the two peaks of the plaid tuning curve are well separated, enabling an unambiguous classification of the unit as a component cell (highlighted by green coloring). B) Same data as in panel A for a lower aspect ratio filter: grating tuning curve gets broader and two plaid tuning curve peaks start to overlap even if they are still distinguishable. The LN neuron is still classified as component (green). C) Same data as in panel A for a low aspect ratio filter (i.e. a "blobby" one): tuning gets so broad that the two plaid tuning curve peaks are no longer distinguishable, so that the LN neuron is now classified as pattern (orange). D) Example of elongated RF poorly matching the local contrast features of the plaid. Vertical RF aligned to component direction (left) sees average luminance within each subunit matching their polarity, producing a component response. Oblique RF aligned to global direction (right) sees approximately mid-gray luminance within each subunit, producing no pattern response. E) Example of blobby RF matching the local contrast features of the plaid. Both vertical and oblique RF see average luminance within each subunit matching their polarity, thus producing a component and a pattern response, respectively. Panels D and E adapted from Tinsley et al.[22].

Crucially, the tuning curve for the plaids peaked at +/- 30° (i.e., half plaid cross-angle) from the preferred direction of the gratings. Such difference is the hallmark of component cells. For an intermediate aspect ratio (notice the less elongated shape of the filter of Fig. 1B),

the result was similar, even if all tuning curves were broader (i.e., the less sharp tuning). Nevertheless, the two peaks of the tuning curve obtained with the plaids were still well separated, enabling the proper classification of such neuron as a component cell. Finally, when the aspect ratio was lowered further (the width of the Gabor filter shown in Fig. 1C is as large as its height), the tuning curves became very broad. Because of this, the two peaks of tuning curve obtained for the plaids (solid line) merged into a single peak centred at the midpoint between the original positions of the two. Since such position coincides with the global direction of the plaid, the model unit displayed the landmark property of pattern cells (i.e., similar tuning curves when probed with gratings or plaids). This is why, in this latter condition of low aspect ratio, our simple linear unit would be misclassified as an integrative pattern cell solving the aperture problem. Fig. 1D and E provide a graphical description of the process leading a unit with a linear Gabor RF to look more like a component or more like a pattern cell, depending on its aspect ratio.

To measure the relative proportion of pattern and component cells across rat visual cortex, as well as to directly test whether the blobbyness effect could explain pattern responses previously reported across rodent visual cortex, we performed extracellular recordings from primary visual (V1, n = 258), lateromedial (LM, n = 187) and rostrolateral (RL, n = 184) cortex of anesthetized rats. Animals were passively exposed to a stimulus set including moving grating and plaids (with a 120° cross-angle) of all combinations of two spatial and two temporal frequencies, drifting in 12 equi-spaced directions (from 0° to 330°), as well as spatiotemporally correlated noise movies (see Materials and methods). Grating and plaid responses were used to compute direction tuning curves and classify the recorded single units as pattern or component cells, based on the standard approach developed in cat and monkey studies[4], and also used in previous rodent studies[12–14]. Noise movies were used to estimate the linear RF of each neuron (i.e., to find the best linear filter that approximated the stimulus-response function) by using the Spike-Triggered Average (STA) technique[30,31].

Only neurons significantly and reliably responsive to gratings or plaids were included in the analysis described below (see Material and methods). “Patternness” and “componentness” were quantified by computing Fisher-transformed partial correlation between the observed responses to the plaids and the responses inferred from the observed responses to the gratings, assuming either an ideal pattern or component selectivity (these correlations are referred to as Zp and Zc respectively; see Materials and methods for a definition). Direction selective neurons showing one of the two Z indexes significantly higher than 0 and bigger than the other were classified as pattern or component, as usually done in the literature (see Material and methods). Neurons that did not meet at least one of these requirements were labelled as unclassified.

Fig. 2 shows tuning curves and raster plots that were obtained for a few example neurons recorded from the three targeted areas and classified as either pattern or component. Fig. 3 shows instead the distribution of the Zp and Zc indices for the neuronal populations sampled from the three areas. In agreement with Palagina et al.[13] and Muir et al.[12], but in contrast with Juavinett et al.[14], our recordings yielded a sizable amount of pattern cells in V1 (6% of the 78 units meeting the responsivity and direction selectivity criteria in V1). In qualitative agreement with the latter study[14], on the other hand, we found a significant fraction of pattern cells in LM (8% of the 74 units meeting the plaid responsivity and

direction selectivity criteria in LM). By contrast, in disagreement with Juavinett et al.[14] we found the RL population completely devoided of pattern cells (0% of the 65 units meeting the plaid responsivity and direction selectivity criteria in RL).

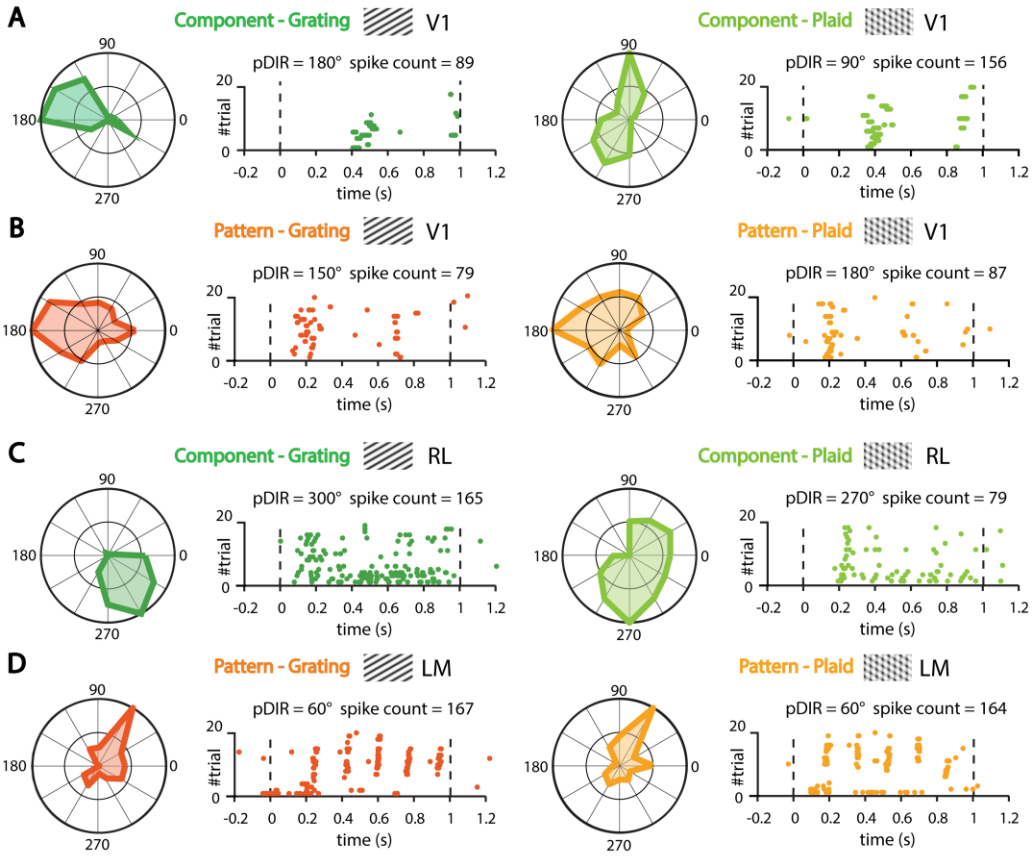

Figure 2. Pattern and component neuron examples from V1, LM and RL. In each subpanel of each panel, both preferred direction raster plot (right) and polar direction tuning curve (left) are shown. A) V1 component cell response to gratings (left subpanel, darker green) and plaid (right subpanel, lighter green). B) V1 pattern cell response to gratings (left subpanel, darker orange) and plaid (right subpanel, lighter orange). C) RL component cell response to gratings (left subpanel, darker green) and plaid (right subpanel, lighter green). D) LM pattern cell response to gratings (left subpanel, darker orange) and plaid (right subpanel, lighter orange).

The fraction of component cells in V1 (27% out of the 78 units meeting the plaid responsivity and direction selectivity criteria in V1), in turn, was found to be in qualitative agreement with Muir et al.[12] and Juavinett et al.[14], but different from that reported by Palagina et al.[13] (where it amounted to nearly half the fraction we observed). On the other hand, whereas the proportion of component cells we observed in RL (23% out of the 74 units meeting the responsivity and direction selectivity criteria in RL) was similar to the one observed by Juavinett et al.[14], the one in LM (16% out of the 65 units meeting the responsivity and direction selectivity criteria in LM) was roughly half of what they reported. Finally, consistently with previous rodent studies, the larger fraction of single units included in the analysis fell into the *unclassified* category (67% in V1, 76% in LM and 77% in RL).

As already anticipated, the next step of our analysis aimed at comparing the responses predicted for pattern and component cells on the basis of their linear RFs, with the responses that were experimentally observed. The goal was to test the blobbyness hypothesis, which implies that observed pattern responses could trivially originate from the RF geometry of

linear, non-integrative neurons. To this aim, linear RFs, inferred by computing STA images from the responses to noise movies, were used as the input stage of LN models to predict grating and plaid tuning curves (see Materials and methods).

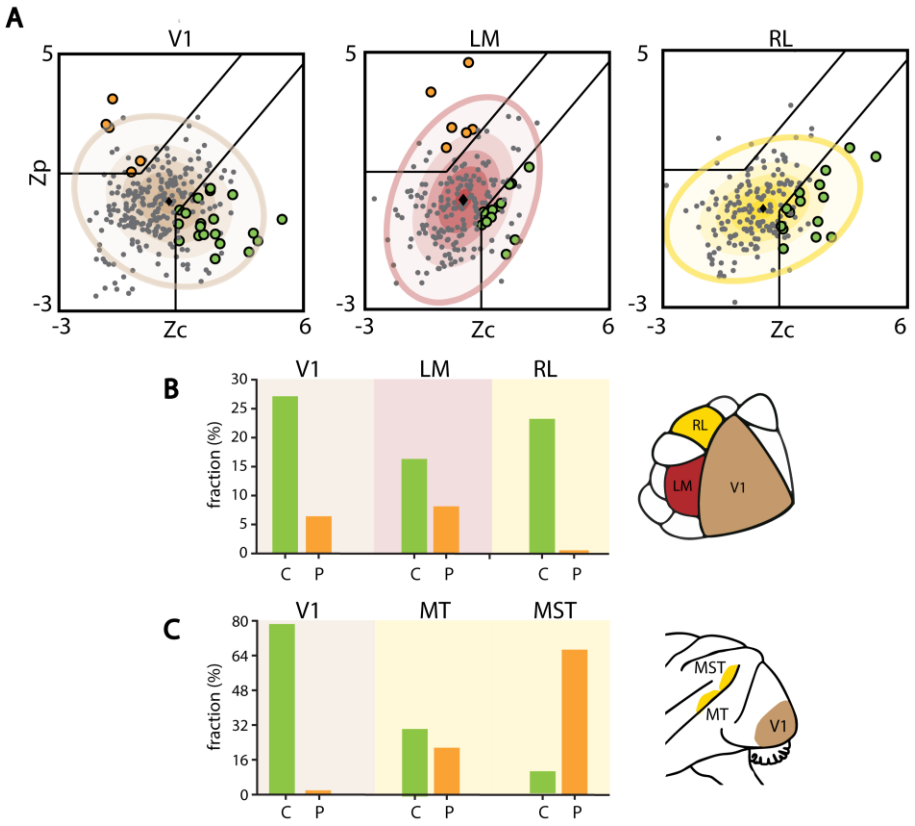

Figure 3. Pattern and component cells across V1, LM and RL. A) Zp and Zc index distributions in V1 (left subpanel), LM (central subpanel) and RL (right subpanel). Gray dots mark unclassified units (either for lack of sufficient direction selectivity or for insufficient Zp and Zc values). Green dots represent component cells whereas orange ones represent pattern. Solid black lines highlight decision boundaries for pattern and component classification. Colored shaded contours represent the 95% quantile of best fit gaussian distributions to the Zp and Zc observed distributions. B) Observed fraction of component cells (green bars) and pattern cells (orange bars) in V1 (left), LM (center) and RL (left). C) Average (across studies reviewed by Palagina et al.[14]) fraction of component cells (green bars) and pattern cells (orange bars) in macaque V1 (left), MT (center) and MST (left).

In Fig. 4A, observed (solid, coloured line) and LN-predicted (solid, black line) tuning curves for the same V1 neurons previously shown in Fig. 2A-B are plotted along with a spatial snapshot of their STA-based RFs (Fig. 4B). Comparing observed and LN-predicted tuning curves for the component cell (in green, on the left) revealed a fairly good match between the two. On the contrary, for the pattern cell (in orange, on the right), a qualitatively different behavior emerged. Whereas the observed and predicted tuning curves for the gratings were, again, reasonably well matched, the observed and predicted tuning curves for the plaids were not. Specifically, the predicted tuning curve for the plaids was substantially broadened on the sides of the peak. As a result, the predicted tuning curve became more component-like. Crucially, the effect was strong enough to switch the classification of the LN-predicted tuning curve, with respect to the measured one: based on its linear-RF-predicted response, the pattern cell would be classified as a component cell. As illustrated in Fig. 1C, this is at odd with the blobbyness hypothesis, which would predict instead a pattern-like tuning curve, simply based on output of a linear Gabor-like filter.

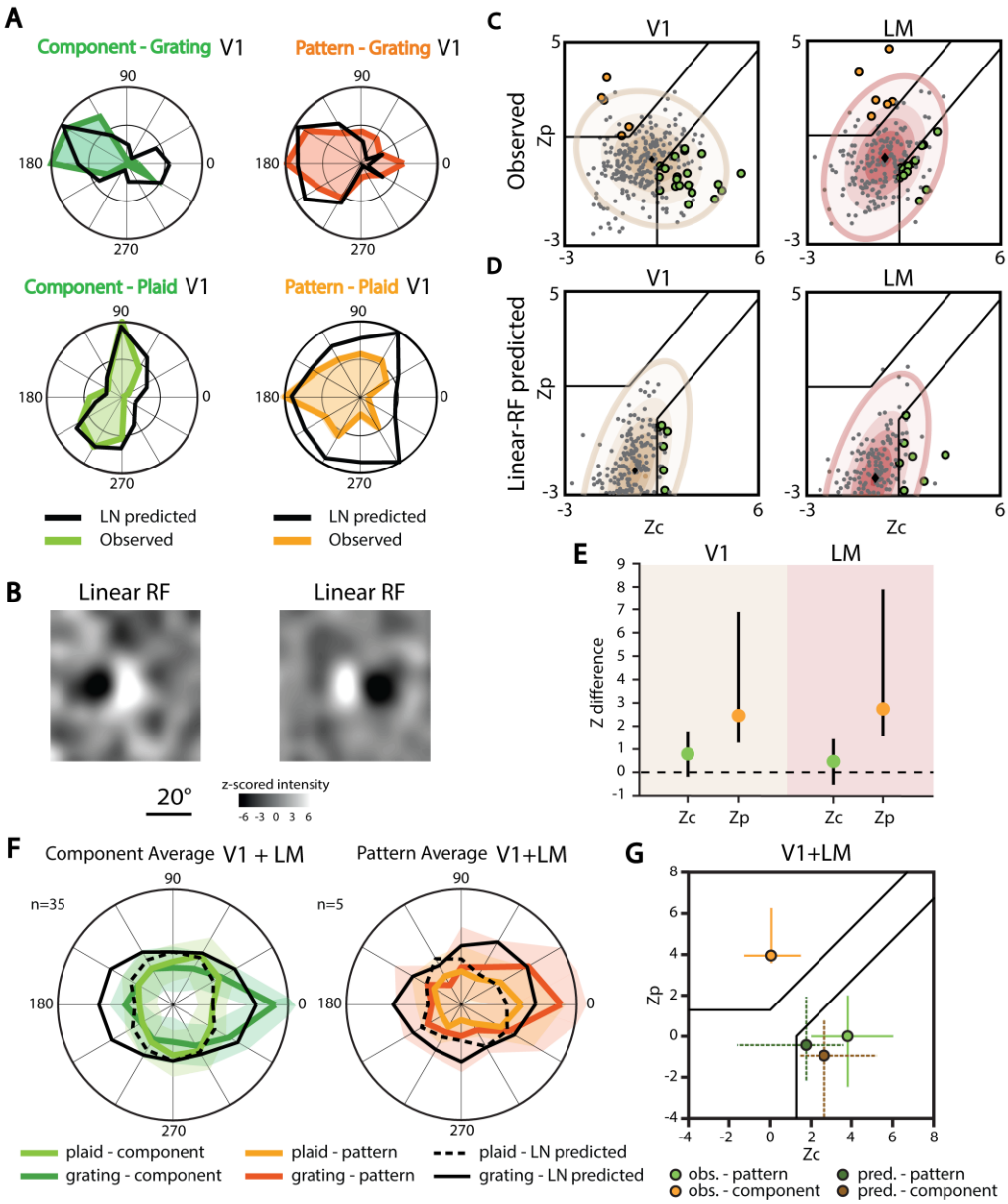

Figure 4. Linear RF does not account for observed pattern responses in V1 and LM. A) Observed (colored) and LN-predicted (black) grating (top) and plaid (bottom) tuning curves. Same pattern and component V1 neuron examples shown in Fig. 2. Notice strong enlargement of LN-predicted tuning curve with respect to the observed one in the bottom right polar plot (i.e. pattern cell plaid tuning curve): according to the LN-predicted curve, the neuron would be no longer classified as a pattern cell. B) Spatial snapshots of the STA-estimated linear RF structure of the two neurons from panel A (left: component, right: pattern). C) Observed Zp and Zc index distributions in V1 and LM (same as in Fig. 2, panel A). D) LN-predicted Zp and Zc index distributions in V1 and LM. Notice that whereas a sizeable number of component cells is predicted on the basis of the linear RFs, no pattern cell is predicted. E) Plot showing the average observed vs. LN-predicted difference (i.e. "Z difference") of Zp (orange) and Zc (green) values in V1 (left) and LM (right), along with bootstrap-estimated 95% C.I. (black whiskers). Notice that difference in componentness is compatible with 0 whereas difference in patternness is not. F) Average pattern and component tuning curves aligned to their peak. Only neurons with more that 0.3 correlation coefficient between observed and LN-predicted grating tuning curves are included (while no requirement on direction selectivity was enforced). G) Position of observed (light colors) and LN-predicted (dark colors) average tuning curves in the Zc and Zp plane. Notice how, in the component case, the predicted datapoint is still firmly in the component region of the plane. Crucially, on the other hand, in the pattern case the predicted datapoint switches from the pattern to the component region of the plane.

To quantify this effect in the sampled V1 and LM neural populations, we compared the observed (Fig. 4C) and LN-predicted (Fig. 4D) Zp and Zc distributions in the two areas. To this aim, we computed the mean difference between observed and LN-predicted Zp and Zc values by bootstrapping over the neurons to get 95% confidence intervals for such difference (denoted as "Z-difference" in Fig. 4E). Such analysis revealed no significant change in componentness (difference compatible with 0 at 95% confidence), but a significant reduction in patternness (difference incompatible with 0 at 95% confidence).

To better understand whether the behaviour illustrated for the example neurons in Fig. 4 generalized to all neurons categorized as pattern or component, we computed an average pattern and an average component tuning curve by re-aligning each curve to its grating preferred direction. This analysis was inspired by the one carried out, in a different context, by Ibrahim et al.[32] To replicate the analysis carried out on the single neurons, we first had to make sure that the STA had a fair amount of predictive power over the tuning curves for the gratings of neurons included in each ensemble. For this reason, we included in the pattern and component ensembles for this analysis only neurons with more than 0.3 correlation coefficient between observed and LN-predicted tuning curves. However, doing so, a very small number of units among the already few pattern cells would satisfy such criterion in each area. To overcome this issue, thus improving the statistical power of the analysis, we increased the number of units included in each ensemble by dropping the usual requirement on direction selectivity and merging V1 and LM populations (i.e., the two areas containing pattern cells).

Average re-aligned tuning curves for pattern (orange) ad component (green) units are shown in Fig. 4F, along with their 95% bootstrap-estimated confidence intervals (bootstrap carried out over neurons included in each ensemble). As it can be appreciated by comparing light and dark colored traces in both subpanels, average re-aligned tuning curves, as expected, still looked like typical pattern and component curves, respectively. Consistently, computing Zp and Zc indexes for these average curves, so as to carry out the usual classification, yielded the expected result of the average curve of the pattern cells being classified as pattern, and the average curve of the component cells being classified as component. This can be appreciated by looking, in Fig. 4G, how the pairs of Zp and Zc indexes computed for these average curves (light orange and green dot for, respectively, the average pattern and component curves) fell in the matching region of the Zp and Zc plane.

When considering the average of the LN-predicted, re-aligned tuning curves for the component ensemble, we could still observe a clear component-like behavior (black curves in Fig. 4F, left). Crucially, however, when considering the average of the LN-predicted, re-aligned tuning curves for the pattern ensemble, we observed a clear difference with respect to the observed curves: the LN-predicted curves looked way more component-like than pattern-like (black curves in Fig. 4F, right). Such observation was quantified by looking at the position of the LN-predicted averages in the Zp and Zc plane. As shown in Fig. 4G, both the LN-predicted pattern and component datapoints (dark green and dark orange dots respectively) fell in the component region of the plane. This means that LN models based on linear STA filters predict a component behavior not only for the actual component cells but also for the pattern cells. This, in turn, implies that the linear RF structure of the recorded units does not account for pattern motion responses we observed in rat V1 and LM.

Having demonstrated the non-trivial nature of pattern responses in V1 and LM, we further assessed whether either area featured the other key response property predicted by the computational model described in the previous chapter to explain rat perception of plaids and gratings – i.e., whether cross-orientation suppression would differentially affect pattern and component cells. To test this, we computed a relative peak response difference index, quantifying the amount of cross-orientation suppression or facilitation displayed by a neuron. This index is defined as the relative firing rate difference between peak plaid and grating responses: a negative value indicates cross-orientation suppression (i.e., larger peak response for grating than for plaids), while a positive value indicates facilitation (i.e., larger peak response for plaids than for gratings).

As can be seen in Fig. 5, V1 and LM populations (gray clouds) were distributed mainly across negative values, with an average of -21.7% in V1 and -35.9% in LM. These values indicate that the populations in the two areas are significantly suppressed, on average, when presented with plaids, as compared to the case in which they are presented with gratings ($p<0.001$ in both cases; t-test). This result could, in principle, be consistent with our computational model. However, when considering pattern and component cells separately in both areas, we found that not only the component cells were largely suppressed on average (green dots), similarly to the whole population, but the pattern cells too (V1 - component cells: -27.9%, pattern cells: -30.9%; LM – component cells: -45.6, pattern cells: -42.8%). This is at odd with the asymmetric suppression assumption needed to explain our behavioural data in chapter 4.

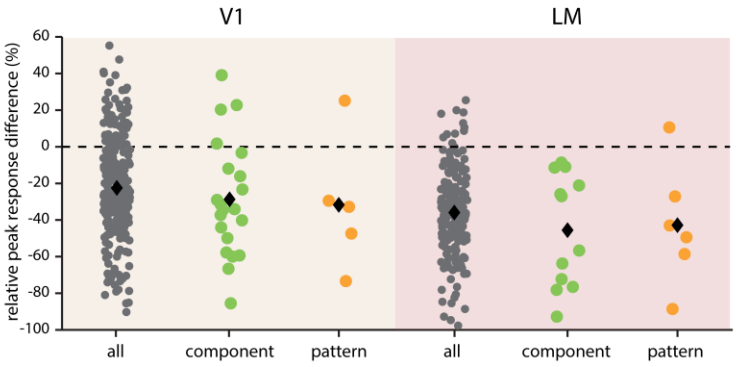

Figure 3. Cross-orientation suppression is common in V1 and LM pattern and component cells. Scatter plots in figure display the relative peak response difference distributions in V1 and LM areas. Gray clouds represents the distribution of all neurons in each area (irrespectively to their classification), green clouds display component cells whereas orange ones display pattern cells. Black diamonds mark average values. Negative values of relative peak response difference indicate cross-orientation suppression (i.e. smaller peak response to plaids than to gratings) while positive values indicate cross-orientation facilitation (i.e. smaller peak response to plaids than to gratings). Both areas display a predominance of cross-orientation suppression over facilitation, independently of the identity (i.e. pattern or component) of neurons taken into account.

Discussion

This study is the first to investigate the pattern and component cell distribution in the visual cortex of the rat. With our results, we corroborate the notion that rodent V1 hosts a sizable fraction of neurons that are sensitive to global motion[12,13]. At the same time we also provide a

confirmation to the view of LM as an extrastriate area that is rich in pattern cells[14]. On the other hand, our data are at odd with the findings of Juavinett et al.[14] about the presence of pattern cells in RL (see Table 1 for a summary of rodent studies).

source study	area	#units	method	anesthesia type	pattern fraction	component fraction	unclassified fraction
Muir et al. 2015	V1	4088	ophys	fentanyl + medetomidin	3%	31%	66%
Juavinett et al. 2015	V1	4743	ophys	isofluorane	0%	30%	70%
Palagina et al. 2017	V1	1119	ophys	isofluorane	10%	17%	73%
our study	V1	285	ephys	fentanyl + medetomidin	**6%**	**27%**	**67%**
Juavinett et al. 2015	LM	4577	ophys	isofluorane	6%	38%	56%
our study	LM	308	ephys	fentanyl + medetomidin	**8%**	**16%**	**76%**
Juavinett et al. 2015	RL	5232	ophys	isofluorane	8%	26%	66%
our study	RL	322	ephys	fentanyl + medetomidin	**0%**	**23%**	**77%**

Table 1. Summary of rodent studies investigating the distribution of pattern and component cells across rodent V1, LM and RL. The first column identifies the study. The second column indicates the cortical area of interest. The third column reports the neuronal sample size. The fourth column indicates whether the study was based on electrophysiological (i.e. "ephys") or optical (i.e. "ophys") recording methods. The fifth column reports the type of anesthetic employed. Last three columns report pattern, component and unclassified cell fractions.

The most likely argument we can conceive to explain such disagreement is related to anesthesia. It is well established that pattern cells are almost absent from V1 of anesthetized monkeys[4,6,9,10], although Guo et al.[33] reported a sizable amount of them (9%) being present in awake animals. Such responses were likely enabled by feedback from higher level areas that are cut off during anesthesia. This demonstrates how, in principle, anesthesia levels can affect motion integration computations. Consistently, Palagina et al.[13], who did observe pattern cells in V1, used lower isoflurane concentrations (0.6%) than Juavinett. et al.[14] (from 0.6 to 1.2%), who did not observe them. Interestingly, Muir et al.[12] who used a fentanyl and medetomidin anesthesia similar to our own, reported proportions of pattern and component cells in V1 matching ours (3% and 31% against 6% and 27%).

These considerations would suggest that the anesthesia we and Muir et al.[12] used corresponds more closely to the light levels of the isoflurane-based one used by Palagina et al.[13] than to the deep ones used by Juavinett et al.[14] While this would reasonably reconcile the results for V1, the disagreement between Juavinett et al.[14] and our study on the fraction of RL pattern cells is hard to explain, unless differential area-specific and species-specific effects of anaesthetics are considered. In fact, even if fentanyl and medetomidin anesthesia has been demonstrated to leave unaffected even subtle aspects of visual processing in rodent V1 and extrastriate areas[19,34–36], one way to make sense of the contrasting observations in RL would

be to posit that motion integration is more sensitive to anesthesia in extrastriate dorsal areas than in V1. To confirm this hypothesis, further studies systematically comparing pattern fractions across extrastriate areas in awake and anesthetized rats would be needed.

Beside contributing to the above discussed debate, our study is the first carrying out STA-based linear RF reconstruction and LN-modelling of neurons identified as pattern or component cells. This enabled us to directly test a longstanding question concerning the nature of rodent pattern selective units: whether their observed selectivity may be explained as the result of a trivial linear mechanisms and the geometry of their RF[13,21] (i.e. blobbyness hypothesis). Our results show that this is not the case – while the linear RF accounted at least partially for the tuning of the component cells, it failed to account for the tuning of the pattern cells. This goes against the blobbyness hypothesis, pointing to the role of nonlinear mechanisms in shaping the observed tuning for global motion. Together, these results point to pattern cells we and other investigators observed in V1[12,13] and LM[14] as potentially similar to those found in the primate dorsal stream.

In spite of this conclusion, whether V1 and LM may be the putative neural substrate underlying the execution of the behavioural task described in chapter 4 remains unclear, due to the discrepancy between the asymmetric cross-orientation suppression postulated by our modelling and the symmetric suppression we observed in our neural data. Such disagreement could arguably be explained in two different ways (besides by the lack of robust statistics for pattern cells): either the areas supporting pattern and grating perception in our task were indeed neither V1 nor LM, or those areas were actually involved (as the presence of pattern cells would suggest) but cross-orientation suppression effects might be strongly impacted by anaesthesia and thus be different in the awake behaving animal.

Materials and methods

Animal preparation and surgery

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 neuronal recordings from 29 naïve male Long Evans rats, weighted 300-700 grams and aged 3-12 months. Each rat was anesthetized with an intraperitoneal (IP) injection of a solution of 0.3 mg/kg of fentanyl (Fentanest®, Pfizer) and 0.3 mg/kg of medetomidin (Domitor®, Orion Pharma). During the surgery, we monitored the anesthesia level by checking the animal paw reflex and by measuring the oxygenation and heart rate through a pulse oximeter (Pulsesense-VET, Nonin). To avoid anesthesia-induced hypothermia, temperature was monitored and maintained at 37° C through a heating pad. In order to prevent hypoxia, a constant flux of oxygen was delivered to the animal throughout the surgery. A constant level of anesthesia was maintained by continuously delivering an IP injection of the same aesthetic solution used for the induction, but at a lower concentration (0.1 mg/kg/h Fentanyl and 0.1 g/kg/h Medetomidin), by means of a syringe pump (NE-500;

New Era Pump Systems). Once deeply anesthetized, the animal was secured to a stereotaxic apparatus (Narishige, SR-5R) and we performed a craniotomy on the left hemisphere, over the selected target (typically, a ~ 4 mm^2 window). Stereotaxic coordinates of the center of the craniotomy were 6.5 mm AP and 4 mm ML for V1-targeted sessions, 7 mm AP and 5 mm ML for LM-targeted sessions and 5 mm AP and 4.5 mm ML for RL-targeted sessions.

To keep eyes hydrated during the surgery, we protected them by delivering an ophthalmic ointment (Epigel®, Ceva Vet). Once completed the surgery, the rat was placed over a rotating platform, with the right eye just in front of the center of the screen (distance = 30 cm) and the left eye covered with black tape. The right eye was fixed through a metal eye-ring.

Neuronal recordings

Extracellular recordings were performed in anesthetized rats passively exposed to visual stimulation using either single shank 32-channels or double shank 64-channels silicon probes (NeuroNexus technologies). The insertion of the electrode into the cortex was performed through an oil hydraulic micromanipulator (Narishige, MO-10). The insertion depth was different for each area: for V1 and RL, it was ~900 μm with an electrode angle relative to the cortical surface of ~20°, for LM ~1500 μm, with an electrode slope of ~25°. Our recordings sampling rate was 25 kHz, and the extracellular signal was acquired and pre-amplified through a system three TDT (Tucker-Davis Technologies) workstation. To reconstruct the path of the electrode insertion through a post-mortem histology, we coated the probe with Vybrant® DiI cell-labelling solution (Invitrogen, Oregon, USA).

Single unit isolation

Single-units (n = 258 in V1, n = 187 in LM and n = 184 in RL) were isolated offline using the KlustaKwik-Phy software package[37]. After the automatic spikes detection and features extraction, we performed a manual refinement of the sorting through the "Kwik-GUI" interface. The manual refinement of the automatic output was sorted based on the following criteria: i) the compactness of the clusters in the space of the principal components of the waveforms; ii) the shape of the auto- and cross-correlogram (the latter was used to decide whether to merge or not two clusters); iii) the variation of the principal components of the waveform over time – this was especially useful to take into account possible electrode drifts; iv) the shape of the average waveform – this was especially useful to detect artefacts or non-physiological signals. When we suspected that a cluster was composed by multiple single units, we ran the "reclustering" function of the GUI, in an attempt to split it into the component single units. In order to be included in the following step of the analysis, Single Units (SUs) should meet these criteria: i) they should show a clear refractory period (less than 0.5% of the spikes present in <2 ms in the autocorrelogram); ii) they should be responsive – a response firing rate exceeding 2 Hz.

Visual Stimuli

Our stimulation protocol was composed of two different sub-protocols: a first one, aimed at coarsely mapping the receptive field of units across the electrode shank length (needed to

identify in an online manner the identity of the insertion area based on its retinotopy), and the main protocol containing the stimuli used to characterize neurons as pattern or component neurons and to reconstruct their linear RFs.

The first sub-protocol consisted in a ~15 minutes long coarse RF mapping procedure. The aim of this protocol was to infer the retinotopy at the recording site in order to establish the identity of the insertion area (to obtain feedback on whether it matched the intended target area or not). In this small sub-protocol, stimuli consisted of 10° long drifting bars presented in different orientations (0°, 45°, 90°, 135°). The bar, a white rectangle over a black background, was presented in 66 different positions, in 6 rows (spanning vertically 50°) and 11 columns (spanning horizontally of 100°) grid. An online plotting of the Multi-Unit-Activity (MUA) allowed us to identify the RF center of the insertion site. The online identification of the area could be performed thanks to the "inversion" (i.e. from nasal to temporal instead of from temporal to nasal) of the retinotopy present in LM and RL compared with V1.

The second, main, sub-protocol included full-field, full-contrast sinewave drifting gratings and plaids (120° of aperture angle) of different spatial (0.02 and 0.04 cpd) and temporal (2 or 6 Hz) frequency, as well as 12 equi-spaced drift directions (from 0° to 330°). Those stimuli were presented at 60 Hz frame rate and repeated 20 times (trials) across each recording experiment. The other kind of stimulus present in the main sub-protocol was a 0.04 cpd-matched-contrast modulated Gaussian noise, presented in 40 60-seconds-long chunks at 30 Hz frame rate. Grating and plaid trials as well as noise chunks were randomly interleaved with an Inter Stimulus Interval (ISI) of 1 second. During the ISI, a uniform, luminance-matched grey coloured screen was shown (this was instrumental to estimate the spontaneous firing rate of each unit). The total duration of the main sub-protocol was 102 minutes.
Stimulus presentation was controlled in Matlab through Psychtoolbox-based custom-made functions. Stimuli were displayed on a 47-inch, gamma-corrected, LCD monitor (SHARP PNE471R) with 1920x1080 pixel resolution, placed at 30 cm from the eye of the animal.

Classification of neurons

In order to categorize a neuron as pattern or component direction selective, the first, preliminary, step is to select direction selective units among the recorded neurons. To do so we have to quantify direction selectivity of single units in each area computing the Direction Selectivity Index (DSI):

$$\mathrm{DSI} = \left(R_{\mathrm{pref}} - R_{\mathrm{opposite}}\right)/\left(R_{\mathrm{pref}} + R_{\mathrm{opposite}}\right)$$

where R_{pref} is the response of the neuron to the preferred direction and R_{opposite} is the response to the opposite direction. Neurons with a DSI > 0.33 were categorized as direction selective. Among those neurons only plaid-responsive ones were considered viable for the classification.
Plaid-responsiveness was quantitatively defined as having a z-scored plaid tuning curve with a peak exceeding 5σ. After this second preparatory step we were ready to classify the neurons that passed the two previous selections as pattern, component or unclassified on the basis of their Fisher transformed partial correlation indexes (Zp and Zc, defined below).

$$R_c = \frac{(r_c - r_p r_{pc})}{\sqrt{\left((1 - r_p^2)(1 - r_{pc}^2)\right)}} \qquad R_p = \frac{(r_p - r_c r_{pc})}{\sqrt{\left((1 - r_c^2)(1 - r_{pc}^2)\right)}}$$

$$Z_c = \frac{\sqrt{(n-3)}}{2} \ln \frac{1 + R_c}{1 - R_c} \qquad Z_p = \frac{\sqrt{(n-3)}}{2} \ln \frac{1 + R_p}{1 - R_p}$$

where r_c is the correlation coefficient between observed and component predicted plaid tuning curve, r_p is the correlation coefficient between observed and pattern predicted plaid, r_{pc} is the correlation coefficient between the two predictions and *n* is the number of elements of the tuning curves. Pattern prediction is simply equal to the observed grating tuning curve. Component prediction is obtained from observed grating tuning curve shifting it leftwards and rightwards of half plaid cross angle and averaging the resulting curves. At 90% confidence (as usually done in the literature) the Z critical value adopted was 1.28 defining the pattern and component regions of the Zc Zp plane.

pattern condition:

$$Z_c - 0 > 1.28 \text{ if } Z_c < 0$$
$$Z_p - Z_c > 1.28 \text{ if } Z_c > 0$$

component condition:

$$Z_c - Z_p > 1.28 \text{ if } Z_p > 0$$
$$Z_c - 0 > 1.28 \text{ if } Z_p < 0$$

Neurons which did not meet these criteria were categorized as unclassified.

RF reconstruction and LN modelling

Reconstruction of linear RFs underlying selectivity of recorded neurons was achieved by means of the Spike-Triggered Average (STA) analysis[30]. 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. To take into account the correlation structure of our stimulus ensemble and reduce artefactual blurring of the reconstructed filters, we "decorrelated" the resulting STA images by dividing them by the covariance matrix of the whole stimulus ensemble[30], using Tikhonov regularization to handle covariance matrix inversion.

Statistical significance of the STA images was then assessed pixelwise, by applying the following 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 SD of this null distribution. The temporal span of the spatiotemporal linear kernel reconstructed via STA extended from spike generation time up to 330 ms before. This corresponds to a duration of 10 frames of the noise movie played at 30 Hz. These procedures were performed on downsampled noise frames (16x32 pixels). Resulting STA images that were later spline interpolated at higher resolution for better visualization and for LN prediction.

After estimating the linear spatiotemporal structure of a RF with STA, we used it as a filter in the input stage of a classical LN model of stimulus-response relationship[30]. 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 correlation coefficient between the two.

Cross-orientation suppression quantification

To this aim we devised an index defined as the relative firing rate difference between peak plaid and grating responses, leading to a negative value for cross-orientation suppressed units and to a positive value for cross-orientation facilitated units.
Formally the index (COI) was defined as:

$$\mathrm{COI} = \frac{(k_p - k_g)}{max(k_p, k_g)}$$

with k_p and k_c indicating peak response firing rate to plaids and gratings respectively.

References

1. Marr, D. *Vision: a computational investigation into the human representation and processing of visual information*. (MIT Press, 2010).
2. Fennema, C. L. & Thompson, W. B. Velocity determination in scenes containing several moving objects. *Comput. Graph. Image Process.* **9**, 301–315 (1979).
3. Wuerger, S., Shapley, R. & Rubin, N. "On the Visually Perceived Direction of Motion" by Hans Wallach: 60 Years Later. *Perception* **25**, 1317–1367 (1996).
4. Movshon, J. A. *et al.* Pattern recognition mechanisms. (1985).
5. Rodman, H. R. & Albright, T. D. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp. Brain Res.* **75**, (1989).
6. Movshon, J. A. & Newsome, W. T. Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys. *J. Neurosci.* **16**, 7733–7741 (1996).

7. Smith, M. A., Majaj, N. J. & Movshon, J. A. Dynamics of motion signaling by neurons in macaque area MT. *Nat. Neurosci.* **8**, 220–228 (2005).
8. Rust, N. C., Mante, V., Simoncelli, E. P. & Movshon, J. A. How MT cells analyze the motion of visual patterns. *Nat Neurosci* **9**, 1421–31 (2006).
9. Khawaja, F. A., Tsui, J. M. G. & Pack, C. C. Pattern Motion Selectivity of Spiking Outputs and Local Field Potentials in Macaque Visual Cortex. *J. Neurosci.* **29**, 13702–13709 (2009).
10. Khawaja, F. A., Liu, L. D. & Pack, C. C. Responses of MST neurons to plaid stimuli. *J. Neurophysiol.* **110**, 63–74 (2013).
11. Solomon, S. S. *et al.* Visual motion integration by neurons in the middle temporal area of a New World monkey, the marmoset: Pattern motion sensitivity in middle temporal area of marmosets. *J. Physiol.* **589**, 5741–5758 (2011).
12. Muir, D. R., Roth, M. M., Helmchen, F. & Kampa, B. M. Model-based analysis of pattern motion processing in mouse primary visual cortex. *Front. Neural Circuits* **9**, (2015).
13. Palagina, G., Meyer, J. F. & Smirnakis, S. M. Complex Visual Motion Representation in Mouse Area V1. *J. Neurosci.* **37**, 164–183 (2017).
14. Juavinett, A. L. & Callaway, E. M. Pattern and Component Motion Responses in Mouse Visual Cortical Areas. *Curr. Biol.* **25**, 1759–1764 (2015).
15. Marques, T. *et al.* A Role for Mouse Primary Visual Cortex in Motion Perception. *Curr. Biol. CB* **28**, 1703-1713.e6 (2018).
16. Rasmussen, R. & Yonehara, K. Circuit Mechanisms Governing Local vs. Global Motion Processing in Mouse Visual Cortex. *Front. Neural Circuits* **11**, 109 (2017).
17. Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional Specialization of Mouse Higher Visual Cortical Areas. *Neuron* **72**, 1025–1039 (2011).
18. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional Specialization of Seven Mouse Visual Cortical Areas. *Neuron* **72**, 1040–1054 (2011).
19. Tafazoli, S. *et al.* Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6**, e22794 (2017).
20. Wang, Q., Sporns, O. & Burkhalter, A. Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J. Neurosci.* **32**, 4386–4399 (2012).
21. Tinsley, C. J. *et al.* The Nature of V1 Neural Responses to 2D Moving Patterns Depends on Receptive-Field Structure in the Marmoset Monkey. *J. Neurophysiol.* **90**, 930–937 (2003).
22. De Valois, R. L., William Yund, E. & Hepler, N. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 531–544 (1982).
23. De Valois, R. L., Albrecht, D. G. & Thorell, L. G. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559 (1982).
24. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. *J. Neurosci.* **22**, 5639–5651 (2002).

25. Gizzi, M. S., Katz, E., Schumer, R. A. & Movshon, J. A. Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *J. Neurophysiol.* **63**, 1529–1543 (1990).
26. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
27. Girman, S. V., Sauvé, Y. & Lund, R. D. Receptive field properties of single neurons in rat primary visual cortex. *J. Neurophysiol.* **82**, 301–311 (1999).
28. Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B. & Zoccolan, D. Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J. Neurosci.* **39**, 1649–1670 (2019).
29. Simoncelli, E. P. & Heeger, D. J. A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–761 (1998).
30. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *J. Vis.* **6**, (2006).
31. Aljadeff, J., Lansdell, B. J., Fairhall, A. L. & Kleinfeld, D. Analysis of Neuronal Spike Trains, Deconstructed. *Neuron* **91**, 221–259 (2016).
32. Ibrahim, L. A. *et al.* Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. *Neuron* **89**, 1031–1045 (2016).
33. Guo, K., Benson, P. J. & Blakemore, C. Pattern motion is present in V1 of awake but not anaesthetized monkeys: Motion processing in V1. *Eur. J. Neurosci.* **19**, 1055–1066 (2004).
34. Zhu, Y. & Yao, H. Modification of Visual Cortical Receptive Field Induced by Natural Stimuli. *Cereb. Cortex* **23**, 1923–1932 (2013).
35. Pecka, M., Han, Y., Sader, E. & Mrsic-Flogel, T. D. Experience-Dependent Specialization of Receptive Field Surround for Selective Coding of Natural Scenes. *Neuron* **84**, 457–469 (2014).
36. Froudarakis, E. *et al.* Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* **17**, 851–857 (2014).
37. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).

Conclusion

While the studies described in this thesis where ongoing, many papers were published that deserve to be mentioned in order to thoroughly frame the results we reported in the broader context of the current literature.

Notably, the last few months saw what is arguably the main industrial-scale "big science" initiative in neuroscience get to the data-release phase. First results of Allen Institute's large-scale survey of the coding properties of mouse visual system[128] where reported in two preprints[129,130]. Recent Allen Institute findings are particularly relevant for the identification of the hierarchical organization of rodent visual areas as well as for the ventral/dorsal homology hypothesis.

The first preprint[129] reported the results from a two-photon calcium imaging survey of 60.000 neurons of 6 visual areas (V1, LM, AL, RL, PM and AM) of the mouse cortex under presentation of gratings, noise and natural stimuli. A 70% of neurons was significantly responsive to stimulation showing very high trial to trial variability (i.e. response to ~ 50% of trials in the preferred condition on average) and sparse responses, in particular to natural scenes. Furthermore, when fitting the responses with a model using, as features, spatiotemporal wavelets followed by either rectifying or quadratic nonlinearities, the authors find out that the latter always had more weight. This was interpreted, at odd with numerous previous findings[79,94,97,98,131] (including our rat results), as the mouse visual cortex containing complex cells only. As a last step, the authors analysed the similarity of the observed neural representations with those in VGG16 (the same network we used in chapter I study) showing a peak of V1 correlation with middle layers of the artificial network and a peak correlation of LM, PM and AL with slightly higher ones. This is consistent with those areas being located at two close but different levels in a functional hierarchy.

The second preprint describes the results of an electrophysiological survey with high density Neuropixel probes[132] across the same 6 areas, totalling ~ 40.000 single units. Nearly 60% of those units displayed significant visual receptive fields. Notably, one of the four analyses described in the paper was explicitly inspired by our work, described in chapter I. The authors used our MI index to categorize neurons as simple or complex on the ground of the phase dependency of their grating responses. In contrast with what suggested by the calcium imaging survey cited above, they found in V1 a bimodal distribution of MI values consistent with the coexistence of simple and complex cells in this area, as evidenced by our and previous studies. Interestingly, the increase in position tolerance observed across extrastriate areas was highly correlated with their "anatomical hierarchy score" (derived from another recent Allen Institute study probing the mesoscale connectome of mouse visual cortex[133]). This further validates the MI as a useful measure for hierarchical level.

Besides the industrial scale science efforts described above, some recent papers targeted specifically putative ventral areas, adding new pieces of evidence to the picture delineated so far. Cadena et al.[134] applied an approach, inspired to the one used by Yamins et al.[135] in the primate ventral stream, to predict visual responses observed imaging mouse V1, LM, AL and RL. In order to do so they regressed neural responses with visual features extracted from different layers of VGG16. To their surprise, they found that activity in all areas was optimally

predicted by 4th-layer VGG features, in contrast with what would be expected assuming the existence of a processing hierarchy among those areas. Even more surprisingly, the authors found that the same network with randomized weights performed equally better in predicting the neural activity. One key consideration that could explain the first puzzling finding (and maybe partially even the second) is the choice of the extrastriate areas targeted for the study. In fact, the authors didn't target putative higher-level shape processing areas such as LI, LL and POR but instead an area that seems to be shared at the start of both processing streams (LM) and two plainly dorsal ones (AL, RL). Also, according to the latest Allen Institute anatomical survey cited above[133], those areas are among the lowest in terms of hierarchical scoring. All these considerations make the inconclusive result reported by Cadena et al. less surprising considering the less than ideal choice of areas to be targeted. A different study, by Nishio et al.[136], found in ventral area POR, evidence of invariance to stimulus size, consistently with the tendency observed in lateral areas to progressively discard low level luminance information[127]. Han et al.[137] imaged mouse visual cortex while presenting noise stimuli with different orientation content. In doing so they evidenced a decrease in preference for stimuli containing a single orientation from V1 to LI, corroborating our result obtained using gratings[131]. Finally, a new study from our lab showed an increase in the activity time-scale and object decoding temporal stability across the V1→LM→LI→LL hierarchy under natural movie stimulation[138]. This is consistent with the idea that increasing temporal slowness is the signature of a shape processing hierarchy progressively building up transformation invariance.

On the other hand, other recent studies advanced significantly the characterization of putative dorsal areas in mice. Diamanti et al.[139], while documenting the widespread spatial modulation of neurons across the whole mouse visual cortex, showed how navigation increases the reliability of neural responses in AM, PM and A. This is consistent with their tentative placement in the dorsal subnetwork concerned with optic flow analyses. Sit et al.[140] used two photon calcium imaging to probe mouse extrastriate cortex with random dot kinematograms (RDK) and natural movies. They found that a high (>70%) proportion of units is responsive to coherent motion in all areas, with a bias towards the lower visual field. The strongest motion responsivities were observed in areas AM, PM and AL, whereas the lower ones in LI and LM. Such result is consistent with the proposed dorsal and ventral assignation of those areas. RL, on the other hand, violated expectations displaying high coherent motion responsivity under movie stimulation, but not under RDK one. In another recent work[141], reporting a high fraction (>50%) of binocular disparity tuned units across all mouse visual cortex, RL emerged as an area specialized for processing visual objects very close to the animal (between ~ 2 and 7.0 cm from the eyes against ~ 5 to 20 cm for V1 and LM). Interestingly, RL-preferred distance range falls within reach of the mouse's whiskers. Such new evidence, together with a previously reported retinotopic overrepresentation of the lower visual field[108], corroborates the view of RL as specialized for visuotactile integration[124]. Containing matched visual and somatosensory maps, in fact, RL may form a multimodal representation of near space in front of the animal, likely playing a key role for vision- and active-whisking-mediated object exploration. Such putative role would place RL firmly in the dorsal sensorimotor pathway, with a possible homology with the "parietal reach region" in primate PPC, coordinating vision and arm/hand movements for object interaction[142].

Finally, even on the postnatal unsupervised learning front, a relevant paper was recently published. Crijns et al.[143] set themselves to test the occurrence of unsupervised learning of scale invariance in V1 of adult rats, similarly to what Li et al.[144] did in monkey IT. Animals were repeatedly exposed to a temporal transition between two gratings of different SF and different orientation (or the same orientation in case of controls). A temporal contiguity-based learning principle would predict the neurons originally selective for a given orientation across the two SF would develop responsivity to the new orientation at the new SF. This would mean effectively "breaking" the SF-invariant orientation selectivity of those neuron trough exposure to an altered temporal visual statistic. The experiment however yielded negative results, showing such manipulation to be ineffective in altering the selectivity of the neuros. This could be likely due to the choice of performing the experiment in V1 of an adult animal where the levels of plasticity are likely quite low, no longer enabling such dramatic experience-dependent changes of the basic building blocks of perceptual representations.

The discussion of the recent literature reported above shows the rich research landscape in which the research work described in this thesis sits.

Summing up, with the first subproject (chapter I) we provided new functional evidence supporting the homology of rat lateral extrastriate cortex with the monkey ventral stream. With the second subproject (chapter II) we developed and validated a novel, automated, method for laminar identification of multichannel silicon probes recording sites. With the third subproject (chapter III) we provided a causal demonstration that temporal continuity of the early postnatal visual experience plays a key role in the learning of position tolerance in V1. With the fourth subproject (chapter IV) we reported the first direct behavioral assessment of the ability of rats to spontaneously integrate motion direction across multiple moving elements. With the last, fifth subproject (chapter V) we investigated the neuronal substrate of such ability, showing how in both rat V1 and LM, pattern cell exist whose behavior is not accounted for by their linear receptive field.

Overall, we hope that the body of work described in this thesis may have provided a vivid example of how rats can serve as an extremely interesting, innovative and useful animal model for the study of cortical visual information processing.

References

1. Cox, D. D. Do we understand high-level vision? *Curr. Opin. Neurobiol.* **25**, 187–193 (2014).
2. Ullman, S. *High Level Vision*. (MIT Press, 1996).
3. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
4. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 770–778 (2016).
5. Munasinghe, S., Fookes, C. & Sridharan, S. Human-level face verification with intra-personal factor analysis and deep face representation. *IET Biom.* **7**, 467–473 (2018).
6. Xu, K. *et al.* Show, attend and tell: Neural image caption generation with visual attention. in *International conference on machine learning* 2048–2057 (2015).

7. Fenno, L., Yizhar, O. & Deisseroth, K. The Development and Application of Optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
8. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295 (2013).
9. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
10. Carandini, M. & Churchland, A. K. Probing perceptual decisions in rodents. *Nat. Neurosci.* **16**, 824–831 (2013).
11. Zoccolan, D. Invariant visual object recognition and shape processing in rats. *Behav. Brain Res.* **285**, 10–33 (2015).
12. Reinagel, P. Using rats for vision research. *Neuroscience* **296**, 75–79 (2015).
13. Ungerleider, L. G., Mishkin, M., Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W. Analysis of visual behavior. (1982).
14. Ungerleider, L. G. & Pasternak, T. Ventral and dorsal cortical processing streams. *Vis. Neurosci.* **1**, 541–562 (2004).
15. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25 (1992).
16. Goodale, M. A. & Keith Humphrey, G. The objects of action and perception. *Cognition* **67**, 181–207 (1998).
17. Galletti, C. & Fattori, P. The dorsal visual stream revisited: Stable circuits or dynamic pathways? *Cortex* **98**, 203–217 (2018).
18. Hong, H., Yamins, D. L. K., Majaj, N. J. & DiCarlo, J. J. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* **19**, 613–622 (2016).
19. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
20. Orban, G. A., Janssen, P. & Vogels, R. Extracting 3D structure from disparity. *Trends Neurosci.* **29**, 466–473 (2006).
21. Farivar, R., Blanke, O. & Chaudhuri, A. Dorsal–ventral integration in the recognition of motion-defined unfamiliar faces. *J. Neurosci.* **29**, 5336–5342 (2009).
22. Freud, E., Plaut, D. C. & Behrmann, M. 'What'is happening in the dorsal visual pathway. *Trends Cogn. Sci.* **20**, 773–784 (2016).
23. Yildirim, I., Wu, J., Kanwisher, N. & Tenenbaum, J. An integrative computational architecture for object-driven cortex. *Curr. Opin. Neurobiol.* **55**, 73–81 (2019).
24. Glickfeld, L. L. & Olsen, S. R. Higher-Order Areas of the Mouse Visual Cortex. *Annu. Rev. Vis. Sci.* **3**, 251–273 (2017).
25. Pinto, N., Cox, D. D. & DiCarlo, J. J. Why is real-world visual object recognition hard? *PLoS Comput Biol* **4**, e27 (2008).
26. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How Does the Brain Solve Visual Object Recognition? *Neuron* **73**, 415–434 (2012).
27. Kar, K., Kubilius, J., Schmidt, K., Issa, E. B. & DiCarlo, J. J. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* **22**, 974 (2019).

28. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* **160**, 106–54 (1962).
29. Fukushima, K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* **36**, 193–202 (1980).
30. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat Neurosci* **2**, 1019–25 (1999).
31. Simoncelli, E. P. & Heeger, D. J. A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–761 (1998).
32. Movshon, J. A. *et al.* Pattern recognition mechanisms. (1985).
33. Rust, N. C., Mante, V., Simoncelli, E. P. & Movshon, J. A. How MT cells analyze the motion of visual patterns. *Nat Neurosci* **9**, 1421–31 (2006).
34. Born, R. T. & Bradley, D. C. Structure and function of visual area MT. *Annu Rev Neurosci* **28**, 157–189 (2005).
35. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183 (1954).
36. Barlow, H. B. Possible principles underlying the transformation of sensory messages. *Sens. Commun.* **1**, 217–234 (1961).
37. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–9 (1996).
38. Wiskott, L. & Sejnowski, T. J. Slow feature analysis: unsupervised learning of invariances. *Neural Comput* **14**, 715–70 (2002).
39. Wiskott, L. How does our visual system achieve shift and size invariance. *JL Van Hemmen TJ Sejnowski Ed.* **23**, 322–340 (2006).
40. Levelt, C. N. & Hübener, M. Critical-period plasticity in the visual cortex. *Annu. Rev. Neurosci.* **35**, 309–330 (2012).
41. Berardi, N., Pizzorusso, T. & Maffei, L. Critical periods during sensory development. *Curr. Opin. Neurobiol.* **10**, 138–145 (2000).
42. Hensch, T. K. *et al.* Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science* **282**, 1504–1508 (1998).
43. Bartoletti, A., Medini, P., Berardi, N. & Maffei, L. Environmental enrichment prevents effects of dark-rearing in the rat visual cortex. *Nat. Neurosci.* **7**, 215–216 (2004).
44. Gianfranceschi, L. *et al.* Visual cortex is rescued from the effects of dark rearing by overexpression of BDNF. *Proc. Natl. Acad. Sci.* **100**, 12486–12491 (2003).
45. Cynader, M. & Chernenko, G. Abolition of direction selectivity in the visual cortex of the cat. *Science* **193**, 504–505 (1976).
46. Daw, N. W. & Wyatt, H. J. Kittens reared in a unidirectional environment: evidence for a critical period. *J. Physiol.* **257**, 155–170 (1976).
47. Blasdel, G. G., Mitchell, D. E., Muir, D. W. & Pettigrew, J. D. A physiological and behavioural study in cats of the effect of early visual experience with contours of a single orientation. *J. Physiol.* **265**, 615–636 (1977).
48. Sasaki, K. S. *et al.* Supranormal orientation selectivity of visual neurons in orientation-restricted animals. *Sci. Rep.* **5**, 16712 (2015).
49. Yoshida, T., Ozawa, K. & Tanaka, S. Sensitivity Profile for Orientation Selectivity in the Visual Cortex of Goggle-Reared Mice. *PLOS ONE* **7**, e40630 (2012).

50. Pecka, M., Han, Y., Sader, E. & Mrsic-Flogel, T. D. Experience-Dependent Specialization of Receptive Field Surround for Selective Coding of Natural Scenes. *Neuron* **84**, 457–469 (2014).
51. Richards, B. A. *et al.* A deep learning framework for neuroscience. *Nat. Neurosci.* **22**, 1761–1770 (2019).
52. Krubitzer, L. The Magnificent Compromise: Cortical Field Evolution in Mammals. *Neuron* **56**, 201–208 (2007).
53. Sereno, M. I. & Allman, J. M. Cortical visual areas in mammals. *Neural Basis Vis. Funct.* **4**, 160–172 (1991).
54. Aoki, R., Tsubota, T., Goya, Y. & Benucci, A. An automated platform for high-throughput mouse behavior and physiology with voluntary head-fixation. *Nat. Commun.* **8**, 1196 (2017).
55. Huberman, A. D. & Niell, C. M. What can mice tell us about how vision works? *Trends Neurosci.* **34**, 464–473 (2011).
56. Chen, G., King, J. A., Burgess, N. & O'Keefe, J. How vision and movement combine in the hippocampal place code. *Proc. Natl. Acad. Sci.* **110**, 378–383 (2013).
57. Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M.-B. & Moser, E. I. Object-vector coding in the medial entorhinal cortex. *Nature* **568**, 400–404 (2019).
58. De Franceschi, G., Vivattanasarn, T., Saleem, A. B. & Solomon, S. G. Vision Guides Selection of Freeze or Flight Defense Strategies in Mice. *Curr. Biol.* **26**, 2150–2154 (2016).
59. Hoy, J. L., Yavorska, I., Wehr, M. & Niell, C. M. Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. *Curr. Biol.* **26**, 3046–3052 (2016).
60. Nakashima, S. F., Ukezono, M., Nishida, H., Sudo, R. & Takano, Y. Receiving of emotional signal of pain from conspecifics in laboratory rats. *R. Soc. Open Sci.* **2**, 140381.
61. Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H. & Zoccolan, D. Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy. *Curr. Biol.* **28**, 1005-1015.e5 (2018).
62. Alemi-Neissi, A., Rosselli, F. B. & Zoccolan, D. Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition. *J. Neurosci.* **33**, 5939–5956 (2013).
63. Tafazoli, S., Di Filippo, A. & Zoccolan, D. Transformation-Tolerant Object Recognition in Rats Revealed by Visual Priming. *J. Neurosci.* **32**, 21–34 (2012).
64. Douglas, R. M., Neve, A., Quittenbaum, J. P., Alam, N. M. & Prusky, G. T. Perception of visual motion coherence by rats and mice. *Vision Res.* **46**, 2842–2847 (2006).
65. Samonds, J. M., Lieberman, S. & Priebe, N. J. Motion Discrimination and the Motion Aftereffect in Mouse Vision. *eNeuro* **5**, (2018).
66. Marques, T. *et al.* A Role for Mouse Primary Visual Cortex in Motion Perception. *Curr. Biol. CB* **28**, 1703-1713.e6 (2018).
67. Euler, T. & Wässle, H. Immunocytochemical identification of cone bipolar cells in the rat retina. *J. Comp. Neurol.* **361**, 461–478 (1995).
68. Muniz, J. A. P. C., de Athaide, L. M., Gomes, B. D., Finlay, B. L. & de Lima Silveira, L. C. Ganglion cell and displaced amacrine cell density distribution in the retina of the howler monkey (Alouatta caraya). *PloS One* **9**, e115291 (2014).

69. Chelazzi, L., Rossi, F., Tempia, F., Ghirardi, M. & Strata, P. Saccadic Eye Movements and Gaze Holding in the Head-Restrained Pigmented Rat. *Eur J Neurosci* **1**, 639–646 (1989).
70. Zoccolan, D., Graham, B. J. & Cox, D. D. A self-calibrating, camera-based eye tracker for the recording of rodent eye movements. *Front. Neurosci.* **4**, 193 (2010).
71. Wallace, D. J. *et al.* Rats maintain an overhead binocular field at the expense of constant fusion. *Nature* **498**, 65–69 (2013).
72. Hughes, A. A schematic eye for the rat. *Vision Res.* **19**, 569–588 (1979).
73. Hughes, A. The refractive state of the rat eye. *Vision Res.* **17**, 927–939 (1977).
74. Green, D. G. & Powers, M. K. Mechanisms of light adaptation in rat retina. *Vision Res.* **22**, 209–216 (1982).
75. Jacobs, G. H., Fenwick, J. A. & Williams, G. A. Cone-based vision of rats for ultraviolet and visible lights. *J Exp Biol* **204**, 2439–46 (2001).
76. Burn, C. C. What is it like to be a rat? Rat sensory perception and its implications for experimental design and rat welfare. *Appl. Anim. Behav. Sci.* **112**, 1–32 (2008).
77. Prusky, G. T., Harker, K. T., Douglas, R. M. & Whishaw, I. Q. Variation in visual acuity within pigmented, and between pigmented and albino rat strains. *Behav Brain Res* **136**, 339–48 (2002).
78. Friedman, L. J. & Green, D. G. Ganglion cell acuity in hooded rats. *Vision Res.* **22**, 441–444 (1982).
79. Girman, S. V., Sauvé, Y. & Lund, R. D. Receptive Field Properties of Single Neurons in Rat Primary Visual Cortex. *J. Neurophysiol.* **82**, 301–311 (1999).
80. Baden, T. *et al.* The functional diversity of retinal ganglion cells in the mouse. *Nature* **529**, 345 (2016).
81. Martersteck, E. M. *et al.* Diverse central projection patterns of retinal ganglion cells. *Cell Rep.* **18**, 2058–2072 (2017).
82. Piscopo, D. M., El-Danaf, R. N., Huberman, A. D. & Niell, C. M. Diverse visual features encoded in mouse lateral geniculate nucleus. *J. Neurosci.* **33**, 4642–4656 (2013).
83. Durand, S. *et al.* 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 (2016).
84. Denman, D. J. & Contreras, D. On parallel streams through the mouse dorsal lateral geniculate nucleus. *Front. Neural Circuits* **10**, 20 (2016).
85. Grubb, M. S. & Thompson, I. D. Quantitative characterization of visual response properties in the mouse dorsal lateral geniculate nucleus. *J. Neurophysiol.* **90**, 3594–3607 (2003).
86. Cruz-Martín, A. *et al.* A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex. *Nature* **507**, 358 (2014).
87. Seabrook, T. A., Burbridge, T. J., Crair, M. C. & Huberman, A. D. Architecture, Function, and Assembly of the Mouse Visual System. *Annu. Rev. Neurosci.* **40**, 499–538 (2017).
88. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates*. (Academic Press, 2013).

89. Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
90. Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local Diversity and Fine-Scale Organization of Receptive Fields in Mouse Visual Cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
91. Ringach, D. L. *et al.* Spatial clustering of tuning in mouse primary visual cortex. *Nat. Commun.* **7**, 12270 (2016).
92. Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51 (2013).
93. Espinoza, S. G. & Thomas, H. C. Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. *Brain Res.* **272**, 137–144 (1983).
94. Parnavelas, J. G., Burne, R. A. & Lin, C. S. Receptive field properties of neurons in the visual cortex of the rat. *Neurosci Lett* **27**, 291–6 (1981).
95. Martinez, L. M. & Alonso, J.-M. Complex receptive fields in primary visual cortex. *The neuroscientist* **9**, 317–331 (2003).
96. Reid, R. C. & Alonso, J.-M. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **378**, 281 (1995).
97. Mangini, N. J. & Pearlman, A. L. Laminar distribution of receptive field properties in the primary visual cortex of the mouse. *J. Comp. Neurol.* **193**, 203–222 (1980).
98. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
99. Gilbert, C. D. Laminar differences in receptive field properties of cells in cat primary visual cortex. *J. Physiol.* **268**, 391–421 (1977).
100. Dräger, U. C. Autoradiography of tritiated proline and fucose transported transneuronally from the eye to the visual cortex in pigmented and albino mice. *Brain Res.* **82**, 284–292 (1974).
101. Dräger, U. C. Receptive fields of single cells and topography in mouse visual cortex. *J. Comp. Neurol.* **160**, 269–289 (1975).
102. Wagor, E., Mangini, N. J. & Pearlman, A. L. Retinotopic organization of striate and extrastriate visual cortex in the mouse. *J. Comp. Neurol.* **193**, 187–202 (1980).
103. Montero, V. M. Retinotopy of cortical connections between the striate cortex and extrastriate visual areas in the rat. *Exp Brain Res* **94**, 1–15 (1993).
104. Sanderson, K. J., Dreher, B. & Gayer, N. Prosencephalic connections of striate and extrastriate areas of rat visual cortex. *Exp. Brain Res.* **85**, 324–334 (1991).
105. Coogan, T. A. & Burkhalter, A. Hierarchical organization of areas in rat visual cortex. *J. Neurosci.* **13**, 3749–3772 (1993).
106. Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and Areal Organization of Mouse Visual Cortex. *J. Neurosci.* **34**, 12587–12600 (2014).
107. Rhim, I., Coello-Reyes, G., Ko, H.-K. & Nauhaus, I. Maps of cone opsin input to mouse V1 and higher visual areas. *J. Neurophysiol.* **117**, 1674–1682 (2017).
108. Zhuang, J. *et al.* An extended retinotopic map of mouse cortex. *Elife* **6**, e18372 (2017).

109. Wang, Q., Gao, E. & Burkhalter, A. Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. *J. Neurosci.* **31**, 1905–1918 (2011).
110. Wang, Q., Sporns, O. & Burkhalter, A. Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J. Neurosci.* **32**, 4386–4399 (2012).
111. Tees, R. C. The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. *Behav. Brain Res.* **106**, 55–73 (1999).
112. Aggleton, J. P., Albasser, M. M., Aggleton, D. J., Poirier, G. L. & Pearce, J. M. Lesions of the rat perirhinal cortex spare the acquisition of a complex configural visual discrimination yet impair object recognition. *Behav. Neurosci.* **124**, 55 (2010).
113. Bussey, T. J., Muir, J. L. & Aggleton, J. P. Functionally dissociating aspects of event memory: the effects of combined perirhinal and postrhinal cortex lesions on object and place memory in the rat. *J. Neurosci.* **19**, 495–502 (1999).
114. Davies, M., Machin, P., Sanderson, D., Pearce, J. & Aggleton, J. Neurotoxic lesions of the rat perirhinal and postrhinal cortices and their impact on biconditional visual discrimination tasks. *Behav Brain Res* **176**, 274–283 (2007).
115. Prusky, G. T., Douglas, R. M., Nelson, L., Shabanpoor, A. & Sutherland, R. J. Visual memory task for rats reveals an essential role for hippocampus and perirhinal cortex. *Proc Natl Acad Sci U A* **101**, 5064–8 (2004).
116. Kolb, B. & Walkey, J. Behavioural and anatomical studies of the posterior parietal cortex in the rat. *Behav. Brain Res.* **23**, 127–145 (1987).
117. Kolb, B., Pittman, K., Sutherland, R. J. & Whishaw, I. Q. Dissociation of the contributions of the prefrontal cortex and dorsomedial thalamic nucleus to spatially guided behavior in the rat. *Behav. Brain Res.* **6**, 365–378 (1982).
118. Nassi, J. J. & Callaway, E. M. Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* **10**, 360–372 (2009).
119. Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional Specialization of Mouse Higher Visual Cortical Areas. *Neuron* **72**, 1025–1039 (2011).
120. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional Specialization of Seven Mouse Visual Cortical Areas. *Neuron* **72**, 1040–1054 (2011).
121. Glickfeld, L. L., Andermann, M. L., Bonin, V. & Reid, R. C. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nat. Neurosci.* **16**, 219–226 (2013).
122. Smith, I. T., Townsend, L. B., Huh, R., Zhu, H. & Smith, S. L. Stream-dependent development of higher visual cortical areas. *Nat. Neurosci.* **20**, 200 (2017).
123. Juavinett, A. L. & Callaway, E. M. Pattern and Component Motion Responses in Mouse Visual Cortical Areas. *Curr. Biol.* **25**, 1759–1764 (2015).
124. Olcese, U., Iurilli, G. & Medini, P. Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron* **79**, 579–593 (2013).
125. Vermaercke, B. *et al.* Functional specialization in rat occipital and temporal visual cortex. *J. Neurophysiol.* **112**, 1963–1983 (2014).

126. Vinken, K., Van den Bergh, G., Vermaercke, B., Beeck, O. de & P, H. Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. *Cereb. Cortex* **26**, 3310–3322 (2016).
127. Tafazoli, S. *et al.* Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6**, e22794 (2017).
128. Hawrylycz, M. *et al.* Inferring cortical function in the mouse visual system through large-scale systems neuroscience. *Proc. Natl. Acad. Sci.* **113**, 7337–7344 (2016).
129. de Vries, S. E. J. *et al. A large-scale, standardized physiological survey reveals higher order coding throughout the mouse visual cortex*. http://biorxiv.org/lookup/doi/10.1101/359513 (2018) doi:10.1101/359513.
130. Siegle, J. H. *et al. A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas*. http://biorxiv.org/lookup/doi/10.1101/805010 (2019) doi:10.1101/805010.
131. Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B. & Zoccolan, D. Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J. Neurosci.* **39**, 1649–1670 (2019).
132. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
133. Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
134. Cadena, S. A. *et al.* How well do deep neural networks trained on object recognition characterize the mouse visual system? in *NeurIPS Neuro AI Workshop* (2019).
135. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci.* **111**, 8619–8624 (2014).
136. Nishio, N. *et al.* Higher visual responses in the temporal cortex of mice. *Sci. Rep.* **8**, 11136 (2018).
137. Han, X., Vermaercke, B. & Bonin, V. *Segregated encoding of spatiotemporal features in the mouse visual cortex*. http://biorxiv.org/lookup/doi/10.1101/441014 (2018) doi:10.1101/441014.
138. Piasini, E., Soltuzu, L., Caramellino, R., Balasubramanian, V. & Zoccolan, D. *Intrinsic dynamics enhance temporal stability of stimulus representation along a visual cortical hierarchy*. http://biorxiv.org/lookup/doi/10.1101/822130 (2019) doi:10.1101/822130.
139. Diamanti, E. M. *et al. Spatial encoding in the visual pathway arises in cortex and depends on active navigation*. http://biorxiv.org/lookup/doi/10.1101/832915 (2019) doi:10.1101/832915.
140. Sit, K. K. & Goard, M. J. *Distributed and Retinotopically Asymmetric Processing of Coherent Motion in Mouse Visual Cortex*. http://biorxiv.org/lookup/doi/10.1101/791905 (2019) doi:10.1101/791905.
141. La Chioma, A., Bonhoeffer, T. & Hübener, M. Area-Specific Mapping of Binocular Disparity across Mouse Visual Cortex. *Curr. Biol.* **29**, 2954-2960.e5 (2019).
142. Andersen, R. A. & Buneo, C. A. Intentional Maps in Posterior Parietal Cortex. *Annu. Rev. Neurosci.* **25**, 189–220 (2002).

143. Crijns, E., Kaliukhovich, D. A., Vankelecom, L. & Op de Beeck, H. Unsupervised Temporal Contiguity Experience Does Not Break the Invariance of Orientation Selectivity Across Spatial Frequency. *Front. Syst. Neurosci.* **13**, 22 (2019).
144. Li, N. & DiCarlo, J. J. Unsupervised Natural Visual Experience Rapidly Reshapes Size-Invariant Object Representation in Inferior Temporal Cortex. *Neuron* **67**, 1062–1075 (2010).