SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Neuroscience Area – PhD course in

Cognitive Neuroscience

A study of rat perceptual sensitivity to multipoint correlations and their representation in the visual cortex

Candidate:
Riccardo Caramellino

Advisor:
Prof. Davide Zoccolan

Academic Year 2021-2022

Table of contents

- **Chapter I**: Introduction p. 1
- **Chapter II**: Rat sensitivity to multipoint statistics is predicted by efficient coding of natural scenes p. 31
- **Chapter III**: Coding of multipoint statistics in rat visual cortex p. 56
- **Chapter IV**: Conclusions p. 83
- Acknowledgements p. 85

Chapter I

Introduction

To survive in the environment an organism has to be highly sensitive to the most task-relevant sensory features. So, a fundamental goal of neuroscience has always been understanding how sensory stimuli are encoded in the brain to maximize perceptual sensitivity. Research indicates that evolutionary constraints, in particular metabolic and wiring limitations, have led sensory systems to accurately encode signals while minimizing the computational costs by reducing the redundancy intrinsically present in the signals themselves. Information theory suggests that optimal encoding can be achieved if neural computations are adapted to the statistical properties of the inputs they need to process. This principle is called “efficient coding” hypothesis and it allowed explaining many aspects of sensory processing especially at the level of early computations. Still there is limited evidence regarding the possibility to apply the efficient coding principle to understand properties of cortical sensory processing. Is the extraction of increasingly higher order features by cortical areas accounted for by the efficient coding principle? And, in particular, is this normative principle able to account for the differential tuning for features present in natural images that is found in visual areas?

High order image statistics are defined as complex correlations in the local spatial distribution of light across pixels and are fundamental for the perception of the elements present in a scene. Multiple studies analysed the distribution of these statistics and found that it is not uniform across different natural images or different patches extracted from the same image. It has been shown that, considering groups of up to four pixels, the frequency of some configurations varies more than others. Specifically, it was found that in natural images the variability of multipoint correlations was rank-ordered as:

- 1-point correlations
- 2-point correlations
- 4-point correlations
- 3-point correlations

This result is particularly interesting because the same correlations were found to be uniformly distributed across white noise patches. This means that the ranking reported above is a specific property of natural images. Efficient coding posits that the human visual system should be adapted to the statistical properties of the environment by being more sensitive to the correlations that are more variable across natural scenes and therefore more informative about their identity. By using texture stimuli that allowed isolating specific correlations, it was determined that human sensitivity matches the variability of multipoint correlations found in natural images. This raises the question of whether such specialization of the visual system is hard-wired in the cortical circuitry, as the result of an evolutionary process, or is learned from the statistics of the visual environment during early postnatal development. Testing this hypothesis requires performing invasive experiments in animal

models – ideally rodents, which would allow controlled rearing experiments. However, it is unknown whether other mammalian species share with humans the same sensitivity to multipoint correlations. Trying to tackle this question has been at the core of the two projects I carried out during my PhD:

I. **Project 1**. To understand if rats and humans are similarly adapted to process the multipoint correlations found in natural images, we selected four image statistics (from 1- to 4-point correlations) and trained four groups of rats to discriminate between white noise patterns and binary textures containing variable intensity levels of one of such statistics. An ideal observer model of the task explained well the behavioral performance of individual rats and allowed us to infer their sensitivity to the statistics they were tested on. Results confirmed our hypothesis as we found that rat sensitivity matches human sensitivity and the variability of the statistics in natural images. An article describing the results of this project has been published on Elife in 2021.
II. **Project 2.** Building on the results of Project 1, we decided to investigate the neuronal substrate behind the processing of multipoint correlations in rat visual cortex. We were interested in testing in which area of the rat equivalent of the ventral stream these statistics are encoded and discriminated from noise, and what differences may emerge between texture representations in primary visual cortex (V1) and the highest level area of the hierarchy (area LL). To answer these questions, we performed extracellular recordings on anaesthetized rats in V1 and LL. Animals were presented textures containing the same statistics already investigated in the behavioral study with varying levels of intensity. Current results don't show any clear tuning in neither area for higher-order correlations (3- and 4-point ones), but reveal some key differences between the two areas in terms of their ability to support discrimination of textures defined by 1- and 2-point correlations – with LL displaying a higher degree of invariance (i.e., tolerance to variations across texture exemplars), especially in the case of 2-point correlations.

To provide the necessary background to the next two chapters, the following introductory section will present the theoretical basis that led to the formulation of the efficient coding principle and its expected consequences both in the early and late stages of the visual system. Next, I will analyse some of the most relevant applications of the efficient coding principle to sensory processing to better frame the context around the research presented here. Then, I will present a brief introduction to the use of textures, and in particular maximum entropy textures, as probes for the visual system. Lastly, I will review some relevant literature regarding the rat visual system to display the reasons behind the choice of this animal model and its capabilities.

Efficient Coding Principle

The main goal of sensory systems is to encode their inputs to make them readily available to higher processing areas that use this information to direct the behavior of the organism in the environment (Brinkman et al., 2016). The completion of this task is subjected to strong evolutionary forces that reward encoding of those features that guarantee the best

survivability. In particular, there are 3 major factors that shape the way a sensory system works: the tasks the organism must perform, the computational constraints caused by metabolic and wiring limitations, and the stimulation the organism receives from the environment (Simoncelli and Olshausen, 2001). These considerations are the basis for the formulation of one of the most prominent theories in system neuroscience, the “efficient coding” hypothesis. Rooted in Shannon’s information theory (Shannon, 1948) and originally formulated by Attneave and Barlow, this principle states that sensory systems process environmental inputs by maximizing the information they carry about the world while minimizing the encoding costs by reducing redundancy (Attneave, 1954; Barlow, 1961; Barlow, 2001). This can be achieved only if neural codes are adapted to the statistical properties of the signals they represent, as they have to extract specific relevant features contained in an otherwise highly redundant input. Focusing on the visual system, research has shown that the information received from the outside world can be quantified to be around 25MB/s (Zhaoping, 2019), a stream of data significantly higher than what the limited resources of the visual system allow to process – e.g., attentional mechanisms process sensory data at 5B/s at most (Sziklai, 1956; Sziklai, 1957; Van Essen, 1991). Still, most of the information that enters our visual system is redundant and it doesn’t provide any useful aid for directing behavior. Redundancy, in the framework of information theory, can be measured as the amount of total correlation between multiple variables (in case of two variables this corresponds to the mutual information between them): the amount of information one can obtain on a variable through the observation of another is a direct esteem of how redundant the two variables are. In the hypothesis that the visual system efficiently encodes incoming signals, faithfully representing a stimulus that is highly redundant would be a waste of computational resources. Therefore, the brain has to compress the input data by discarding as much redundant information as possible while preserving relevant features.

Crucial to Barlow hypothesis was the idea that, in order to be efficient, a sensory neural code had to match the statistical properties of the signals coming from the environment. Therefore, it is important to understand what statistics can be measured from natural images. We define as “image statistic” any mathematical quantity that describes one or more statistical properties of an image. Describing images in terms of the value of some of their statistical properties is useful as it provides a simplification of the complexity of natural scenes, giving us a metric to quantify the variability across different scenes. Image statistics can estimate average global properties of an image by taking into account the value of each “pixel” by itself, or they can put in relation properties of multiple neighbouring pixels; the number of locations considered at the same time defines the order of a statistic. The mean luminance value of an image represents a clear example of first order statistic as it requires measuring the intensity of every single pixel; spatial correlations, on the other hand, require computing statistical relationships between two or more pixels at the same time and for this reason are considered higher order statistics. These higher order statistics play an important role in the perception of most of the relevant features that are present in natural images, such as edges or contours (see Figure 1). Under the efficient coding hypothesis, the visual system should be strongly tuned to them.

Until now I only examined efficient coding in the context of pure redundancy reduction without considering another important element that influences how sensory systems operate under the efficient coding hypothesis: noise. Noise makes the relevant features of a signal

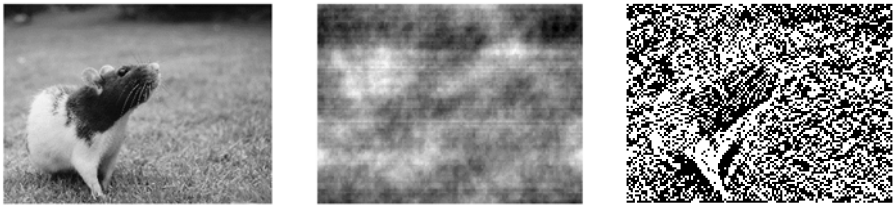


Figure 1. Image statistics in natural scenes. A natural image (left) contains both first order and higher order statistics. To obtain the middle image shown in this example, I first computed the Fourier transform of the natural image on the left, then I destroyed all high level properties by scrambling the phases of the Fourier components, while maintaining the first order statistics intact by keeping constant the Fourier power spectrum. On the right, I instead performed whitening + binarization on the natural image: this transformation removes all low level properties of the image maintaining multipoint correlations intact. It is clear from these examples that removing high order features has a way more destructive impact on shape perception than removing low order ones. This is because high order statistics are fundamental for the perception of features such as edges and contours.

harder to detect and has to be taken into account by neural coding both at the input (noise that is already present in the signal) and output level (noise that is injected in the output signal due to the neural computations). Understanding the interplay between input and output noise is fundamental to comprehend why efficient coding predicts completely different behaviors at different stages of visual processing. In a regime where output noise dominates over input noise, if neuronal responses are correlated and so is the noise they inject in the signal, there is going to be a build-up of the noise level. A transformation that decorrelates the responses will allow to average out the noise. In the condition where input noise is the main constraint, neurons receive an already noisy representation, and the best way to encode it is to allocate more resources to the most variable features as this strategy allows the most reliable discrimination. In the visual periphery where output noise caused, as an example, by the limited band of the optic nerve is greater than noise in the signal coming from the photoreceptors of the retina, the efficient coding framework implies that the most efficient way to encode a stimulus is to minimize redundancy by performing a decorrelation of the image through the whitening process (Srinivasan, 1982). By contrast, in the cortical stages of visual processing input noise becomes the main limiting factor. Here the efficient coding hypothesis predicts that the best way to encode information is to invest more resources in the most variable features, as those are the ones that allow to better discriminate the elements of a noisy scene (Hermundstad, 2014). We therefore expect a direct correlation between the variability of some image statistics across natural scenes and the salience of this statistics (Field, 1987).

In the next section we will examine some of the most relevant application of efficient coding principle to the study of the early visual system to better understand what properties of vision can be explained using the theoretical framework we introduced here.

Efficient Coding in the visual periphery and early cortical stages

Efficient coding has been shown to explain many processes in the early visual system. Srinivasan et al. (1982) measured the number of correlated pixels in natural scenes and then predicted the degree of inhibition that would be required by monopolar cells of the fly

to efficiently remove these correlations. Neighbouring pixels in a natural image are highly correlated, which means that any given value can be reliably predicted by knowing its surrounding values. Therefore, if we want to minimize redundancy in the output the value of each pixel can be exchanged with the difference between this central pixel and its prediction from the linear sum of the neighbouring pixels' values (centre-surround receptive fields). This manipulation decorrelates, i.e. whitens, the image, as I mentioned in the previous section. Authors found a good correspondence between the inhibition level necessary to whiten input images and the one measured in monopolar cells, providing one of the first demonstration of image decorrelation in the sensory periphery. Building on this result, it was later found that retinal ganglion cells perform an analogous operation (Atick & Redlich, 1991, 1992; Balboa & Grzywacz, 2000; Pitkow, 2012). Rao et al. (1999) showed using a computational model that these extra classical surround effects, that make neurons behave as residual error detectors, may be due to the presence of cortico-cortical feedbacks produced by the efficient hierarchical strategy employed by the visual system. To summarize, efficient coding theory assumptions at the periphery predict that if variability of a certain feature of the input is high, the gain of neurons will be low and vice versa. This means that input signal will be 'whitened' by making variance equal in different neuronal populations coding for that input.

Regarding single neurons activity in the periphery, under a constraint on the response intensity, we expect the distribution of a unit responses to match the probability distributions of specific statistics present in natural scenes, with higher sensitivity for the most common outcomes. In general, we can identify three main constraints that shape neuronal responses: a maximum firing rate limit, an average firing rate limit due to metabolic costs, and a boundary on the sparseness of the response (Baddeley, 1997). For instance, the intensity-response function of the monopolar cells of the fly, with a firing rate limited by their reversal potentials, is matched to the intensity distribution of local contrast in natural scenes as shown in Figure 2A (Laughlin, 1981). So, neuronal responses code inputs with an intensity that is proportional to the inputs' expected frequency in order to achieve a flat response distribution (Figure 2B).

We have seen how redundancy reduction and, in general, the need to effectively encode information about the environment drive neurons responses in the retina and early visual system in a way that is in line with information theory and, therefore, with the efficient coding principle. Now we are interested in understanding if it's possible to describe the properties of signal processing in the cortical stages of the visual system using this normative assumption. Li & Atick (1994) provided one of the first demonstrations of the ability of efficient coding hypothesis to explain neurons responses in the visual cortex. They designed a computational model of binocular processing in V1 based on efficient coding. The model was built on the assumption that the most efficient strategy depends on the balance among computational costs, quantified as the weights of the connections between retinal and V1 neurons, the maximization of the removal of signal correlation, i.e. redundancy, and the necessity to keep output noise low. The resulting model was able to correctly predict: the presence of monocular, binocular and disparity selective units; the different tuning properties among these cells types; and correlations between ocular dominance and neurons characteristics.

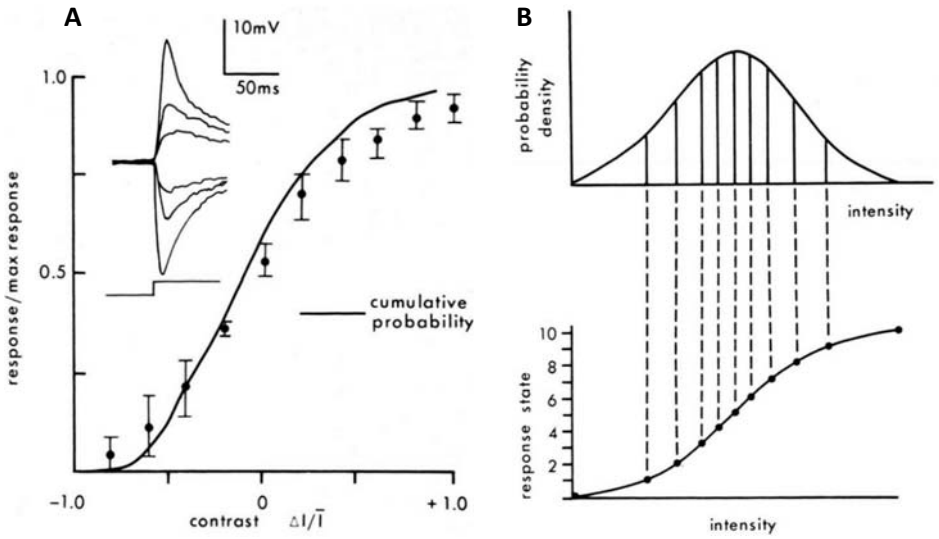


Figure 2. Comparison between the most efficient encoding strategy and monopolar cells contrast response. (**A**) Contrast response of fly's monopolar cells compared to the cumulative probability function for contrast in natural images. (**B**) The most efficient encoding strategy for maximising the information carried by neuron's response (bottom) is to ensures that the interval between each response level encompasses an equal area under the intensity distribution of the stimulus (top), so that each state is used with equal frequency. Adapted from (Laughlin, 1981).

We have seen that, according to the efficient coding principle, neuronal populations have to encode natural scenes in the most compact way possible while maximizing the relevant information about their content. This consideration lead to the discovery of one of the most important strategies employed by neuronal populations: sparse coding (Olshausen & Field, 1996, 1997, 2004; Simoncelli, 2003). Sparse codes are characterized by the strong activation of only a small subset of units in response to any given stimulus; they provide a powerful compromise between "dense codes", where each stimulus is represented by a combination of the activations of all neurons, and "local codes", where each stimulus causes the activation of only one specific unit. Sparse codes are more computationally and metabolically efficient than both dense and local codes (Baddeley, 1998; Willmore, 2011). The sparse coding hypothesis predicts multiple aspects of visual processing both of static and dynamic scenes such as: the maximization of sparseness by simple cells in V1 when presented with natural scenes (Olshausen, 2004; Vinje, 2000) and the emergence of oriented, Gabor-like spatial receptive fields (Olshausen & Field, 1996), or, in the time domain, the tolerance for transformation of the preferred stimulus observable in computational models of the visual system (Wiskott, 2002).

We know that cortical stages of the visual hierarchy have the goal of extracting features that are increasingly more complex, so we expect these areas to focus on higher order statistics compared to what was observed for the retina and LGN. Testing predictions that arise from the efficient coding principle in relation to the processing of higher order statistics is a complex task, as it requires to isolate the contribution of correlations of multiple levels that are entangled in natural images, which also contain other high-level features such as objects shape. This task requires using a set of stimuli that are able to selectively express only certain subsets of higher order statistics found in the natural environment while remaining easy to manipulate and define from a mathematical point of view. In the next

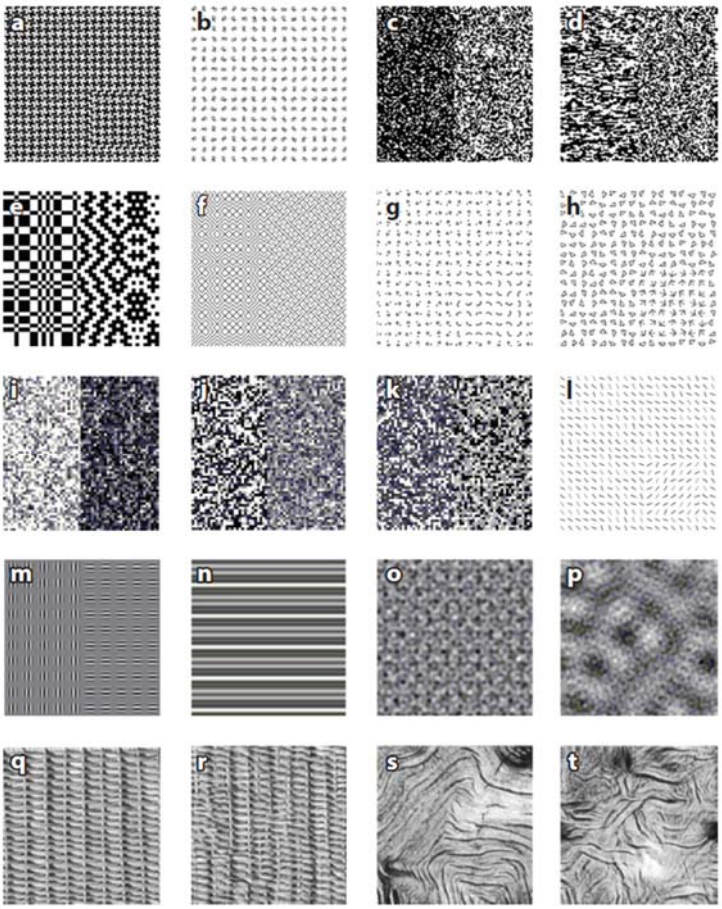


Figure 3. Examples of different textures families. Multiple samples that show the variety of patterns that satisfy the "texture" definition. Among these textures we focus on the most relevant for studies on the processing of high order image statistics: (**c-e**) are maximum entropy textures containing respectively 1-, 2- and 4-point; (**r**) and (**t**) are "naturalistic" artificial textures obtained by sampling from a statistical model, with parameters chosen to match the parameters of a set of visually distinct natural texture images, respectively (**q**) and (**t**). Taken from (Victor, 2017).

section we will look at how artificial textures, and in particular maximum entropy textures, are able to satisfy all these conditions, allowing to test the predictions of the efficient coding principle regarding the saliency of higher order statistics.

Textures as a proxy of natural scenes

Establishing a direct, quantitative, connection between specific statistical properties of natural images and the neural activity is a complex task. The difficulty arises from the fact that mathematical stimuli such as gratings, as well as purely natural scene, despite having many applications, have also important limitations. Stimuli based on purely mathematical models are well defined as they allow to control each single feature they contain, but they do not take into account the complex relationships present in natural scenes, possibly leading to miss some crucial features of neuronal responses. Naturalistic stimuli, on the other hand, provide ecologically relevant results that can be generalized to the real-world environment, but in natural images is difficult to disentangle and control the relationship between different statistics, making impossible to measure individual sensitivities (Victor &

Conte, 2012). Artificial textures represent a convenient compromise between these two categories. They are a class of images in which local patterns are distributed in a repetitive way in order to highlight a specific set of features (see Figure 3 for a wide collection of textures with different statistical properties). Furthermore, an individual sample of a particular texture family is representative for any other sample that is included in that ensemble (Victor, 2017). These features mimic the properties of natural textures which intuitively can be seen as repeated local patterns that, as it is shown in Figure 3q and Figure 3t, provide useful information about a surface orientation, shape and material (Portilla & Simoncelli, 2000). The defining property of a single element being representative of the entire ensemble is also reflected in how the human visual system processes textures, as there is evidence that textures are not represented as individual samples. It was suggested that humans use an internal typified representation of a textures group, that is speculated to be exploited as a Bayesian prior then combined with the perceptual data (Victor & Conte 2004, 2006; Victor, 2017).

Mathematically artificial textures are characterized by two main properties that formalize the idea that each texture sample has to typify for the entire ensemble from which it is collected. The first property is ergodicity, which means that any image statistic, when computed within a sample (i.e., by averaging an estimate across multiple locations within the same image) should return the same result as the one obtained computing that statistic at one specific location for all members of the ensemble and averaging the results. The second feature that defines artificial textures, and is a direct consequence of ergodicity, is spatial homogeneity: averages computed at specific locations across the ensemble of images should all match the sample-wide averages (due to ergodicity) and therefore yield all the same result. An effect of this homogeneity is that textures must be infinite in spatial extent or at least be able to scale infinitely. So, sampling different images from a texture ensemble can be seen as equivalent to selecting discrete regions inside an indefinitely extended “ideal” texture (Victor, 2017).

Despite having such well-defined mathematical properties, developing textures ensembles that are effective probes to understand the characteristics of the visual system is a complex task. The main issue arises from the necessity of controlling for the interdependencies that arise from the parameters that define the textures. Early approaches tried to overcome this problem by restricting the range of spatial correlations present in the textures – e.g. by scrambling pixels. This procedure allowed to build only textures that contained spatial correlations on small spatial scales, as shown in Figure 3i-k (Chubb, 2004; Chubb & Nam, 2000). Other methods based on computer graphic solutions or on working in the frequency domain of the image, lead to textures that contained longer range correlation which intensity could be quantified but not defined, or tuned, before synthesis due to images necessarily having random phases (Xu, 2000; Victor, 2017). One prominent approach to the creation of textures containing fine-tuned values of specific high order statistic is the one proposed by Portilla and Simoncelli (2000) and refined in later studies (Freeman, 2013; Ziemba, 2016). Their generative algorithm is able to produce synthetic textures close to naturalistic ones (Figure 4A) through an iterative process. The procedure starts with a Gaussian noise image that is then processed through a series of filters, modelled on V1 receptive fields properties (Figure 4B), whose responses and energies are correlated across positions, orientations and scales. This process is then repeated in order to match by gradient descent the statistical properties of the output with the desired ones present in a

natural texture (Figure 4C). Despite being able to generate a wide variety of synthetic or naturalistic textures, the main limitation of this technique is that it requires a starting example to be able to generate a set of images with the desired statistics; it is not possible to create textures starting with an arbitrary set of parameters.

In order to solve many of the limitation of the types of textures we described, many studies decided to use an approach based on statistical mechanics by creating artificial textures where one or few correlations are explicitly fixed before generation and the values of all the others are instead implicitly established by the condition that the final textures ensemble has to have maximum entropy. In this context we define entropy as the quantity of "randomness" in the image, therefore, by maximizing it, we impose that the texture is as similar as possible to white noise except for the correlations we fix a priori. The entropy maximization condition has the advantage that it guarantees there aren't hidden interdependencies between the unspecified statistics. In the context of experiments that study specific features of the visual system, the use of these textures prevents the possibility that subjects use strategies that don't depend on the perceptual mechanisms of interest (Victor & Conte 2012).

The main limitation for maximum entropy textures ensembles is that, despite the fact that one texture for any set of self-consistent statistics always exists, it can be operationally

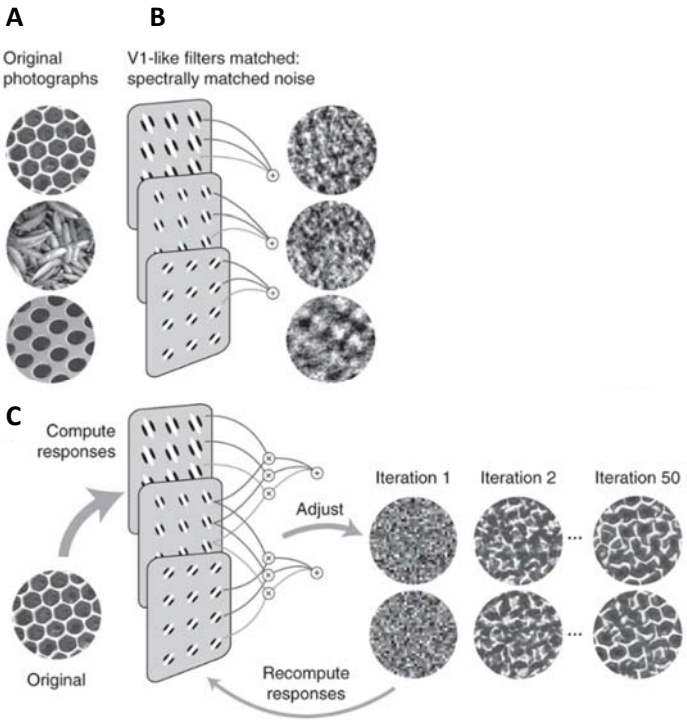


Figure 4. Creation of naturalistic textures and spectrally matched noise from texture photographs. (**A**) Original photographs of natural textures. (**B**) Spectrally matched noise is created by analysing the original image with linear and energy filters that mimic the properties of V1 simple and complex cells. (**C**) Synthesis of naturalistic textures images starts with Gaussian noise. Then, correlations are computed by taking products of linear and energy filter responses across different orientations, spatial frequencies and positions. The result is iteratively adjusted using gradient descent until the properties of the generated images match those of the original texture photograph. Adapted from (Freeman, 2013).

challenging constructing it. Iterative algorithms permit, in many cases, to increase entropy to its limit value by exchanging positions of blocks inside the texture (Zhu, 1998), but they are highly computationally intensive. An efficient alternative to determine the maximum entropy ensembles for binary textures containing a specific subset of multipoint correlations has been developed by Victor & Conte (2012). From this point forward we will refer to these textures when mentioning “maximum entropy textures”. Binary maximum entropy textures for multipoint correlations involving up to 4 pixels can be constructed by randomly initializing a small subset of pixels (by setting them to black or white), and then using a Markov process that depends only on the “parity count” of the number of pixels contained in the glider, which is the basic building-block from which a texture is originated (Victor & Conte, 2012). The concept of glider, a specific window that acts as the building block for the entire pattern, was originally introduced to build isodipole textures –i.e. textures which are not maximum entropy but possess an equal value for 1st and 2nd order statistics (Gilbert, 1980). To clarify the concept of parity count, in maximum entropy textures the specified correlation can take a value between −1 and 1. A value of 1 means that all the possible glider placements across the texture contain an even number of white pixels, a value of 0 means that the texture is pure white noise, while a value of -1 means that all glider placements contain an odd number of white pixels (Victor & Conte, 2012). Maximum entropy textures can be classified based on the amount of points of the glider: γ (gamma) refers to 1-point correlations, β (beta) to 2-point correlations, θ (theta) to 3-point correlations, and α (alpha) is a 4-point correlations. For beta and theta statistics there are four different gliders with different pixels’ arrangements, while for gamma and alpha statistics there is one. This gives rise to ten possible glider types that corresponds to as many image statistics. Figure 5 shows samples of maximum entropy textures for all the four statistics, each with an intensity varying from −1 to 1. Until now we have only considered the case where the intensity of only one statistic is fixed, still it is possible to define multiple of this ten image statistics values when building

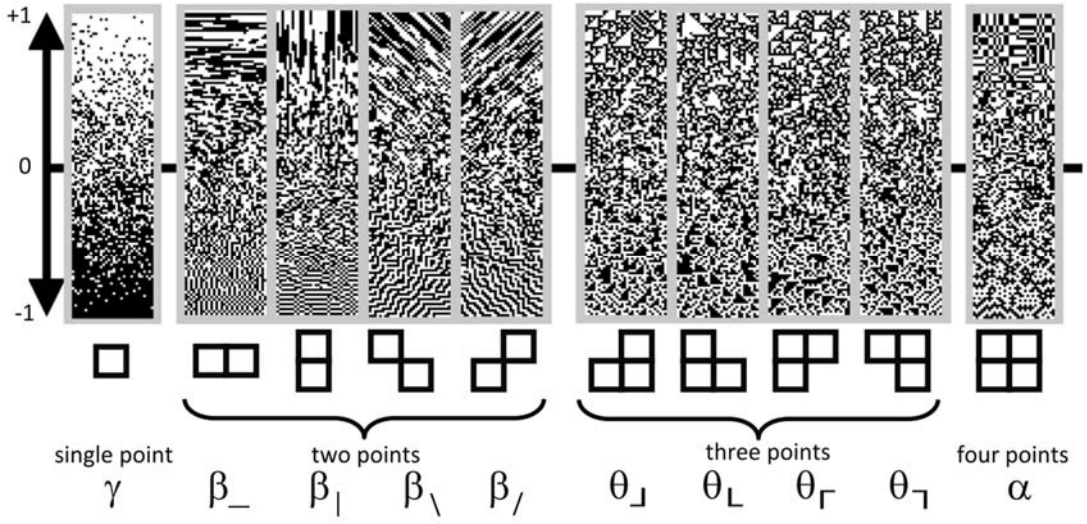


Figure 5. Maximum Entropy Textures for multipoint correlations up to 4 points. Maximum entropy textures are binary textures where one correlation is established and the values of all the others are instead implicitly determined by the condition that the final textures ensemble has to have maximum entropy. Here you can see how textures changes along the range of intensity of each correlation, from -1 to 1. Below each gamut we have the glider used to build textures expressing that specific correlation. Adapted from (Victor, 2015)

a texture ensemble. However, when two or more multipoint correlations are specified at the same time, it is often not possible to use the simple iterative Markov process as the recursion rule does not satisfy the required algebraic constraints that guarantee maximum entropy (Champagnat, 1998; Gilbert, 1980; Pickard, 1980). Still, for combinations of two statistics it's possible to apply a recursive process to reach maximum entropy (Victor & Conte, 2012).

As we already said, maximum entropy textures are a useful tool to study how high order image statistics are encoded in cortical visual areas as they do not contain any undetectable correlations between statistics. However, they lack in "naturalness" due to their major limit: they only allow the exploration of small spatial scales, while in natural images correlations on larger spatial scales may be relevant – e.g. correlations between groups of non-adjacent pixels or inside bigger gliders (Victor, 2017).

In the next section we will examine how naturalistic and, in particular, maximum entropy textures have contributed to shed light on how the visual system processes higher order statistics and the possible connection with the efficient coding hypothesis.

Sensitivity to high order image statistics

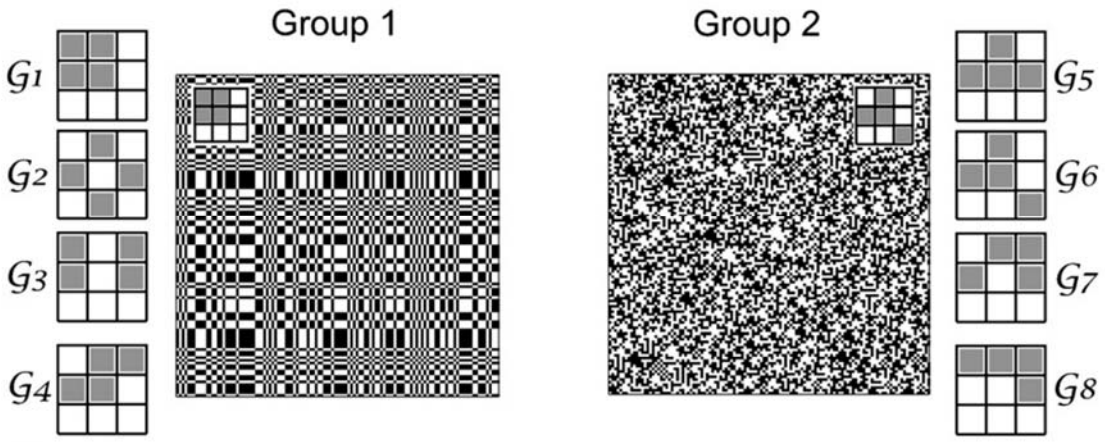
A fundamental claim of the efficient coding hypothesis is that human sensitivity to image statistics follows the variability of those statistics in natural environments. Indeed, in the previous sections, we have seen how many aspects of early visual processing are shaped by the necessity of exploiting the statistical features present in natural images. To extend this analysis to the cortical stages of visual processing vision scientists have relied upon artificial textures, a type of stimuli that allows a fine tuning of their statistical structure. Through the use of texture stimuli, researchers were able to provide evidence for the relevance of efficient coding as a normative principle at stages of the visual hierarchy where nonlinear feature extraction occurs.

Tkačik et al. (2010) analysed the image statistics present in natural grey-scale scenes. First, they built histograms of local pixel intensities by computing the distribution of all pixels at certain distances from a randomly selected one. This analysis showed that nearby pixels have similar intensities for small distances – i.e., the histogram was peaked at the intensity of the central pixel, while for bigger distances pixels tend towards statistical independence. Then, the authors performed principal component analysis (PCA) on the ensemble of histograms to characterize the variations between these distributions of pixels' intensities. They found that approximately 90% of the variance was explained by three PCA components: mean, variance and fraction of dark pixels. They demonstrated that a linear combination of these components approximate the blackshot mechanism, a perceptual mechanism that senses fine variations in the concentration of dark pixels irrespective of the overall luminance of the background (Chubb, 2004). Researchers then repeated the same PCA analysis on synthetic images that matched natural images power spectrum but didn't contain any higher order image statistics. For these images the principal components representing mean and variance explained alone more than 90% of the total variance across local intensity histograms. Furthermore, for synthetic images there was no component associated to the fraction of dark pixels. This result suggested that high order correlations are fundamental to make the blackshot mechanism a salient image statistic. In order to understand which high order correlations contribute the most to blackshot being salient

authors used gliders (see Figure 6A), analogue to the ones used to build isodipole and maximum entropy textures, to parse the distribution of 4th order correlations in patches of natural images after whitening and binarizing them (See Figure 6B). Previous behavioral studies established, using isodipole textures built with the same gliders, that in humans only some correlations were perceptually salient (Victor & Conte, 1991). Tkačik and colleagues found that 4th order correlations that carried the highest information about differences between patches were the same that were salient for the human visual system. Therefore, 4th order correlations can be divided into two families of pixel configurations based on the amount of information they carry: informative and uninformative. Informative configurations are correlations which are not inferable from lower-level statistics and are salient; in contrast uninformative configurations are predictable from lower-level features and therefore not salient as they are less informative according to the efficient coding principle.

Building on these results Victor and Conte (2012) investigated what made specific subtypes of 4th order statistics both informative and visually salient. All the possible arrangements of black and white pixels in a 2x2 grid can be summarized by 10 unique and independent configurations, each associated to a specific glider, that can be seen as coordinate of a ten dimensional image statistics space (see Figure 5 for all the 10 gliders). The goal of the study was to measure human sensitivity independently for each of these configurations – i.e. multipoint correlation from 1-point up to 4-point. To this end, authors built maximum entropy textures, as we described in the previous section, that contained a specified intensity of one correlation while having all the others implicitly fixed to an intensity

A



B

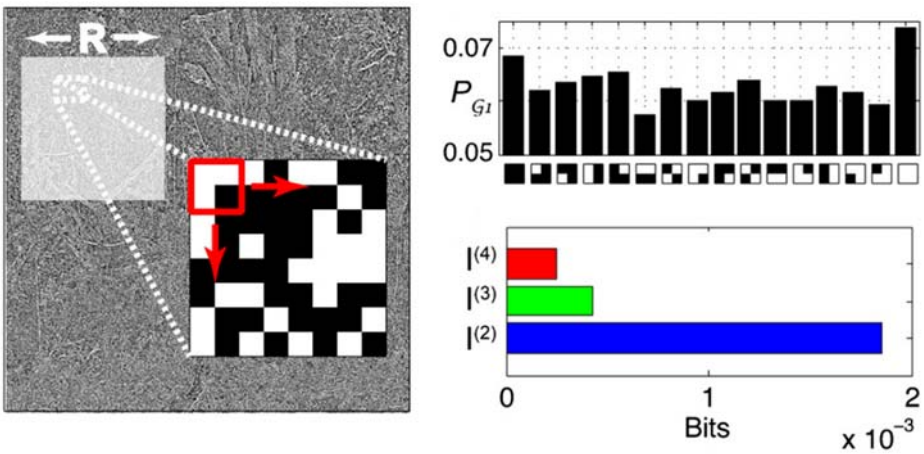


Figure 6. Gliders for 4th order statistic and examples of isodipole textures generate by using them. (**A**) Textures generated from gliders G1-G4 were found to be perceptually salient for humans compared to white noise, while textures from gliders G5-G8 were not. (**B**) Tkačik and colleagues used each gider to scan patches of size R (64 pixels in the example) taken from multiple natural images (left). They then measured the probability to appear of each possible configuration of black and white pixels contained into the glider (right-top). Finally, they measured the information content corresponding to correlations of different order (right bottom). Adapted from (Tkačik, 2010).

that guaranteed that the final texture was as similar as possible to white noise. Subjects had to perform a 4-alternatives forced choice task: they had to identify in one of four possible positions a rectangular target texture containing the correlation inside an otherwise white noise background or, vice versa, they had to discriminate an unstructured target texture from a structured background (See Figure 7A for the results on each statistic). Looking at the sensitivities for each multipoint correlation they found that these were rank ordered in a specific way that was consistent across subjects and constant for both the positive and negative axis of each statistic (Figure 7B): humans are more sensitive to 1-point correlations (gamma), followed by 2-point correlations (beta), 4-point correlations (alpha), and finally 3-point correlations (theta). The higher saliency of 4-point compared to 3-point correlations was an important and unexpected result: perception of 3-point correlation should be easier given that it requires to process less pixels simultaneously. Authors interpreted this result as an effect of an efficient adaptation to the statistical properties of the environment: the salience of each of these statistics must be tied to the amount of information they carry about the content of natural scenes.

The work of Hermundstad et al. (2014) added further evidence to this line of research by combining and expanding the findings obtained by Tkačik et al. (2010) and Victor and Conte (2012). Authors, first, analysed the variability across patches of greyscale natural images of nine of the ten statistics, defined by glider of up to 4 points, already explored in the work of Victor and Conte (2012); the only statistic excluded was 1-point correlations (average luminance) due to processing constraints. As in the work by Tkačik et al. (2010), images were pre-processed to remove the contribution of low-level features: they were block-averaged, divided into patches, binarized and then whitened. This pipeline made it impossible to measure the variability of 1-point correlations as previously mentioned. The pre-processed patches were “scanned” using a 2x2 glider to measure the frequency of appearance of each of the 16 possible configurations of black and white pixels. The occurrence probability of every possible coloring of the window was then translated to a 9-dimensional version of the image statistics space introduced by Victor and Conte (2012). In this space each patch corresponds to a vector with its own coordinates, given by the frequency of each multipoint correlation (Figure 8A). Hermundstad and colleagues, then, computed the degree of variation for each of the 9 correlations in the ensemble of patches. This quantity is defined as the standard deviation of the ensemble along one specific

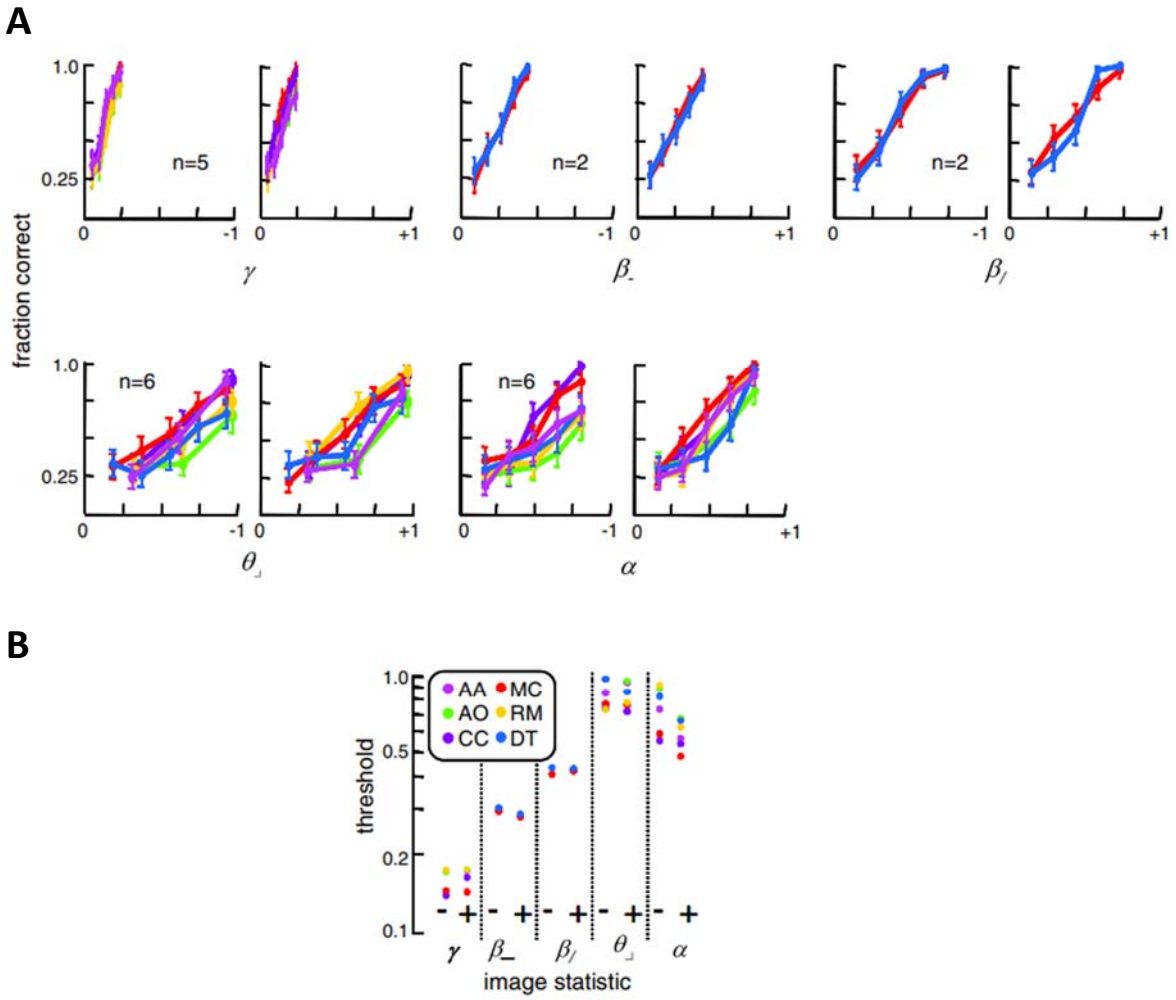
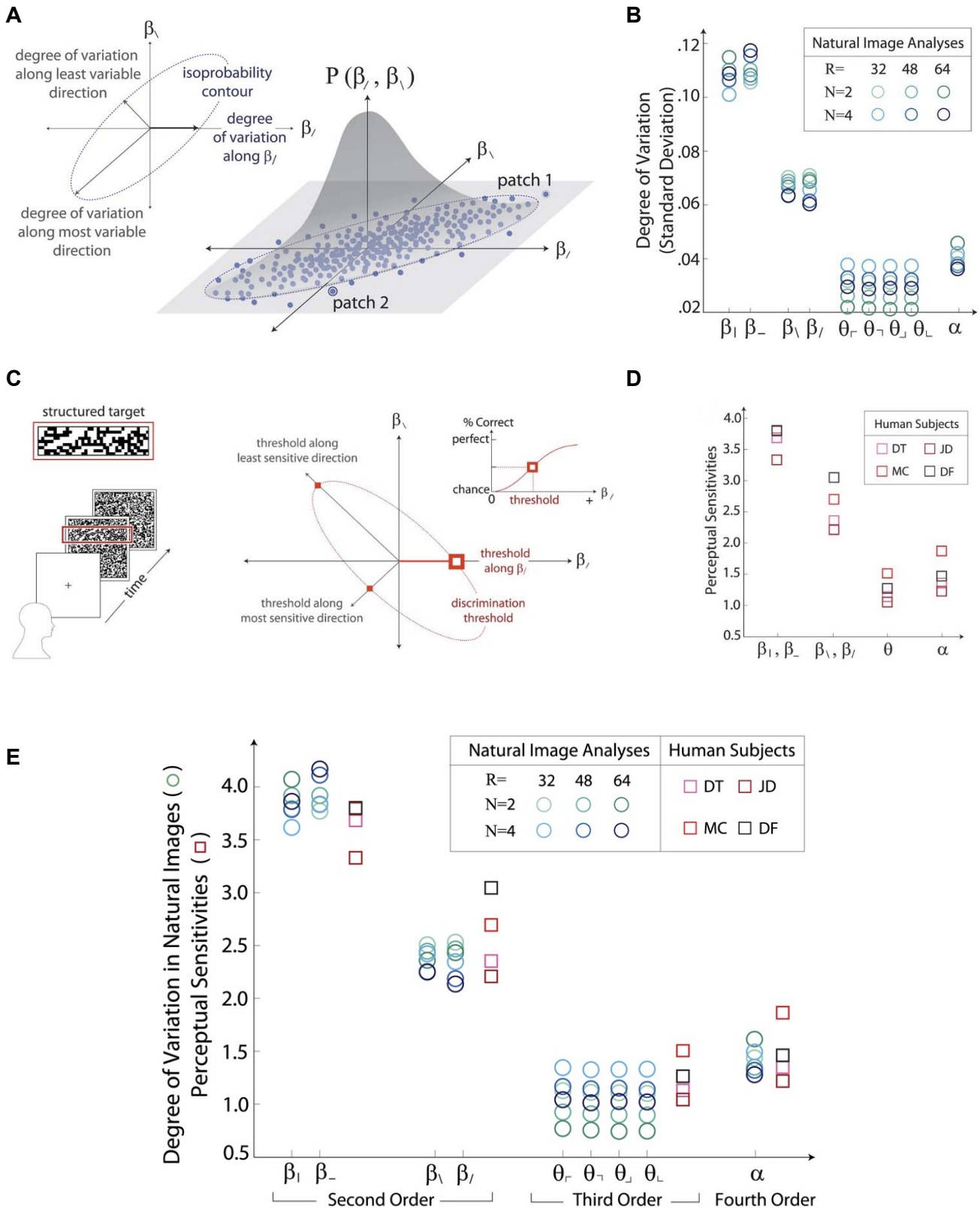


Figure 7. Human sensitivity to multipoint correlations. (**A**) Psychometric functions for 5 multipoint correlations from 1- to 4-point. For each coordinate, psychometric functions are shown for negative intensity levels (left element of each pair) and positive levels (right element of each pair). Chance performance is 0.25; error bars indicate 95% confidence limits based on binomial statistics. Slope of the psychometric. (**B**) Threshold values for sensitivity were estimated, for each subject, by determining the value of the statistic that yielded a performance half-way between chance and perfect. All subjects had the same rank-order of thresholds: 1-point, 2-point cardinal, 2-point diagonal, 3-point, 4-point correlations. From Victor & Conte (2012).

direction of the 9-dimensional space and reflects how much the frequency of the corresponding configuration varies across patches of natural images. The results of this analysis showed that multipoint correlations in natural images are rank ordered in a specific way, with 2-point correlations being more variable than 4-point correlations, and 4-point correlation being more variables than 3-point ones (Figure 8B). Authors also confirmed that this is a specific property of natural images: the same analysis performed on patches of white noise patches lead to a flat distribution of the variabilities with significantly smaller coefficients of variation. Human subjects' sensitivity to the nine correlations in exam was then tested in a way analogue to Victor and Conte's work (Figure 8C), obtaining matching results (Figure 8D). Therefore, Hermundstad and colleagues were able to confirm that human sensitivity to multipoint correlations and the variability of the same correlations in



natural images were rank ordered in the same exact way, as shown in Figure 8E. To show that these results are in line with the efficient coding hypothesis, they developed a theoretical model of efficient coding in the visual system that takes into account the presence of noise. Through this model, the researchers showed that the visual cortex, contrary to the sensory periphery, operates in a regime in which the noisy nature of synaptic inputs creates a

Figure 8. Comparison between multipoint correlations variability in natural images and human sensitivity. (**A**) Hermundstad et al. computed the multidimensional distribution of 9 multipoint correlations in patches of size R obtained from greyscale natural images that were block-averaged, whitened, and binarized. Each patch corresponds to a vector of coordinates in the 9-dimensional space of image statistics (as an example, we have a 2 dimensional projection). (**B**) Variability of multipoint correlations in natural images is then defined as the degree of variation (standard deviation) along different directions of this space (R is the patch size; N is the block-average factor used in image pre-processing). Statistics variability follows the ranking: 2-point, 4-point, 3-point correlations. (**C**) Human sensitivity to multipoint correlations was measured through a segmentation task using maximum entropy textures. Discriminability of each correlation from noise was established as the threshold value of the coordinate required to achieve a performance halfway between chance and perfect. (**D**) The sensitivities were rank ordered as: 2-point, 4-point, 3-point correlations. (**E**) Comparison between statistics scaled degree of variation in natural images and human sensitivities. Adapted from (Hermundstad, 2014).

limitation in the reliability of the incoming signals. In this regime the information gain increases along with signal variability; this means that the degree of variation of the spatial correlations in natural images should be matched by the sensitivity to those same statistics. The results the authors found validate this prediction, and suggest that human visual cortex is adapted to process the statistical structure of natural scenes as expected in the efficient coding framework.

A recent study by Tesileanu et al. (2020) extended the results that emerged in previous works (Hermundstad, 2014; Victor & Conte, 2012) to the domain of grayscale textures. Subjects here were asked to find a target structured texture within a noise background, with both containing three grayscale levels. They used, as the noise background, independent identically-distributed (IID) textures with three grey levels (the analogue of white noise in the non-binary case); the target, instead, was a maximum entropy texture containing varying intensities of one or two multipoint correlations. Results showed that subjects' sensitivity to image statistics corresponded to their variance in natural images expanding the findings presented in Hermundstad et al. (2014) to the space of non-binary images. Moreover, the authors found two other interesting properties of human sensitivity to multipoint correlations that further support the efficient coding hypothesis. First, perceptual thresholds do not change if the texture is flipped symmetrically, even if the texture is not visually symmetrical. This suggests that this invariance arises from the statistics distribution itself rather than from the visual resemblance of symmetry. Authors hypothesized that this feature reflects the visual system adaptation to the symmetries in the real world. Second, results suggested the presence of bright-dark asymmetries in perceptual thresholds for statistic detection, namely, a threshold, in a setting in which there are more dark pixels than light ones, is lower than in the opposite case – i.e. the blackshot mechanism we examined previously (Tkačik, 2010; Chubb, 2004).

We have seen how there is wide evidence showing that human sensitivity to high order image statistics is tuned to their variability across natural scenes. However, what remains to be understood are the neural correlates behind the processing of these statistics in the visual cortex. Yu et al. (2015), building on the results of Hermundstad (2014) and Tkačik (2010), investigated the differential encoding, compared to noise, of multiple 3rd and 4th order statistics in the primary visual cortex (V1) and V2 area of macaques through single-unit recordings. As shown in Figure 9A, the stimuli used were white noise textures and

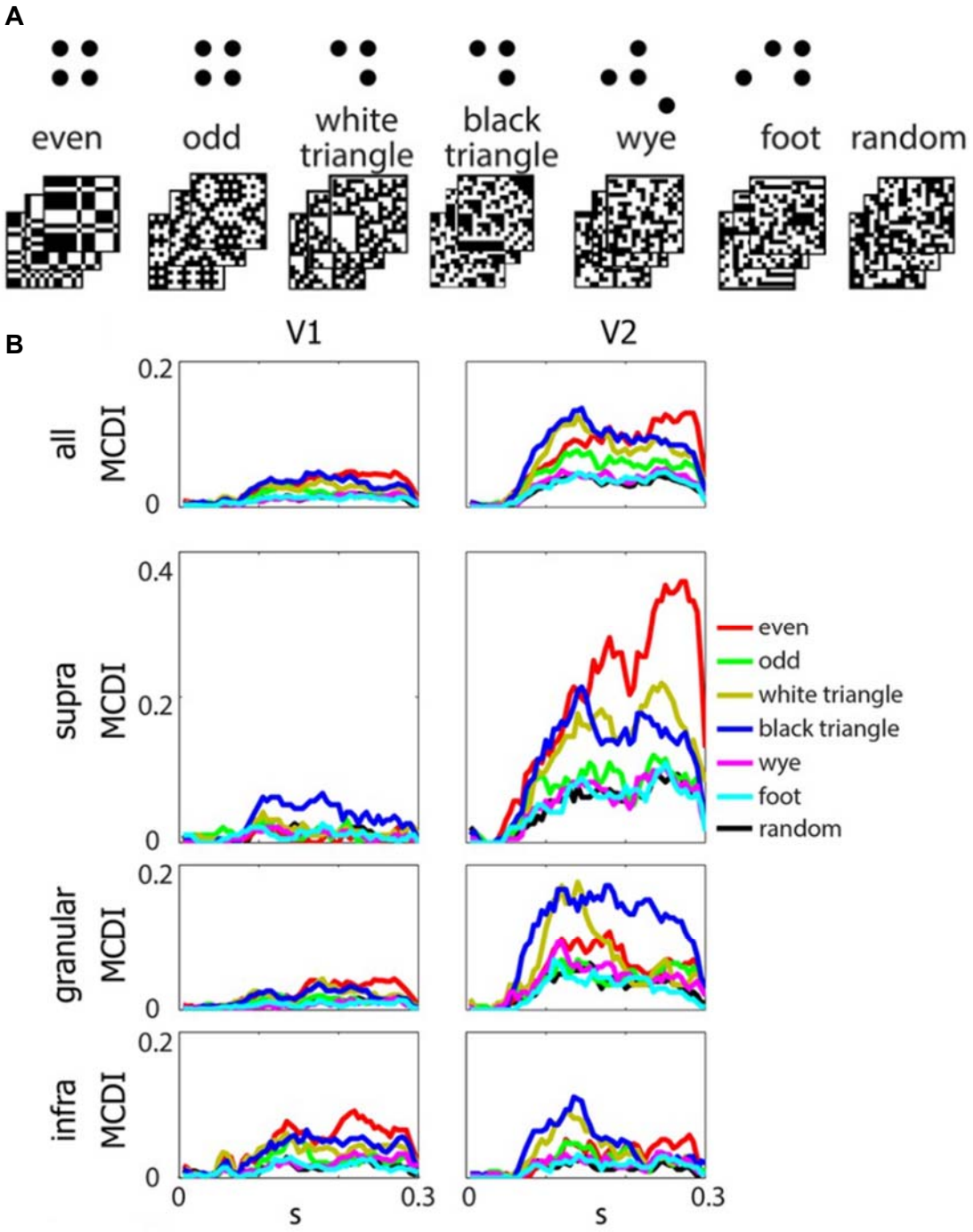


Figure 9. Discriminability of 3rd and 4th order statistics in monkey V1 and V2. (**A**) examples of the stimulus sets used to isolate different statistics with the corresponding glider on top. Six sets consist of 1024 16 × 16 binary isodipole textures containing one specific correlation while the last set consists of 1024 16 × 16 white noise images. (**B**) Mean values of stimulus-specific "multipoint correlations discrimination index" (a measure of the difference between the firing rate for a stimulus and for white noise) show that there is a higher discriminability of stimuli from noise in V2 compared to V1. Furthermore, the highest values of MCDI were found for the same stimuli that are informative in natural images (Tkačik, 2010). Adapted from (Yu, 2015).

isodipole textures containing both informative and uninformative 3rd and 4th order correlations (Tkačik, 2010; Victor & Conte, 1991). Authors found that, in V2, the firing pattern

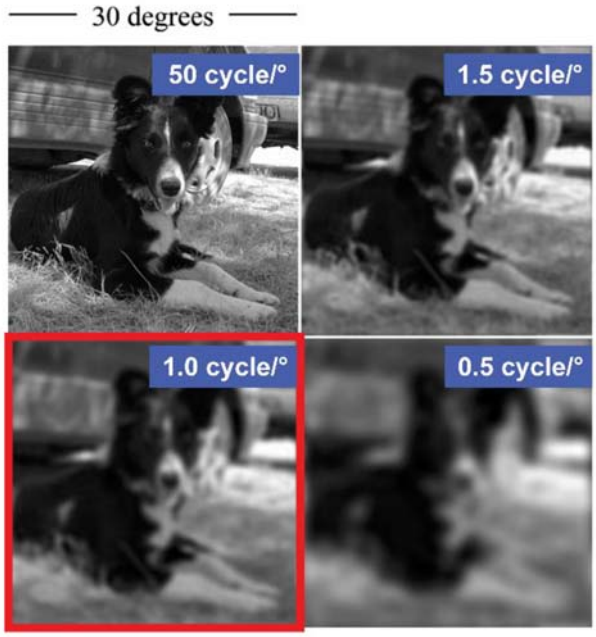


Figure 10. Comparison between human's visual acuity and different rat's strains acuity. An original image (top left), rendered at a resolution corresponding to 50 c/d to represent human's visual acuity, is Gaussianly blurred to give a visual impression of how different rat's strains perceive a visual stimulus (30° wide). Acuity values for Fisher Norway (top-right), Long Evans (bottom-left, red square) and Wistar (bottom-right) rats were obtained by testing the animals in a visual water maze task. Adapted from (Prusky, 2002)

of cortical neurons was significantly different for textures containing informative correlations and noise textures; in V1, instead, this difference was not present in any cortical layer (Figure 9B). This result suggests that the sensitivity for the salient 3rd and 4th order image statistics emerges in V2 and authors speculate that this may be the reason behind many of the differences between the neural characteristics of the two areas.

Other studies found evidence for the processing of higher order image statistics starting from V2. All these studies used as stimuli the naturalistic textures introduced by Portilla and Simoncelli (2000). As we have seen in the previous section, these stimuli differ from maximum entropy or isodipole textures because the higher order image statistics they contain are not isolated but combined in order to obtain artificial textures as close as possible to specific natural ones. Freeman et al. (2013), using fMRI, compared the neuronal responses in areas V1 and V2 of human subjects to naturalistic textures and control textures lacking higher order structure. Neurons in V2 responded more to naturalistic textures compared to unstructured ones, while the same didn't happen for V1 neurons. Furthermore, authors measured human sensitivity to naturalistic textures through a 3-alterantives forced choice "oddity" task and they found that human observers' ability to perceive these stimuli correlated with the strength of fMRI responses in V2 but not in V1. Ziemba et al. (2016) investigated how tolerance for transformations in the arrangement of high order statistics is built along the ventral stream of macaque monkeys. They generated ensembles of naturalistic textures, analogue to those used by Freeman et al. (2013), that contained a set of parameters chosen to match specific natural texture images. Each ensemble contained textures that had the same statistical features but they differed in appearance due to the

location of the features themselves inside the texture. Authors recorded single neurons activity from V1 and V2 areas of anesthetized macaque monkeys. They found that neurons in V2 are more tolerant and selective than V1 neurons for higher-order image statistics when asked to discriminate between textures coming from different ensembles, and so containing different statistical properties. Neurons from V1 instead were more selective when "asked" to discriminate between elements of the same ensemble. Okazawa and colleagues (2017) used naturalistic textures to compare neurons' responses from areas V2 and V4 of macaque monkeys to higher-order image statistics. They showed that both areas respond to higher-order statistics, but V4 demonstrated slightly stronger sensitivity to them. This result seems to suggest that there is a progressive build-up of the information regarding high order statistics along the visual hierarchy. Despite this evidence, the processing of high order correlations by the ventral stream could be more complex: Ziemba and colleagues (2019) showed that, when performing population instead of single unit recordings, there was a higher than expected sensitivity to image statistics in monkey's V1. In particular, this sensitivity was found for neurons in the supragranular and infragranular layers of V1 and it only emerged after selective responses for natural textures appeared in V2. These led Ziemba and colleagues to conclude that sensitivity for high order statistics arises in V1 due to feedback projections from V2.

We have seen that there is evidence for the relevance of the efficient coding principle in the cortical stages of visual processing. Yet, there is not a causal demonstration of the adaptation of the visual cortical hierarchy to the statistical properties of the environment. Rodents, and rats in particular, could allow a direct causal test of this adaptation through controlled rearing experiments in an altered visual environment (Matteucci & Zoccolan, 2020). Furthermore, rats could be the ideal animal model to clarify in which areas of the visual cortex specific multipoint correlations are processed. In the next section we will examine why rats could be a convenient model to study properties of visual processing and examine the properties of the areas that are thought to form the rat analogue of the ventral stream.

Rat visual system

Historically non-human primates have been the most common animal model used in vision research. They are close in the phylogenetic tree to humans (Pozzi et al., 2014), and the functional and anatomical structure of primate's visual system closely resembles our own (Orban et al., 2004; Rajalingham et al., 2015). Primates extensive use in vision research is also justified by their ability to perform advanced visual tasks, such as objects recognition and categorization (DiCarlo et al., 2012; DiCarlo & Cox, 2007; N. C. Rust & DiCarlo, 2010). Still, the primate brain, as our own, is extremely complex (Felleman & Van Essen, 1991). Monkeys are relatively big animals that are difficult to handle, making parallel behavioral training of multiple subjects almost impossible. Furthermore, genetic, molecular, and, in general, invasive techniques such as optogenetics or two-photon imaging are impractical in monkeys. These limitations, over the last two decades, led to small mammals, and in particular rodents, to become an increasingly popular animal model for studying the visual system (Huberman & Niell, 2011; Zoccolan, 2015). Among the main factors behind rodents' popularity there was the continuous growth of the array of experimental techniques, as those mentioned above, that allow a fine exploration and manipulation of the visual system.

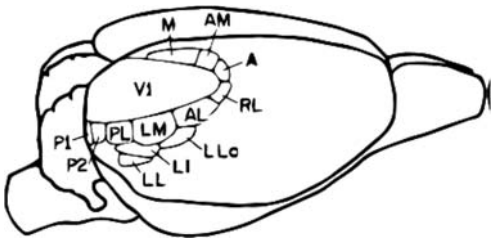


Figure 11. Rat visual areas. Rat visual cortex is formed by 13 areas, with V1 being the widest (15% of neocortex). Adapted from (Sereno & Allman, 1991)

Furthermore, rodents require less housing space and resources, and they are also easier to handle compared to primates. This allows researchers to use large number of subjects in behavioral experiments which make results more statistically robust. Another advantage of rodents is that their brain is relatively small compared to monkeys. This, while making it harder to precisely target specific areas, permits simpler procedures for electrophysiological recordings and allows electrodes to span multiple areas.

For a long time, rodents have been thought not to rely on vision to guide their behavior and this has strongly limited their use as a model for visual research. The main reason behind this notion was that rodents are crepuscular animals with a low spatial resolution, ~50 times lower than primates (Huberman & Niell, 2011). Nonetheless, recent studies have shown that rodents behavior is frequently driven by visual cues. Rodents depend on vision for environment exploration (Høydal, 2019; Reinhold, 2019; Chen, 2013), predation (Hoy, 2016), and even communication (Nakashima, 2015). Looking at perception specifically, rats, in particular, are able to perform complex object discrimination tasks including invariant object recognition – i.e. they are able to correctly recognise visual objects in spite of a wide array of identity-preserving transformations such as scale, position, and pose changes (Djurdjevic, 2019; Alemi-Neissi, 2013; Tafazoli, 2012; Zoccolan, 2009).

Among rodent species, rats were shown to be able to perform the most complex visual tasks as we have mentioned above; for this reason, they are the rodent model of choice for this work. We will now look more in details at some important properties of the rat visual system.

Thanks to their eye position rats possess a wide viewing angle that spans close to 180° degrees, with a small region of binocular overlap amounting to 40° on the horizontal plane and 60° on the vertical plane (Wallace, 2013). Rat visual acuity is relatively low, ranging from ~1.5 c/d (cycles/degree) in pigmented strains (1.0 c/d in Long Evans) to ~0.5 c/d in certain albino strains (Burn, 2008; Prusky, 2002). These values are remarkably lower than human's visual acuity that reaches more than 30 c/d in the fovea. See Figure 10 for a comparison between human and rat's acuity. One of the main reasons for rat's poor visual acuity is that its eyes lack a fovea and the retina has a roughly constant density of photoreceptors (Whishaw & Kolb, 2004). Moreover, rat eyes lack the ability to accommodate – i.e. they are unable to adjust to keep an object in focus on the retina when its distance is changing. Despite this, they display, in high pupil constriction states, a high depth of focus that allows them to maintain a sharp retinal image for small distance changes without the need for accommodation (Hughes, 1977). Rat retina, as expected from their crepuscular nature, is able to function in low light conditions due to the preponderance of rods, which

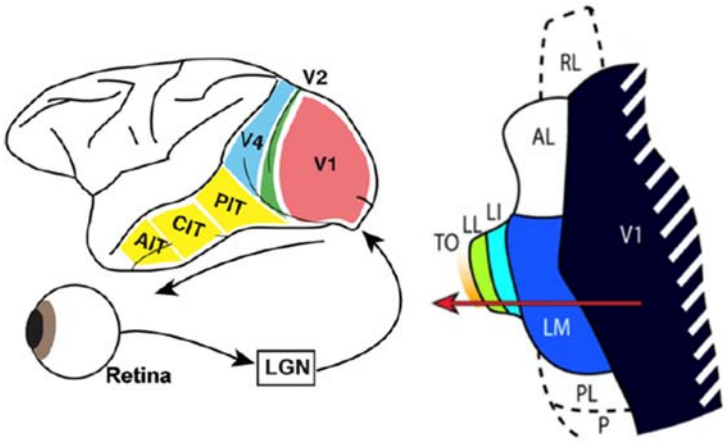


Figure 12. Comparison between monkey's ventral stream and rat's putative one. In non-human primates (left), object recognition has its neural correlates in the so-called ventral visual pathway (Kravitz, 2010; Logothetis & Sheinberg, 1996). Information is processed hierarchically in a feed-forward manner (DiCarlo, 2012). Signals coming from the retina pass through LGN and arrive in the ventral pathway formed by areas: V1, V2, V4 and IT (inferotemporal cortex). These areas are characterized by an increase of transformation tolerance and a tuning for progressively more complex features. Functional studies have showed that in rats (right) a similar type of hierarchical processing is found in the sequence of areas: V1, LM, LI, LL. Adapted from (DiCarlo, 2012) and (Vermaercke, 2015).

account for more than 99% of the photoreceptors. The remaining 1% of receptors is made by two types of cones: one, corresponding to the 93% of the total, has a peak response at 510 nm (blue-green light), while the other type has its peak at about 360 nm (in the UV range) (Sefton et al., 2004). It's interesting to note that, due to the cones distribution, the white light emitted by RGB monitors does not appear as white to rats (Burn, 2008).

Moving from the properties of the retina to visual cortex, rat's visual system, as shown in Figure 11, is formed by 13 areas with V1 being the largest, occupying 15% of the neocortex (Sereno & Allman, 1991). Primary visual cortex in rat has a precise retinotopic organization: the upper visual field (lower retina) is represented caudally, the nasal visual field (temporal retina) is represented laterally, and the vertical meridian is at the lateral border of V1. The lateral part constitutes the binocular field, whereas in its medial part only the monocular field from the contralateral eye is represented; approximately 80% of V1 cells are binocular (Sefton et al., 2004). Neurons in V1 responds to spots, bars, edges and light flashes with an average receptive field size of 15° (Espinoza & Thomas, 1983). They are tuned for orientations and, to a lesser extent, direction (Matteucci et al., 2019; Girman, 1999). As in other species, V1 neurons can be classified in simple and complex cells based on the response to drifting gratings (Matteucci & Zoccolan, 2020; Matteucci et al., 2019; Girman, 1999; Parnavelas, 1981).

In rats V1 is bordered by a large number of high level visual areas (HVAs). These are extrastriate areas receiving input from other visual cortical area (intracortical pathway) and from the lateral posterior nucleus (extrageniculate pathway) which, in turn, receives its input from the high-order visual thalamus (Sereno and Allman, 1991; Dräger, 1974). High level visual areas can be discriminated based on their retinotopic maps (Dräger, 1975; Wagor, 1980), in particular, Espinoza and Thomas (1983) were able to identify a group of six HVA that contains a full representation of the visual field through serial electrode recordings.

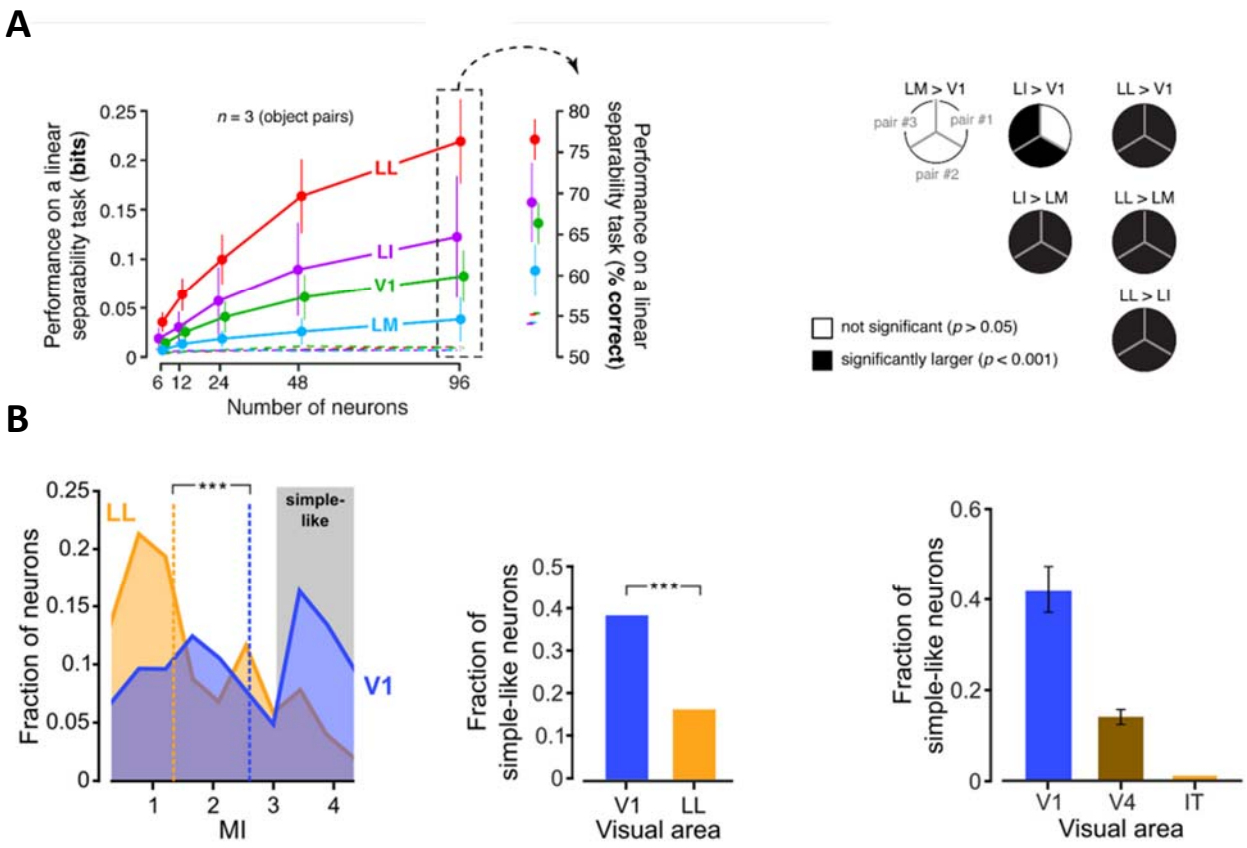


Figure 13. Evidence for a hierarchical processing stream in rats. (**A**) Classification performance of the binary linear decoders in the test for linear separability between objects pairs, as a function of the number of neurons used to build the population response space (left). Performances were computed for three pairs of objects with matched luminance. The statistical significance of each pairwise comparison between areas is reported for individual objects pairs (1-tailed U-test, Holm-Bonferroni corrected). In the pie charts, a black slice indicates that the test was significant. From Tafazoli et al. (2017). (**B**) Distribution of modulation index (MI) for units in V1 and LL (left). Modulation index is standardized measure of the difference between the power of the response at the stimulus frequency and the average value of the power spectrum in units of its SDs. Values of MI>3 imply a strong modulation of the FR at the stimulus frequency (typical of simple-like cells). The fraction of simple-like cells (i.e., units with MI>3) in V1 (blue) compared to LL (orange) shows a significant (p<0.001) difference between the two areas (centre). The drop in simple-like units is analogue to the one found in primate's ventral stream (right). From Matteucci et al. (2019).

Extrastriate areas receives strong projections from V1 in a retinotopically organized way (Montero, 1993), and send feedback connection to V1 and the lateral posterior nucleus (Sanderson, 1991). Moreover, asymmetric projections exist within lateral extrastriate cortex that testify the presence of feedforward and feedback connections (Coogan & Burkhalter, 1993). Lesion studies, additionally, demonstrated that damage to lateral visual areas impairs rats' ability to recognise visual objects (Tees, 1999). Together all these results, combined with the increasing size of receptive fields in lateral areas (Espinoza & Thomas, 1983), suggest that extrastriate areas in rat are organized in a hierarchical way.

Strong evidence for a hierarchical organization of extrastriate areas in a "dorsal like" and "ventral like" pathways, partially analogues to those found in primates, came from

functional studies (see Figure 12 for a comparison between monkey and rat's ventral stream). Multiple works, in particular, contributed to establish the existence of an object processing stream in rats (Tafazoli, 2017; Vinken, 2017; Vermaercke, 2015; Vermaercke, 2014). Tafazoli and colleagues (2017), specifically, were able to demonstrate an increase of tolerance to identity preserving transformations along the hierarchy of areas V1, LM, LI, LL (Figure 13A). The latero-intermediate area (LI) and the latero-lateral area (LL) emerged as the best candidate for rat's analogue to primates V4 and IT (Figure 13A). Using a different experimental design, relying on presentation of drifting gratings rather than visual objects, Matteucci et al (2019) confirmed these results, showing that neuronal tuning from V1 to LL becomes less sensitive to the phase of the gratings (i.e., more translation invariant), thus displaying a behavior increasingly more consistent with the one of complex cells (Figure 13B).

The results discussed provide strong basis for new studies aimed at further testing the similarities between rat's extrastriate areas and primates ventral stream. The third chapter of this thesis will focus on the search for the neural correlates of sensitivity to multipoint correlations comparing the responses of neurons populations in rat's V1 and LL. This will further contribute not only to the understanding behind the processing of high order statistics, but also to the characterization of the properties of the rat putative ventral stream, with a particular focus on the increase in the invariant encoding of visual stimuli along the hierarchy.

References

Alemi-Neissi, A., Rosselli, F. B., & Zoccolan, D. (2013, April). Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition. *Journal of Neuroscience, 33*, 5939–5956. doi:10.1523/jneurosci.3629-12.2013

Atick, J. J., & Redlich, A. N. (1991, January). Predicting ganglion and simple cell receptive field organizations. *International Journal of Neural Systems, 01*, 305–315. doi:10.1142/s0129065791000340

Atick, J. J., & Redlich, A. N. (1992, March). What Does the Retina Know about Natural Scenes? *Neural Computation, 4*, 196–210. doi:10.1162/neco.1992.4.2.196

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183–193. doi:10.1037/h0054663

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. T. (1997, December). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 264*, 1775–1783. doi:10.1098/rspb.1997.0246

Balboa, R. M., & Grzywacz, N. M. (2000, September). Occlusions and their relationship with the distribution of contrasts in natural images. *Vision Research, 40*, 2661–2669. doi:10.1016/s0042-6989(00)00099-7

Barlow, H. (2001, January). Redundancy reduction revisited. *Network: Computation in Neural Systems, 12*, 241–253. doi:10.1080/net.12.3.241.253

Barlow, H. B. (1961, September). Possible Principles Underlying the Transformations of Sensory Messages. In *Sensory Communication* (p. 216–234). doi:10.7551/mitpress/9780262518420.003.0013

Brinkman, B. A., Weber, A. I., Rieke, F., & Shea-Brown, E. (2016, October). How Do Efficient Coding Strategies Depend on Origins of Noise in Neural Circuits? (J. Beck, A cura di) *PLOS Computational Biology, 12*, e1005150. doi:10.1371/journal.pcbi.1005150

Burn, C. C. (2008, July). What is it like to be a rat? Rat sensory perception and its implications for experimental design and rat welfare. *Applied Animal Behaviour Science, 112*, 1–32. doi:10.1016/j.applanim.2008.02.007

Champagnat, F., Idier, J., & Goussard, Y. (1998). Stationary Markov random fields on a finite rectangular lattice. *IEEE Transactions on Information Theory, 44*, 2901–2916. doi:10.1109/18.737521

Chen, G., King, J. A., Burgess, N., & O\textquotesingleKeefe, J. (2012, December). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences, 110*, 378–383. doi:10.1073/pnas.1215834110

Chubb, C., & Nam, J.-H. (2000, June). Variance of high contrast textures is sensed using negative half-wave rectification. *Vision Research, 40*, 1677–1694. doi:10.1016/s0042-6989(00)00007-9

Chubb, C., Landy, M. S., & Econopouly, J. (2004, December). A visual mechanism tuned to black. *Vision Research, 44*, 3223–3232. doi:10.1016/j.visres.2004.07.019

Coogan, T. A., & Burkhalter, A. (1993, September). Hierarchical organization of areas in rat visual cortex. *The Journal of Neuroscience, 13*, 3749–3772. doi:10.1523/jneurosci.13-09-03749.1993

DiCarlo, J. J., & Cox, D. D. (2007, August). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*, 333–341. doi:10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012, February). How Does the Brain Solve Visual Object Recognition? *Neuron, 73*, 415–434. doi:10.1016/j.neuron.2012.01.010

Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H., & Zoccolan, D. (2018, April). Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy. *Current Biology, 28*, 1005–1015.e5. doi:10.1016/j.cub.2018.02.037

Dräger, U. C. (1974, December). Autoradiography of tritiated proline and fucose transported transneuronally from the eye to the visual cortex in pigmented and albino mice. *Brain Research, 82*, 284–292. doi:10.1016/0006-8993(74)90607-6

Dräger, U. C. (1975, April). Receptive fields of single cells and topography in mouse visual cortex. *Journal of Comparative Neurology, 160*, 269–289. doi:10.1002/cne.901600302

Espinoza, S. G., & Thomas, H. C. (1983, August). Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. *Brain Research, 272*, 137–144. doi:10.1016/0006-8993(83)90370-0

Essen, D. V., Olshausen, B. A., Anderson, C. H., & Gallant, J. T. (1991, July). Pattern recognition, attention, and information bottlenecks in the primate visual system. In B. P. Mathur, & C. Koch (A cura di), *SPIE Proceedings.* SPIE. doi:10.1117/12.45537

Felleman, D. J., & Essen, D. C. (1991, January). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1*, 1–47. doi:10.1093/cercor/1.1.1

Field, D. J. (1987, December). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*, 2379. doi:10.1364/josaa.4.002379

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013, May). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience, 16*, 974–981. doi:10.1038/nn.3402

Gilbert, E. N. (1980, June). Random Colorings of a Lattice of Squares in the Plane. *SIAM Journal on Algebraic Discrete Methods, 1*, 152–159. doi:10.1137/0601018

Girman, S. V., Sauvé, Y., & Lund, R. D. (1999, July). Receptive Field Properties of Single Neurons in Rat Primary Visual Cortex. *Journal of Neurophysiology, 82*, 301–311. doi:10.1152/jn.1999.82.1.301

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., & Tkačik, G. (2014, November). Variance predicts salience in central sensory processing. *eLife, 3*. doi:10.7554/elife.03722

Hoy, J. L., Yavorska, I., Wehr, M., & Niell, C. M. (2016, November). Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. *Current Biology, 26*, 3046–3052. doi:10.1016/j.cub.2016.09.009

Høydal, Ø. A., Skytøen, E. R., Moser, M.-B., & Moser, E. I. (2018, March). Object-vector coding in the medial entorhinal cortex. doi:10.1101/286286

Huberman, A. D., & Niell, C. M. (2011, September). What can mice tell us about how vision works? *Trends in Neurosciences, 34*, 464–473. doi:10.1016/j.tins.2011.07.002

Hughes, A. (1979, January). A schematic eye for the rat. *Vision Research, 19*, 569–588. doi:10.1016/0042-6989(79)90143-3

Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010, March). High-Level Visual Object Representations Are Constrained by Position. *Cerebral Cortex, 20*, 2916–2925. doi:10.1093/cercor/bhq042

Laughlin, S. (1981, October). A Simple Coding Procedure Enhances a Neuron\textquotesingles Information Capacity. *Zeitschrift für Naturforschung C, 36*, 910–912. doi:10.1515/znc-1981-9-1040

Li, Z., & Atick, J. (1994, May). Efficient stereo coding in the multiscale representation∗. *Network: Computation in Neural Systems, 5*, 157–174. doi:10.1088/0954-898x/5/2/003

Logothetis, N. K., & Sheinberg, D. L. (1996, March). Visual Object Recognition. *Annual Review of Neuroscience, 19*, 577–621. doi:10.1146/annurev.ne.19.030196.003045

Matteucci, G., & Zoccolan, D. (2020, May). Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. *Science Advances, 6*. doi:10.1126/sciadv.aba3742

Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B., & & Zoccolan, D. (2019). Nonlinear processing of shape information in rat lateral extrastriate cortex. *Journal of Neuroscience, 39(9)*, 1649-1670.

Montero, V. (1993, May). Retinotopy of cortical connections between the striate cortex and extrastriate visual areas in the rat. *Experimental Brain Research, 94*. doi:10.1007/bf00230466

Okazawa, G., Tajima, S., & Komatsu, H. (2016, September). Gradual Development of Visual Texture-Selective Properties Between Macaque Areas V2 and V4. *Cerebral Cortex*. doi:10.1093/cercor/bhw282

Olshausen, B. A., & Field, D. J. (1996, June). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609. doi:10.1038/381607a0

Olshausen, B. A., & Field, D. J. (1997, December). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research, 37*, 3311–3325. doi:10.1016/s0042-6989(97)00169-7

Olshausen, B., & Field, D. (2004, August). Sparse coding of sensory inputs. *Current Opinion in Neurobiology, 14*, 481–487. doi:10.1016/j.conb.2004.07.007

Orban, G. A., Essen, D. V., & Vanduffel, W. (2004, July). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences, 8*, 315–324. doi:10.1016/j.tics.2004.05.009

Parnavelas, J. G., Burne, R. A., & Lin, C.-S. (1981, December). Receptive field properties of neurons in the visual cortex of the rat. *Neuroscience Letters, 27*, 291–296. doi:10.1016/0304-3940(81)90445-6

Piasini, E., Soltuzu, L., Muratore, P., Caramellino, R., Vinken, K., Op De Beeck, H., . . . Zoccolan, D. (2021, July 21). Temporal stability of stimulus representation increases along rodent visual cortical hierarchies. *Nature Communications, 12*, 4448. doi:10.1038/s41467-021-24456-3

Pickard, D. K. (1980, September). Unilateral Markov fields. *Advances in Applied Probability, 12*, 655–671. doi:10.2307/1426425

Pitkow, X., & Meister, M. (2012, March). Decorrelation and efficient coding by retinal ganglion cells. *Nature Neuroscience, 15*, 628–635. doi:10.1038/nn.3064

Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision, 40*, 49–70. doi:10.1023/a:1026553619983

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014, June). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution, 75*, 165–183. doi:10.1016/j.ympev.2014.02.023

Prusky, G. T., Harker, K. T., Douglas, R. M., & Whishaw, I. Q. (2002, November). Variation in visual acuity within pigmented, and between pigmented and albino rat strains. *Behavioural Brain Research, 136*, 339–348. doi:10.1016/s0166-4328(02)00126-2

Purpura, K., Victor, J., & Katz, E. (1994, August 30). Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences, 91*, 8482-8486. doi:10.1073/pnas.91.18.8482

Rajalingham, R., Schmidt, K., & DiCarlo, J. J. (2015, September). Comparison of Object Recognition Behavior in Human and Monkey. *Journal of Neuroscience, 35*, 12127–12136. doi:10.1523/jneurosci.0573-15.2015

Rao, R. P., & Ballard, D. H. (1999, January). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*, 79–87. doi:10.1038/4580

Reinhold, A. S., Sanguinetti-Scheck, J. I., Hartmann, K., & Brecht, M. (2019, September). Behavioral and neural correlates of hide-and-seek in rats. *Science, 365*, 1180–1183. doi:10.1126/science.aax4705

Rust, N. C., & DiCarlo, J. J. (2010, September). Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT. *Journal of Neuroscience, 30*, 12978–12995. doi:10.1523/jneurosci.0179-10.2010

Sanderson, K. J., Dreher, B., & Gayer, N. (1991, June). Prosencephalic connections of striate and extrastriate areas of rat visual cortex. *Experimental Brain Research, 85*. doi:10.1007/bf00229410

Sefton, A. J., Dreher, B., Harvey, A. R., & Martin, P. R. (2015). Visual System. In *The Rat Nervous System* (p. 947–983). Elsevier. doi:10.1016/b978-0-12-374245-2.00030-9

Sereno, M. I., & Allman, J. M. (1991). Cortical visual areas in mammals. *The neural basis of visual function, 4*, 160–172.

Shannon, C. E. (1948, July). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Simoncelli, E. P. (2003, April). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology, 13*, 144–149. doi:10.1016/s0959-4388(03)00047-3

Simoncelli, E. P., & Olshausen, B. A. (2001, March). Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience, 24*, 1193–1216. doi:10.1146/annurev.neuro.24.1.1193

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982, November). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 216*, 427–459. doi:10.1098/rspb.1982.0085

Sziklai, G. (1956, September). Some studies in the speed of visual perception. *IEEE Transactions on Information Theory, 2*, 125–128. doi:10.1109/tit.1956.1056814

Sziklai, G. C. (1957, February). Some studies in the speed of visual perception. *Electrical Engineering, 76*, 150–152. doi:10.1109/ee.1957.6442982

Tafazoli, S., Filippo, A. D., & Zoccolan, D. (2012, January). Transformation-Tolerant Object Recognition in Rats Revealed by Visual Priming. *Journal of Neuroscience, 32*, 21–34. doi:10.1523/jneurosci.3932-11.2012

Tafazoli, S., Safaai, H., Franceschi, G. D., Rosselli, F. B., Vanzella, W., Riggi, M., . . . Zoccolan, D. (2017, April). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife, 6*. doi:10.7554/elife.22794

Tees, R. (1999, December). The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. *Behavioural Brain Research, 106*, 55–73. doi:10.1016/s0166-4328(99)00092-3

Tesileanu, T., Conte, M. M., Briguglio, J. J., Hermundstad, A. M., Victor, J. D., & Balasubramanian, V. (2020, August). Efficient coding of natural scene statistics predicts discrimination thresholds for grayscale textures. *eLife, 9*. doi:10.7554/elife.54347

Tkačik, G., Prentice, J. S., Victor, J. D., & Balasubramanian, V. (2010, October). Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences, 107*, 18149–18154. doi:10.1073/pnas.0914916107

Vermaercke, B., den Bergh, G. V., Gerich, F., & de Beeck, H. O. (2015, May). Neural discriminability in rat lateral extrastriate cortex and deep but not superficial primary visual cortex correlates with shape discriminability. *Frontiers in Neural Circuits, 9*. doi:10.3389/fncir.2015.00024

Vermaercke, B., Gerich, F. J., Ytebrouck, E., Arckens, L., de Beeck, H. P., & den Bergh, G. V. (2014, October). Functional specialization in rat occipital and temporal visual cortex. *Journal of Neurophysiology, 112*, 1963–1983. doi:10.1152/jn.00737.2013

Victor, J. D., & Conte, M. M. (1991, January). Spatial organization of nonlinear interactions in form perception. *Vision Research, 31*, 1457–1488. doi:10.1016/0042-6989(91)90125-o

Victor, J. D., & Conte, M. M. (2004, March). Visual working memory for image statistics. *Vision Research, 44*, 541–556. doi:10.1016/j.visres.2003.11.001

Victor, J. D., & Conte, M. M. (2006, November). Encoding and stability of image statistics in working memory. *Vision Research, 46*, 4152–4162. doi:10.1016/j.visres.2006.07.023

Victor, J. D., & Conte, M. M. (2012, June). Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America A, 29*, 1313. doi:10.1364/josaa.29.001313

Victor, J. D., Conte, M. M., & Chubb, C. F. (2017, September). Textures as Probes of Visual Processing. *Annual Review of Vision Science, 3*, 275–296. doi:10.1146/annurev-vision-102016-061316

Victor, J. D., Thengone, D. J., Rizvi, S. M., & Conte, M. M. (2015, December). A perceptual space of local image statistics. *Vision Research, 117*, 117–135. doi:10.1016/j.visres.2015.05.018

Vinje, W. E., & Gallant, J. L. (2000, February). Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision. *Science, 287*, 1273–1276. doi:10.1126/science.287.5456.1273

Vinken, K., Vogels, R., & de Beeck, H. O. (2017). Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. *Current Biology, 27*, 914-919. doi:https://doi.org/10.1016/j.cub.2017.02.024

Wagor, E., Mangini, N. J., & Pearlman, A. L. (1980, September). Retinotopic organization of striate and extrastriate visual cortex in the mouse. *The Journal of Comparative Neurology, 193*, 187–202. doi:10.1002/cne.901930113

Wallace, D. J., Greenberg, D. S., Sawinski, J., Rulla, S., Notaro, G., & Kerr, J. N. (2013, May). Rats maintain an overhead binocular field at the expense of constant fusion. *Nature, 498*, 65–69. doi:10.1038/nature12153

Whishaw, I. Q., & Kolb, B. (A cura di). (2004, September). *The Behavior of the Laboratory Rat.* Oxford University PressNew York. doi:10.1093/acprof:oso/9780195162851.001.0001

Willmore, B. D., Mazer, J. A., & Gallant, J. L. (2011, June). Sparse coding in striate and extrastriate visual cortex. *Journal of Neurophysiology, 105*, 2907–2919. doi:10.1152/jn.00594.2010

Wiskott, L., & Sejnowski, T. J. (2002, April). Slow Feature Analysis: Unsupervised Learning of Invariances. *Neural Computation, 14*, 715–770. doi:10.1162/089976602317318938

Yu, Y., Schmid, A. M., & Victor, J. D. (2015, April). Visual processing of informative multipoint correlations arises primarily in V2. *eLife, 4*. doi:10.7554/elife.06604

Zhaoping, L. (2019, October). A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology, 58*, 1–10. doi:10.1016/j.conb.2019.06.001

Zhu, S. C., Wu, Y., & Mumford, D. (1998). Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling. *International Journal of Computer Vision, 27*, 107–126. doi:10.1023/a:1007925832420

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016, May). Selectivity and tolerance for visual texture in macaque V2. *Proceedings of the National Academy of Sciences, 113*. doi:10.1073/pnas.1510847113

Ziemba, C. M., Perez, R. K., Pai, J., Hallum, L. E., Shooner, C., Kelly, J. G., & Movshon, J. A. (2019, July). Laminar differences in responses to naturalistic texture in macaque V1 and V2. doi:10.1101/710426

Zoccolan, D. (2015, May). Invariant visual object recognition and shape processing in rats. *Behavioural Brain Research, 285*, 10–33. doi:10.1016/j.bbr.2014.12.053

Zoccolan, D., Oertelt, N., DiCarlo, J. J., & Cox, D. D. (2009, May). A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences, 106*, 8748–8753. doi:10.1073/pnas.0811583106

Chapter II

Rat sensitivity to multipoint statistics is predicted by efficient coding of natural scenes

Abstract

Efficient processing of sensory data requires adapting the neuronal encoding strategy to the statistics of natural stimuli. Previously, Hermundstad et al. 2014 showed that local multipoint correlation patterns that are most variable in natural images are also the most perceptually salient for human observers, in a way that is compatible with the efficient coding principle. Understanding the neuronal mechanisms underlying such adaptation to image statistics will require performing invasive experiments that are impossible in humans. Therefore, it is important to understand whether a similar phenomenon can be detected in animal species that allow for powerful experimental manipulations, such as rodents. Here we selected four image statistics (from single- to four-point correlations) and trained four groups of rats to discriminate between white noise patterns and binary textures containing variable intensity levels of one of such statistics. We interpreted the resulting psychometric data with an ideal observer model. We found a sharp decrease in sensitivity from 2- to 4-point correlations and a further decrease from 4- to 3-point, these results were also confirmed by within-subject comparisons of rat sensitivity to pairs of statistics. This ranking fully reproduces the trend it was previously observed in humans, thus extending a direct demonstration of efficient coding to a species where neuronal and developmental processes can be interrogated and causally manipulated.

Introduction

It is widely believed that the tuning of sensory neurons is adapted to the statistical structure of the signals they must encode (Sterling and Laughlin, 2015). This normative principle, known as efficient coding, has been successful in explaining many aspects of neural processing in vision (Atick and Redlich, 1990; Fairhall et al., 2001; Laughlin, 1981; Olshausen and Field, 1996; Pitkow and Meister, 2012), audition (Carlson et al., 2012; Smith and Lewicki, 2006) and olfaction (Teşileanu et al., 2019), including adaptation (Młynarski and Hermundstad, 2020) and gain control (Schwartz and Simoncelli, 2001). In Hermundstad et al., 2014, the authors analysed the relationship between human sensitivity to multipoint correlations and their variability in natural images. They first scanned an ensemble of natural images patches using a 2x2 window and measured the frequency of each correlation in each patch. From this data they extracted the variability of each correlation across all the patches. Then, they compared this variability to the human sensitivity estimated through a behavioral experiment where subjects were asked to identify patches of visual textures containing the statistic inside a white noise background. The authors reported that human sensitivity to this visual textures defined by local multipoint correlations depends on the variability of such correlations across natural scenes. This allocation of resources to features

that are the most variable in the environment, and thus more informative about its state, is accounted for by efficient coding, demonstrating its role as an organizing principle also at the perceptual level (Hermundstad et al., 2014; Tesileanu et al., 2020; Tkačik et al., 2010). However, it remains unknown whether this preferential encoding of texture statistics that are the most variable across natural images is a general principle underlying visual perceptual sensitivity across species. Although some evidence exists for differential neural encoding of multipoint correlations in macaque V2 (Yu et al., 2015) and V1 (Purpura et al., 1994), the sensitivity ranking we previously reported in Hermundstad et al., 2014 has not been investigated in any species, other than humans (Hermundstad et al., 2014; Tesileanu et al., 2020; Tkačik et al., 2010; Victor and Conte, 2012). Moreover, while monkeys are standard models of advanced visual processing (DiCarlo et al., 2012; Kourtzi and Connor, 2011; Lehky and Tanaka, 2016; Nassi and Callaway, 2009; Orban, 2008), they are less amenable than rodents to causal manipulations (e.g., optogenetic or controlled rearing) to interrogate how neural circuits may adapt to natural image statistics. On the other hand, rodents have emerged as powerful model systems to study visual functions during the last decade (Glickfeld et al., 2014; Glickfeld and Olsen, 2017; Huberman and Niell, 2011; Katzner and Weigelt, 2013; Niell and Scanziani, 2021; Reinagel, 2015; Zoccolan, 2015). Rats, in particular, are able to employ complex shape processing strategies at the perceptual level (Alemi-Neissi et al., 2013; De Keyser et al., 2015; Djurdjevic et al., 2018; Vermaercke and Op de Beeck, 2012), and rat lateral extrastriate cortex shares many defining features with the primate ventral stream (Kaliukhovich and Op de Beeck, 2018; Matteucci et al., 2019; Piasini et al., 2021; Tafazoli et al., 2017; Vermaercke et al., 2014; Vinken et al., 2017). More importantly, it was recently shown that rearing newborn rats in controlled visual environments allows causally testing long-standing hypotheses about the dependence of visual cortical development from natural scene statistics (Matteucci and Zoccolan, 2020). Establishing the existence of a preferential encoding of less predictable statistics in rodents is therefore crucial to understand the neural substrates of efficient coding and its relationship with postnatal visual experience.

Results

To address this question, we measured rat sensitivity to visual textures defined by local multipoint correlations, training the animals to discriminate binary textures containing structured noise from textures made of white noise (Figure 1A). The latter were generated by independently setting each pixel to black or white with equal probability, resulting in no spatial correlations. Structured textures, on the other hand, were designed to enable precise control over the type and intensity of the correlations they contained. To generate these textures we built and published a software library (Piasini, 2021) that implements the method developed in (Victor and Conte, 2012). Briefly, for any given type of multipoint correlation (also termed a *statistic* in what follows), we sampled from the distribution over binary textures that had the desired probability of occurrence of that statistic, but otherwise contained the least amount of structure (i.e., had maximum entropy). The probability of occurrence of the pattern was parametrized by the *intensity* of the corresponding statistic, determined by a parity count of white or black pixels inside tiles of 1, 2, 3 or 4 pixels (termed *gliders*) used as the building blocks of the texture (Victor and Conte, 2012). When the intensity is zero, the texture does not contain any structure – it is the same as white noise (Figure 1A, left). When the intensity is +1, every possible placement of the glider across the texture contains an

even number of white pixels, while a level of -1 corresponds to all placements containing an odd number of white pixels. Intermediate intensity levels correspond to intermediate fractions of gliders containing the even parity count. The structure of the glider and the sign of the intensity level dictate the appearance of the final texture. For instance (see examples in Figure 1A, right), for positive intensity levels, a 1-point glider produces textures with increasingly large luminance, a 2-point glider produces oriented edges and a 4-point glider produces rectangular blocks. A 3-point glider produces L-shape patterns, either black or white depending on whether the intensity is negative or positive.

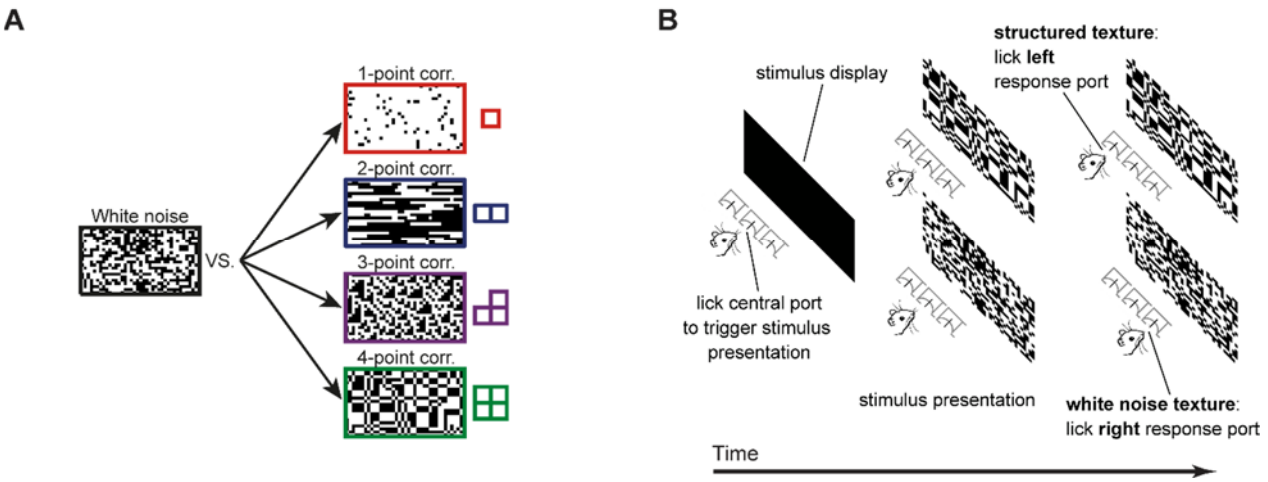


Figure 1. Visual stimuli and behavioral task. (**A**) Schematic of the four kinds of texture discrimination tasks administered to the four groups of rats in our study. Each group had to discriminate unstructured binary textures containing white noise (example on the left) from structured binary textures containing specific types of local multipoint correlations among nearby pixels (i.e., 1-, 2-, 3- or 4-point correlations; examples on the right). The textures were constructed to be as random as possible (maximum entropy), under the constraint that the strength of a given type of correlation matched a desired level. The strength of a correlation pattern was quantified by the value (intensity) of a corresponding statistic (see main text), which could range from 0 (white noise) to 1 (maximum possible amount of correlation). The examples shown here correspond to intensities of 0.85 (1- and 2-point statistics) and 0.95 (3- and 4-point statistics). (**B**) Schematic representation of a behavioral trial. Left and center: animals initiated the presentation of a stimulus by licking the central response port placed in front of them. This prompted the presentation of either a structured (top) or an unstructured (bottom) texture. Right: in order to receive the reward, animals had to lick either the left or right response port to report whether the stimulus contained the statistic (top) or the noise (bottom).

Notably, 2-point and 3-point gliders are associated to multiple distinct multipoint correlations, corresponding to different spatial glider configurations. For instance, 2-point correlations can arise from horizontal (–), vertical (|) or oblique gliders (/, \), while 3-point correlations can give rise to L patterns with various orientations ($\theta_{\urcorner}$, $\theta_{\neg}$, $\theta_{\llcorner}$, $\theta_{\lrcorner}$). In their previous study with human participants, (Hermundstad et al., 2014) tested all these 2-point, 3-point and 4-point configurations, as well as 11 of their pairwise combinations, for a total of 20 different texture statistics. In that set of experiments, they did not test textures defined by 1-point correlations because, by construction, the method used to measure the variability of texture statistics across natural images could not be applied to the 1-point statistic. In our current study, practical and ethical constraints prevented us from

measuring rat sensitivity to a large number of statistic combinations, because a different group of animals had to be trained with each tested statistic (see below), meaning that the number of rats required for the experiments increased rapidly with the number of statistics studied. Therefore, we chose to test the 4-point statistic, as well as one each of the 2-point and 3-point statistics (those shown in Figure 1A). One of the 3-point statistics (corresponding to the glider θ⌟) was randomly selected among the four available, since in the previous study on humans, no difference was found among the variability of distinct 3-point textures across natural images, and aggregate human sensitivity to 3-point correlations was measured without distinguishing among glider configurations. As for the 2-point statistic, we selected one of the two gliders (the horizontal one) that yielded the largest sensitivity in humans, so as to include in our stimulus set at least an instance of both the most discriminable (2-point –) and least discriminable (3-point θ⌟) textures. In addition, we also tested the 1-point statistic because, given the well-established sensitivity of the rat visual system to luminance changes (Minini and Jeffery, 2006; Tafazoli et al., 2017; Vascon et al., 2019; Vermaercke and Op de Beeck, 2012), performance with this statistic served as a useful benchmark against which to compare rat discrimination of the other, more complex textures. Finally, while in Hermundstad et al., 2014, both positive and negative values of the statistics were probed against white noise, here we tested only one side of the texture intensity axis (either positive, for 1-, 2- and 4-point configurations, or negative, for 3-point ones) – again, with the goal of limiting the number of rats used in the experiment (see Methods for more details on the rationale behind the choice of statistics and their polarity, and see Discussion for an assessment of the possible impact of these choices on our conclusions).

For each of the four selected image statistics, we trained a group of rats to discriminate between white noise and structured textures containing that statistic with nonzero intensity (Figure 1A). Each trial of the experiment started with the rat autonomously triggering the presentation of a stimulus by licking the central response port within an array of three (Figure1B). The animal then reported whether the texture displayed over the monitor placed in front of him contained the statistic (by licking the left port) or white noise (by licking the right port). The rat received liquid reward for correct choices and was subjected to a time-out period for incorrect ones (Figure 1B). In the initial phase of the experiment, the intensity of the statistic was set to a single level, close to the maximum (or minimum, in case of the 3-point statistic, for which we used only negative values), to make the discrimination between structured textures and white noise as easy as possible for naïve rats that had to learn the task from scratch. The learning curves of four example rats, one per group, are shown in Figure 2A. In the following phase of the experiment, the intensity of the statistic was gradually reduced using an adaptive staircase procedure (see Materials and methods) to make the task progressively harder. The asymptotic levels of the statistics reached across consecutive training sessions by four example rats, one per group, are shown in Figure 2B. Following this training, rats were subjected to: 1) a main testing phase, where textures were sampled at regular intervals along the intensity level axis and were randomly presented to the animals; and 2) a further testing phase, where rats originally trained with a given statistic were probed with a different one (see Materials and methods for details on training and testing).

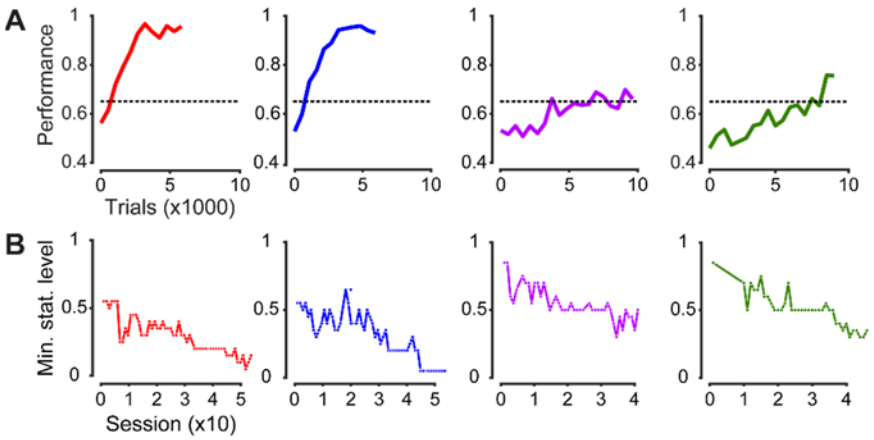


Figure 2. Learning curves of four example rats during the two initial training phases. **(A)** Learning curves of four example animals belonging to the first batch of rats (see Materials and methods), each trained with one of the four texture statistics: 1-point (red), 2-point (blue), 3-point (purple) and 4-point (green). During this phase, animals had to discriminate textures containing a single high intensity level of the assigned statistic (see examples in the Figure 1A, right) from white noise. Each curve shows the performance of a rat (i.e., the fraction of correct choices), as computed in consecutive blocks of 500 trials. The dashed line indicates the 65% criterion performance that rats of the first batch had to reach in order to be moved to the next experimental phase (see Materials and methods). As exemplified by these four animals, rats trained on different statistics required different lengths of time to learn the discrimination – animals trained on 1- and 2-point statistics reached and maintained a performance above criterion considerably earlier than those trained on 3- and 4-point. Specifically, considering only those rats that reached criterion performance, animals trained on 1-point correlations required on average 3.5±0.3 blocks of 500 trials (n=9) to reach criterion; those trained on 2-point correlations 5.3±1.3 (n=8); those trained on 3-point statistic 3.4±1.6 (n=5); and those trained on 4-point correlations 11.8±1.6 (n=10). **(B)** Minimal intensity levels of the statistics that the same animals shown in A were able to discriminate from white noise across consecutive sessions (colored dots) of the second phase of the experiment, when the rats were exposed to progressively lower intensities through a staircase procedure (see Materials and methods). Sessions where the number of trials performed by the animals was below 50 are not shown in these graphs.

The main test phase yielded psychometric curves showing the sensitivity of each animal in discriminating white noise from the structured texture with the assigned statistic (example in Figure 3A, black dots). To interpret results, we developed an ideal observer model, in which the presentation of a texture with a level of the statistic equal to *s* produces a percept *x* sampled from a truncated Gaussian distribution centered on the actual value of the statistic (*s*) with a fixed standard deviation σ (Fleming et al., 2013; Geisler, 2011). Here, σ measures the "blurriness" in the animal's sensory representation for a particular type of statistic (i.e., the perceptual noise) and, consequently, its inverse 1/σ captures its resolution, or sensitivity – i.e., the perceptual threshold for discriminating a structured texture from white noise. As detailed in the Materials and methods, our ideal observer model yields the

psychometric function giving the probability of responding “noise” at any given level of the statistic *s* as

$$p(\text{report noise}|s) = \frac{\Phi\left[\frac{x^*(\alpha,\sigma)-s}{\sigma}\right]-\Phi\left[\frac{-1-s}{\sigma}\right]}{\Phi\left[\frac{1-s}{\sigma}\right]-\Phi\left[\frac{-1-s}{\sigma}\right]},$$

where $\Phi(x)$ is the standard Normal cumulative density function, α captures the animal’s prior choice bias and $x^*(\alpha,\sigma)$ is the decision boundary used by the animal to divide the perceptual axis into “noise” and “structured texture” regions. The two free parameters of the model (α and σ) parametrize the psychometric function (example in Figure 3A, blue curve) and can be estimated from behavioral data by maximum likelihood. Prior bias (α) and sensitivity (1/σ) are related, respectively, to the horizontal offset and slope of the curve.

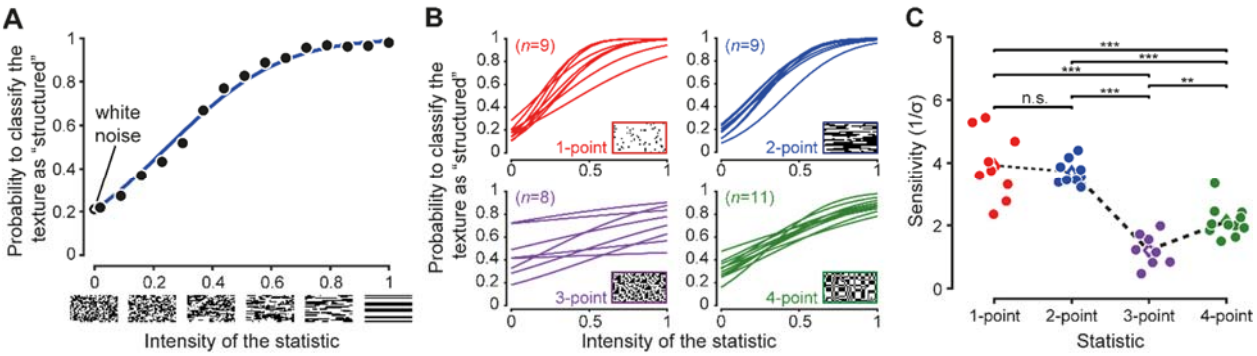


Figure 3. Rat sensitivity to multipoint correlations. (**A**) Psychometric data for an example rat trained on 2-point correlations. Black dots: fraction of trials in which a texture with the corresponding intensity of the statistic was correctly classified as “structured”. Empty black circle: fraction of trials the rat has judged a white noise texture as containing the statistic. Blue line: psychometric function corresponding to the fitted ideal observer model (see main text). (**B**) Psychometric functions obtained for all the rats tested on the four statistics (*n* indicates the number of animals in each group). (**C**) Values of the perceptual sensitivity 1/σ to each of the four statistics. Filled dots: individual rat estimates. Empty diamonds: group averages. The dashed line emphasizes the sensitivity ranking observed for the four statistics. Significance markers ** and *** indicate, respectively, $p < 0.01$ and $p < 0.001$ for a two-sample t-test with Holm-Bonferroni correction. The same analysis was repeated in Figure 2–figure supplement 1 including only the rats that reached a certain performance criterion during the initial training.

Fitting this model to the behavioral choices of rats in the four groups led to psychometric functions with a characteristic shape, which depended on the order of the multipoint statistic an animal had to discriminate (Figure 3B). In particular, the sensitivity 1/σ followed a specific ranking among the groups (Figure 3C), being higher for 1- and 2-point than for 3-point (p_1<0.001 and p_2<0.001, two sample t-test with Holm-Bonferroni correction) and 4-point (p_1<0.001, p_2<0.001) correlations, and larger for 4-point than 3-point correlations (p<0.01). When focusing on the texture statistics that had been also tested in our previous study (i.e., 2-point horizontal, 3-point and 4-point correlations), this sensitivity ranking was the same as the one observed in humans and as the variability ranking measured across natural images (Hermundstad et al., 2014): 2-point horizontal > 4-point > 3-point. To make

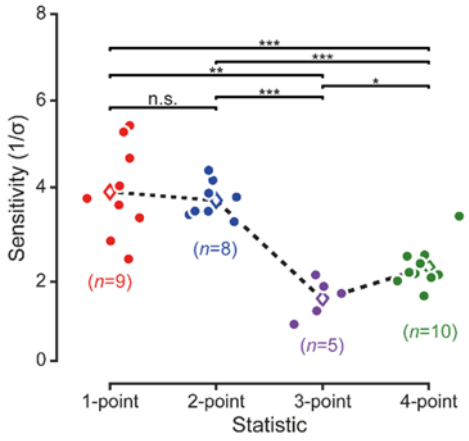


Figure 4. Rat sensitivity to multipoint correlations after excluding animals that did not reach the 65% correct discrimination criterion in phase I of the training. Same data as in Figure 2C, but after excluding 1 rat from the 2-point group, 3 rats from the 3-point group and 1 rat from the 4-point group. Significance markers as in Figure 2C.

sure that this result was not due to the inclusion in the analysis of rats that didn't learn the task, we repeated the comparison shown in Figure 3C keeping only the animals that in the first learning phase reached a 65% correct discrimination criterion. Figure 4 shows that the resulting sensitivity ranking didn't change, as well as the significance of the different pairwise comparisons.

For the set of statistics that were studied both here and in Hermundstad 2014 (2-, 3- and 4-point correlations), the actual values of the rat sensitivity matched, up to a scaling factor, both the human sensitivity and the standard deviation of the statistics in natural images (Figure 5). This match was quantified with the "degree of correspondence", defined in Hermundstad 2014, which takes on values between 0 and 1, with 1 indicating perfect quantitative match (see Materials and methods for details). The degree of correspondence was 0.986 between rat sensitivity and image statistics (p-value: 0.07, Monte Carlo test), and 0.990 between rat sensitivity and human sensitivity (p-value: 0.05, Monte Carlo test). For reference, Hermundstad 2014 reported values between 0.987 to 0.999 for the degree of correspondence between human sensitivity and image statistics. This indicates not only a qualitative but also a quantitative agreement between our findings and the pattern of texture sensitivity predicted by efficient coding.

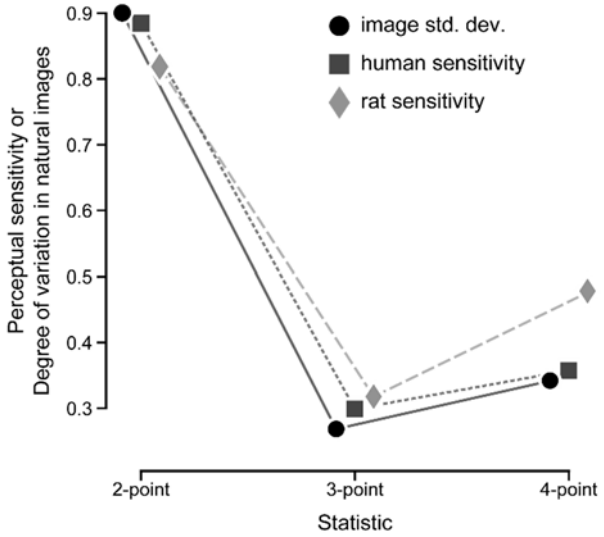


Figure 5. Quantitative match between rat sensitivities, human sensitivities and texture variability across natural images. For the three texture statistics that were tested both in our current study with rats and in our earlier study with humans (Hermundstad et al., 2014), rat average sensitivities are compared to human average sensitivities and to the variability of these statistics across natural scenes (data from Hermundstad et al. 2014). The three sets of data points have been scaled in such a way that each triplet of sensitivity values had Euclidean norm = 1, so as to allow an easier qualitative comparison (for a quantitative comparison see main text).

To further validate these findings, we performed additional within-group and within-subject comparisons. To this end, each group of animals was either tested with a new statistic or was split into two subgroups, each tested with a different statistic. Results of these additional experiments are reported in Figure 6, comparing the sensitivity to the new statistic(s) with the sensitivity to the originally learned statistic (colored symbols without and with halo, respectively) for each group/subgroup. Rats trained on 1- and 2-point statistics (the most discriminable ones; see Figure 3C) performed poorly with higher-order correlations (compare the green and purple star with the red star, and the green and purple cross with the blue cross in Figure 6), while animals trained on the 4-point statistic performed on 2-point correlations as well as rats that were originally trained on those textures (compare the blue square to the blue cross). This shows that the better discriminability of textures containing lower-order correlations is a robust phenomenon, which is independent of the history of training and observable within individual subjects. Moreover, performance on 4-point correlations was higher than performance on 3-point correlations for each group of rats (compare the green to the purple symbols connected by a line). This was true, in particular, not only for rats trained on 4-point and switching to 3-point (green vs. purple square, $p < 0.01$, paired one-tailed t-test) but even for rats trained on 3-points and switching to 4-point (green vs. purple triangle, $p < 0.05$, paired one-tailed t-test). This means that the larger

discriminability of the 4-point statistic, as compared to the 3-point one, is a statistically robust phenomenon within individual subjects.

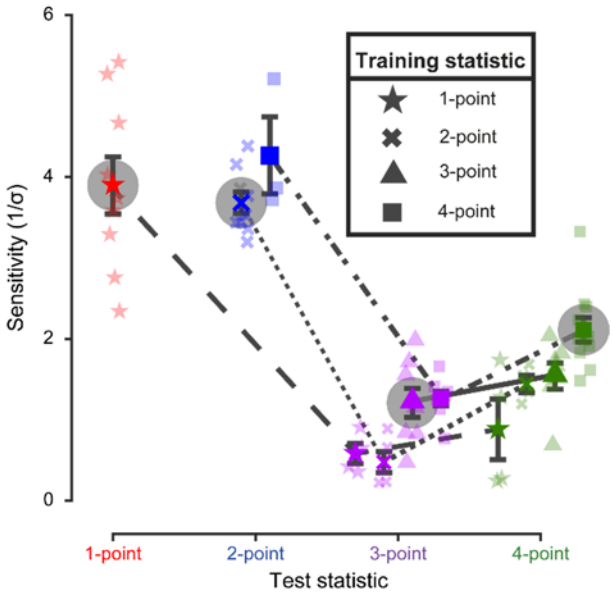


Figure 6. Rat sensitivity to multipoint correlations – dependence on training history and within-subject analysis. Colored points with halo show the average sensitivities (with SEM) of the four groups of rats to the statistics (indicated by the symbols in the key) they were originally trained on (i.e., same data as the colored diamonds in Figure 5C). The other colored symbols connected by a line show the average sensitivities (with SEM) obtained when subgroups of rats originally trained on a given statistic (as indicated by the symbol in the key) were tested with different statistics (as indicated in abscissa). Specifically: 1) out of the nine rats originally trained/tested with 1-point correlations (star), four were tested with 3-point (purple star) and four with 4-point (green star) correlations (1 rat did not reach this test phase); 2) out of the nine rats originally trained/tested with 2-point correlations (cross), five were tested with 3-point (purple cross) and four with 4-point (green cross) correlations; 3) out of the eight rats originally trained/tested with 3-point correlations (triangle), seven were tested with 4-point (green triangle) correlations (1 rat did not reach this test phase); and 4) out of the eleven rats originally trained/tested with 4-point correlations (square), eight were tested with 3-point (purple square) and three with 2-point correlations (blue square). Sensitivities achieved by individual animals are represented as shaded data points with the corresponding symbol/color combination.

Discussion

Overall, our results show that rat sensitivity to multipoint statistics is similar to the one it was previously observed in humans and to the variability of multipoint correlations it was previously measured across natural images (Hermundstad et al., 2014; Tkačik et al., 2010). This agreement holds both qualitatively and quantitatively (Figures 3-6). Importantly, we found the expected sensitivity ranking (2-point horizontal > 4-point > 3-point) to be robust not only across groups (Figure 3C) but also for animals that were sequentially tested with multiple texture statistics (Figure 6) – and even at the within-subject level for the crucial 3-

point vs. 4-point comparison. Moreover, we found a high degree of correspondence between rat and human sensitivities (Figure 5).

A potential limitation of our study is related to our stimulus choices, both in terms of selected texture statistics and polarity (i.e., negative vs. positive intensity). A first possible issue is whether the three texture statistics that were tested in both the present study and in (Hermundstad et al, 2014) are sufficient to allow a meaningful comparison between rat and human sensitivities, as well as rat sensitivity and texture variability in natural scenes. We addressed this matter at the level of experimental design, by carefully choosing the three statistics that, based on the sensitivity ranking observed in humans, would have yielded the cleanest signature of efficient coding (Hermundstad et al., 2014). That is, we selected two statistics that were, respectively, maximally and minimally variable across natural images, and yielded the largest and lowest sensitivities in humans: horizontal 2-point correlations and one of the 3-point correlations. The 4-point correlation was then a natural choice as the third statistic, as it was the only one characterized by a differently-shaped glider. Additionally, human sensitivity to this statistic, as well as its variability across natural images, is only slightly larger than for the 3-point configurations. Therefore, finding a reliable sensitivity difference between 3-point and 4-point textures also for rats would have provided strong evidence for matching texture sensitivity across the two species. Due to the experimental limitations discussed in the Results and the Materials and Methods sections, we were unable to analyse one of the oblique 2-point statistics, for which human sensitivity takes on an intermediate value between the 2-point horizontal and 3-point correlations, and that in humans allows one to differentiate between the predictions of efficient coding and those stemming from an oblique effect for patterns that are rotated versions of each other (Hermundstad et al., 2014).

The second potential limitation is related to the choice of polarity (positive or negative intensity values for the examined statistics). This choice was guided by different considerations depending on the kind of statistic. For 1-point correlations we chose positive intensity values because they yield patterns that are brighter than white noise. Since previous work from our group has shown that rat V1 neurons are very sensitive to increases of luminance (Tafazoli et al., 2017; Vascon et al., 2019), our choice ensured that 1-point textures were highly distinguishable from white noise (as indeed observed in our data; see Figure 3 B-C), which was the key requirement for our benchmark statistic. This enabled us to guard against issues in our task design: if the animals had failed to discriminate 1-point textures, this would have suggested an overall inadequacy of the behavioral task rather than a lack of perceptual sensitivity to luminance changes. For 2-point and 4-point statistics we also used positive intensity values – a choice dictated by the need of testing a rodent species that has much lower visual acuity than humans (Keller et al., 2000; Prusky et al., 2002; Zoccolan, 2015). Positive 2-point and 4-point correlations give rise to large features (thick oriented stripes and wide rectangular blocks made of multiple pixels with the same color), while negative intensities produce higher spatial frequency patterns, where color may change every other pixel (see Fig. 2A in Hermundstad et al., 2014). Therefore, using negative 2-point and 4-point statistics would have introduced a possible confound, since low sensitivity to these textures could have been simply due to the low spatial resolution of rat vision. For 3-point correlations, polarity does not affect the shape and size of the emerging visual patterns, but it determines their contrast. Positive and negative intensities yield L-shaped patches that are, respectively, white and black. In this case, we chose the latter to

make sure that the well-known dominance of OFF responses observed across the visual systems of many mammal species would not play in favor of finding the lowest sensitivity for the 3-point statistic. In fact, several studies have shown that primary visual neurons of primates and cats respond more strongly to black than to white spots and oriented bars (Liu and Yao, 2014; Xing et al., 2010; Yeh et al., 2009). A very recent study has shown that this is the case also for the central visual field of mice, although in the periphery OFF and ON response are more balanced (Williams et al., 2021). Indeed, the asymmetry begins already in the retina where there are more OFF cells than ON cells (Ratliff et al., 2010). Since in our behavioral rigs rats face frontally the stimulus display (Figure 1B) and maintain their head oriented frontally during stimulus presentation (Vanzella et al., 2019), it was important that the L-shaped patterns produced by 3-point correlations had the highest saliency. Choosing negative intensity values ensured that this was the case, thus excluding the possibility that the low-sensitivity found for 3-point textures (Figures 3-6) was partially due to presentation at a suboptimal contrast. Notwithstanding these considerations, one could wonder whether probing also the opposite polarities of those tested in our study would be desirable for a tighter test of the efficient coding principle. Previous studies, however, found human sensitivity to be nearly identical for negative and positive intensity variations of each of the statistic tested in our study: 1-point, 2-point, 3-point and 4-point correlations (Victor and Conte, 2012), even in the face of asymmetries of the distribution of the corresponding statistic in natural images (see Figure 3 – figure supplement 9 in Hermundstad et al 2014). In the present work, we have accordingly decided to focus the available resources on the differences between different statistics, rather than between positive and negative intensities of the same statistic.

In summary, our choices of texture types and their polarity were all dictated by the need of adapting to a rodent species texture stimuli that, so far, have only been used in psychophysics studies with humans (Hermundstad et al., 2014; Tesileanu et al., 2020; Tkačik et al., 2010; Victor and Conte, 2012) and neurophysiology studies in monkeys (Purpura et al., 1994; Yu et al., 2015). Our goal was to maximize the sensitivity of the comparison with humans and natural image statistics, while reducing the possible impact of phenomena (such as rat low visual acuity and the dominance of OFF responses) that could have acted as confounding factors. Thanks to these measures, our findings provide a robust demonstration that a rodent species and humans are similarly adapted to process the statistical structure of visual textures, in a way that is consistent with the computational principle of efficient coding. This attests to the fundamental role of natural image statistics in shaping visual processing across species, and opens a path towards a causal test of efficient coding through the altered-rearing experiments that small mammals, such as rodents, allow (Hunt et al., 2013; Matteucci and Zoccolan, 2020; White and Fitzpatrick, 2007).

Materials and Methods

Psychophysics experiments

Subjects

A total of 42 male adult Long Evans rats (Charles River Laboratories) were tested in a visual texture discrimination task. Animals started the training at 10 weeks, after 1 week of

quarantine upon arrival in our institute and 2 weeks of handling to familiarize them with the experimenters. Their weight at arrival was approximately 300g and they grew to over 600g over the time span of the experiment. Rats always had free access to food but their access to water was restricted in the days of the behavioral training (5 days a week). They received 10–20ml of diluted pear juice (1:4) during the execution of the discrimination task, after which they were also given free access to water for the time needed to reach at least the recommended 50ml/Kg intake per day.

The number of rats was chosen in such a way to yield meaningful statistical analyses (i.e., to have about 10 subjects for each of the texture statistic tested in our study), under the capacity constraint of our behavioral rig. The rig allows to simultaneously test 6 rats, during the course of 1-1.5 hours (Zoccolan 2015; Djurdjevic et al. 2018). Given the need for testing four different texture statistics, we started with a first batch of 24 animals (i.e., 6 per statistics), which required about 6 h of training per day. This first batch was complemented with a second one of 18 more rats, again divided among the four statistics (see below for details), so as to reach the planned number of about 10 animals per texture type. The first batch arrived in November 2018 and was tested throughout most of 2019; the second group arrived in September 2019 and was tested throughout most of 2020. In the first batch, 4 animals did not reach the test phase (i.e., the phase yielding the data shown in Fig. 2A, B), because three of them did not achieve the criterion performance during the initial training phase (see below) and one died shortly after the beginning of the study. In the second batch one rat died before reaching the test phase and two more died before the last test phase with switched statistics (i.e., the phase yielding the data of Fig. 2C).

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 25271, submitted on December 1, 2014 and approved on September 4, 2015, approval 940/2015-PR).

Experimental setup

Rats were trained in a behavioral rig consisting of two racks, each equipped with three operant boxes (a picture of the rig and a schematic of the operant box can be found in previous studies (Zoccolan 2015; Djurdjevic et al. 2018)). Each box was equipped with a 21.5” LCD monitor (ASUS VEZZHR) for the presentation of the visual stimuli and an array of three stainless-steel feeding needles (Cadence Science), serving as response ports. To this end, each needle was connected to a led-photodiode pair to detect when the nose of the animal approached and touched it (a Phidgets 1203 input/output device was used to collect the signals of the photodiodes). The two lateral feeding needles were also connected to computer-controlled syringe pumps (New Era Pump System NE-500) for delivery of the liquid reward. In each box, one of the walls bore a 4.5 cm-diameter viewing hole, so that a rat could extend its head outside the box, face the stimulus display (located at 30cm from the hole) and reach the array with the response ports.

Choice of image statistics to be used in the experiment

As mentioned in the main text, in our experiment we studied the 1-point and 4-point statistic, as well as one of the 2-point and one of the 3-point statistics. In the nomenclature introduced by Victor and Conte 2012, these are, respectively, the γ, α, β₋ and θ statistics. By

comparison, in humans, Victor and Conte 2012 studied a total of 5 statistics (the same we tested, plus β_l), while Hermundstad et al. 2014 tested many more, including combinations of statistic pairs, although they did not investigate γ. Our choice of which statistics to test was constrained on practical and ethical grounds by the need to use the minimum possible number of animals in our experiments, which led us to study one representative statistic per order of the glider. We note also that we decided to test the γ statistic, even though this was omitted by Hermundstad et al. 2014 (as explained in that paper, the method used to assess the variability of all other multipoint correlation patterns in natural images can't be applied to γ by construction, because the binarization threshold used for images is such that γ = 0 for all images in the dataset). The reason for including γ was that it provided a useful control on the effectiveness of our experimental design, as (unlike for the other visual patterns) it was already known that rats can discriminate simple stimuli that differ by average luminosity (Tafazoli et al. 2017; Vascon et al. 2018). As mentioned in the Discussion, failure of the rats to discriminate 1-point textures would have indicated a likely issue in the design of the task.

Human sensitivity to multipoint correlation patterns does not distinguish between positive and negative values of the statistics (Victor and Conte 2012). Therefore, again in order to minimize the number of animals necessary to the experiment, we only collected data for positive values of the γ, β and α statistics, and negative values of the θ statistic (see below for the specific values used). Unlike 2- or 4-point statistics, θ statistics change contrast under a sign change (namely, positive θ values correspond to white triangular patterns on a black background, and negative θ values correspond to black triangular patterns on a white background). On the other hand, dominance of OFF responses (elicited by dark spots on a light background) has been reported in mammals, including primates, cats, and rodents (Liu and Yao 2014; Xing, Yeh, and Shapley 2010; Yeh, Xing, and Shapley 2009; Williams et al. 2021). Therefore, we reasoned that if rats, unlike humans, were to have a different sensitivity to positive and negative θ values, the sensitivity to negative θ would be the higher of the two.

Finally, for the sake of simplicity, whenever in the text we refer to the "intensity" of a statistic, this should be interpreted as the absolute value of the intensity as defined by Victor and Conte 2012. This has no effect when describing γ, β, or α statistics, and only means that any value reported for θ should be taken with a sign flip (i.e., negative instead of positive values) if trying to connect formally to the system of coordinates in Victor and Conte 2012.

Visual stimuli

Maximum-entropy textures were generated using the methods described by Victor and Conte (Victor and Conte 2012). To this end, we implemented a standalone library and software package that we have since made publicly available as free software (Piasini 2021). In the experiment, we used white noise textures as well as textures with positive levels of four different multipoint statistics, as described above (see also Fig. 1A). It should be noted that, with the exception of the extreme value of the γ statistic (γ = 1 corresponds to a fully white image), the intensity level of a given statistic does not specify deterministically the resulting texture image. In our experiment, for any intensity level of each statistic, multiple, random instances of the textures were built to be presented to the rats during the discrimination task (see below for more details).

Subjects had to discriminate between visual textures containing one of the 4 selected statistics and white noise. Each texture had a size of 39x22 pixels and occupied the entire monitor (full- field stimuli). The pixels had a dimension of about 2 degrees of visual angle. Given that the maximal resolution of rat vision is about 1 cycle per degree (Keller et al. 2000; Prusky, West, and Douglas 2000; Prusky et al. 2002), such a choice of the pixel size guaranteed that the animals could discriminate between neighboring pixels of different color. Textures were showed at full-contrast over the LCD monitors that were calibrated in such a way to have minimal luminance of 0.126±0.004 cd/mm (average ± SD across the six monitors), maximal luminance of 129±5 cd/mm, and an approximately linear luminance response curve.

Discrimination task

Each rat was trained to: 1) touch the central response port to trigger stimulus presentation and initiate a behavioral trial; and 2) touch one of the lateral response ports to report the identity of the visual stimulus and collect the reward (all the animals were trained with the following stimulus/response association: structured texture left response port; white noise texture right response port). The stimulus remained on the display until the animal responded or for a maximum of 5 seconds, after which the trial was considered as ignored. In case of a correct response the stimulus was removed, a positive reinforcement sound was played and a white (first animal batch) or gray (second batch) background was shown during delivery of the reward. In case of an incorrect choice, the stimulus was removed and a 1-3s time-out period started, during which the screen flickered from middle-gray to black at a rate of 10 Hz, while a “failure” sound was played. During this period the rat was not allowed to initiate a new trial. To prevent the rats from making impulsive random choices, trials where the animals responded in less than 300 or 400ms were considered as aborted: the stimulus was immediately removed and a brief sound was played. In each trial, the visual stimuli had the same probability (50%) of being sampled from the pool of white noise textures or from the pool of structured textures, with the constraint that stimuli belonging to the same category were shown for at most n consecutive trials (with n varying between 2 and 3 depending on the animal and on the session), so as to prevent the rats from developing a bias towards one of the response ports.

Stimulus presentation, response collection and reward delivery were controlled via workstations running the open source suite MWorks (https://mworks.github.io).

Experimental design

Each rat was assigned to a specific statistic, from 1- to 4-point, for which it was trained in phases I and II and then tested in phase III. Generalization to a different statistic from the one the rat was trained on was assessed in phase IV. Out of the 42 rats, 9 were trained with 1-point statistics, 9 with 2-point, 12 with 3-point and 12 with 4-point. The animals that reached phase III were 9, 9, 8 and 11, respectively, for the four statistics.

Phase I

Initially, rats were trained to discriminate unstructured textures made of white noise from structured textures containing a single high-intensity level of one of the statistics (for 1-point and 2-point: 0.85; for 3-point and 4-point: 0.95). To make sure that the animals learned a general distinction between structured and unstructured textures (and not between specific

instances of the two stimulus categories), in each trial both kinds of stimuli were randomly sampled (without replacement) from a pool of 350 different textures. Since the rats typically performed between 200 and 300 trials in a training session, every single texture was not shown more than once. A different pool of textures was used in each of the five days within a week of training. The same five texture pools were then used again (in the same order) the following week. Therefore, at least 7 days had to pass before a given texture stimulus was presented again to a rat.

For the first batch of rats, we moved to the second phase of the experiment all the animals that were able to reach at least an average performance of 65% correct choices over a set of 500 trials (collected across a variable number of consecutive sessions). Based on this criterion, 2 rats tested with 3-point textures and 1 rat tested with 4-point textures were excluded from further testing. For the second batch of rats, we decided to admit all the animals to the following experimental phases after a prolonged period of training in the first phase. In fact, we reasoned that, in case some texture statistic was particularly hard to discriminate, imposing a criterion performance in the first phase of the experiment would bias the pool of rats tested with such very difficult statistic towards including only exceptionally proficient animals. This in turn, could lead to an overestimation of rat typical sensitivity to such difficult statistic. On the other hand, the failure of a rat to reach a given criterion performance could be due to intrinsic limitations of its visual apparatus (such as a malfunctioning retina or particularly low acuity). Therefore, to make sure that our result did not depend on including in our analysis some animals of the second batch that did not reach 65% correct discrimination in the first training phase, the perceptual sensitivities were re-estimated after excluding those rats (i.e., after excluding 1 rat from the 2-point, 3 rats from the 3-point and 1 from the 4-point groups). As shown in Figure 4, the resulting sensitivity ranking was unchanged (compare to Fig. 2C) and all pairwise comparisons remained statistically significant (two sample t-test with Holm-Bonferroni correction).

Phase II

In this phase, we introduced progressively lower levels of intensity of each statistic, bringing them gradually closer to the zero-intensity level corresponding to white noise. To this end, we applied an adaptive staircase procedure to update the minimum level of the statistic to be presented to a rat based on its current performance. Briefly, in any given trial, the level of the multipoint correlation in the structured textures was randomly sampled between a minimum level (under the control of the staircase procedure) and a maximum level (fixed at the value used in phase I). Within this range, the sampling was not uniform, but was carried out using a geometric distribution (with the peak at the minimum level), so as to make much more likely for rats to be presented with intensity levels at or close to the minimum. The performance achieved by the rats on the current minimum intensity level was computed every ten trials. If such a performance was higher than 70% correct, the minimum intensity level was decreased by a step of 0.05. By contrast, if the performance was lower than 50%, the minimum intensity level was increased of the same amount.

This procedure allowed the rats to learn to discriminate progressively lower levels of the statistic in a gradual and controlled way. At the end of this phase, the minimum intensity level reached by the animal in the three groups was: 0.21 ± 0.12, 0.2 ± 0.2, 0.70 ± 0.22, and 0.56 ± 0.18 (group average ± SD) for, respectively, 1-, 2-, 3- and 4-point correlations.

Phase III

After the training received in phases I and II, the rats were finally moved to the main test phase, where we measured their sensitivity to the multipoint correlations they were trained on. In each trial of this phase, the stimulus was either white noise or a patterned texture with equal probability. If it was a patterned texture, the level of the statistic was randomly selected from the set {0.02, 0.09, 0.16, . . . , 0.93, 1} (i.e., from 0.02 to 1 in steps of 0.07) with uniform probability. The responses of each rat over this range of intensity levels yielded psychometric curves (see example in Fig. 1B), from which rat sensitivity was measured by fitting the Bayesian ideal observer model described below (Fig. 2A, B).

Phase IV

To verify the sensitivity ranking observed in phase III, we carried out an additional test phase, where each rat was tested on a new statistic, which was different from the one the animal was previously trained and tested on. The two groups of rats that were originally trained with the statistics yielding the highest sensitivity in phase III (i.e., 1- and 2-point correlations; see Fig. 2B) were split in approximately equally-sized subgroups and each of these subgroups was tested with the less discriminable statistics (i.e., 3- and 4-point correlations; leftmost half of Fig. 2C). This allowed assessing that, regardless of the training history, sensitivity to 4-point correlations was slightly but consistently higher than sensitivity to 3-point correlations. For the group of rats originally tested with the 3-point statistic, all the animals were switched to the 4-point (third set of points in Fig. 2C). This allowed comparing the sensitivities to these statistics at the within-subject level (notably, these rats were found to be significantly more sensitive to the 4-point textures than to the 3-point, despite the extensive training they had received with the latter). For the same reason, most of the rats (8/11) of the last group (i.e., the animals originally trained/tested with the 4-point correlations; last set of points in Fig. 2C) were switched to the 3-point statistic, which yielded again the lowest discriminability. A few animals (3/11) were instead tested with the 2-point statistic, thus verifying that the latter was much more discriminable than the 4-point one (again, despite the extensive training the animals of this group had received with the 4-point textures).

Data availability

Experimental data will be made freely and permanently available at (Caramellino et al. 2021).

Ideal observer model

In this section we describe the ideal observer model we used to estimate the sensitivity of the rats to the different textures. The approach is a standard one and is inspired by that in (Fleming, Maloney, and Daw 2013). Because our intention is to use an ideal observer as a model for animal behavior, we will write interchangeably “rat”, “animal” and “ideal observer” in the following.

Preliminaries

The texture discrimination task is a two-alternative forced choice (2AFC) task, where the stimulus can be either a sample of white noise or a sample of textured noise, and the goal of the animal is to correctly report the identity of each stimulus. On any given trial, either

stimulus class can happen with equal probability. The texture class is composed of K discrete, positive values of the texture. In practice, K = 14, and these values are {0.02, 0.09, ..., 0.93, 1}, but we'll use a generic K in the derivations for clarity. The texture statistics are parametrized such that a statistic value of zero corresponds to white noise. Therefore, if we call *s* the true level of the statistic, the task is a parametric discrimination task where the animal has to distinguish $s = 0$ from $s > 0$.

Key assumptions

1. Each trial is independent from those preceding and following it (both for the generated texture and for the animal's behavior);

2. On any given trial, the nominal (true) value of the statistic is some value s. Because the texture has finite size, the empirical value of the statistic in the texture will be somewhat different from *s*. We lump this uncertainty together with that induced by the animal's perceptual process, and we say that any given trial results on the production of a percept *x*, sampled from a truncated Normal distribution centered around the nominal value of the statistic and bounded between $a = -1$ and $b = 1$:

$$p(x|s,\sigma,a,b) = \frac{1}{\sigma}\frac{\phi\left(\frac{x-s}{\sigma}\right)}{\Phi\left(\frac{b-s}{\sigma}\right)-\Phi\left(\frac{a-s}{\sigma}\right)}$$

   where φ (.) is the probability density function of the standard Normal and Φ(.) is its cumulative density function. Setting the bounds to -1 and 1 allows us to account for the fact that the value of a statistic is constrained within this range by construction. We will keep *a* and *b* in some of the expressions below for generality and clarity, and we will substitute their values only at the end.

3. We assume that each rat has a certain prior over the statistic level that we parametrise by the log prior odds:

$$\alpha := \ln\frac{p(s=0)}{p(s>0)}$$

   Where α depends on the rat. More specifically, we assume that each rat assigns a prior probability $p(s = 0) = 1/(1 + e^{-\alpha})$ to the presentation of a noise sample, and a probability of $1/[K(1 + e^{\alpha})]$ to the presentation of a texture coming from any of the K nonzero statistic values. In formulae:

$$p(s) = \frac{\delta_{s,0}}{1+e^{-\alpha}} + \frac{1}{K}\sum_{k=1}^{K}\frac{\delta_{s,s_k}}{1+e^{\alpha}}$$

   where δ is Kronecker's delta, and $s_k > 0$, $k \in \{1, ..., K\}$ are the *K* possible nonzero values of the statistic. Note that this choice of prior matches the distribution actually used in generating the data for the experiment, except that α is a free parameter instead of being fixed at 0.

4. We assume that the true values of α, σ, *a* and *b* are accessible to the decision making process of the rat.

Derivation of the ideal observer

For a particular percept, the ideal observer will evaluate the posterior probability of

noise vs texture given that percept. It will report “noise” if the posterior of noise is higher than the posterior of texture, and “texture” otherwise.

More in detail, for a given percept *x* we can define a decision variable *D* as the log posterior ratio:

$$D(x) := \ln \frac{p(s=0|x)}{p(s>0|x)} = \ln \frac{p(x|s=0)}{p(x|s>0)} + \ln \frac{p(s=0)}{p(s>0)} \qquad (1)$$

With this definition, the rat will report “noise” when $D > 0$ and “texture” otherwise.

By plugging in the likelihood functions and our choice of prior, we get:

$$D(x) = \alpha + \ln \left[\frac{1}{\sigma} \frac{\phi(x/\sigma)}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)} \right] - \ln \left[\frac{1}{K} \sum_k \frac{1}{\sigma} \frac{\phi\left(\frac{x-s_k}{\sigma}\right)}{\Phi\left(\frac{b-s_k}{\sigma}\right) - \Phi\left(\frac{a-s_k}{\sigma}\right)} \right] \qquad (2)$$

Now, remember that given a value of the percept *x*, the decision rule based on *D* is fully deterministic (maximum a posteriori estimate). But on any given trial we don’t know the value of the percept — we only know the nominal value of the statistic. On the other hand, our assumptions above specify the distribution *p(x|s)* for any *s*, so the deterministic mapping *D(x)* means that we can compute the probability of reporting “noise” as

$$p(\text{report noise}|s) = p(D > 0|s) = \int_{x:D(x)>0} p(x|s)\,dx \qquad (3)$$

We note at this point that D(x) is monotonic: indeed,

$$\begin{aligned} \frac{dD(x)}{dx} &= -\frac{x}{\sigma^2} + \\ &- \left[\frac{1}{K} \sum_k \frac{\exp\left[-\frac{(x-s_k)^2}{2\sigma^2}\right]}{\Phi\left(\frac{b-s_k}{\sigma}\right) - \Phi\left(\frac{a-s_k}{\sigma}\right)} \right]^{-1} \frac{1}{K} \sum_k \frac{\exp\left[-\frac{(x-s_k)^2}{2\sigma^2}\right]\left(-\frac{x-s_k}{\sigma^2}\right)}{\Phi\left(\frac{b-s_k}{\sigma}\right) - \Phi\left(\frac{a-s_k}{\sigma}\right)} \\ &= -\left[\frac{1}{K} \sum_k \frac{\exp\left[-\frac{(x-s_k)^2}{2\sigma^2}\right]}{\Phi\left(\frac{b-s_k}{\sigma}\right) - \Phi\left(\frac{a-s_k}{\sigma}\right)} \right]^{-1} \frac{1}{K} \sum_k \frac{\exp\left[-\frac{(x-s_k)^2}{2\sigma^2}\right]\frac{s_k}{\sigma^2}}{\Phi\left(\frac{b-s_k}{\sigma}\right) - \Phi\left(\frac{a-s_k}{\sigma}\right)} \\ &< 0 \text{ for all } x \end{aligned} \qquad (4)$$

where for the last inequality we have used the fact that $a < b$ and so $\Phi((b - s_k)/\sigma) > \Phi((a - s_k)/\sigma)$. This result matches the intuitive expectation that a change in percept in the positive direction (i.e., away from zero) should always make it less likely for the observer to report “noise”.

Because *D(x)* is monotonic, there will be a unique value of x such that *D(x)* = 0, and the integration region *x* : *D(x)* > 0 will simply consist of all values of *x* smaller than that. More formally, if we define:

$$x^* = x^*(\alpha, \sigma) \text{ such that } D(x^*) = 0 \tag{5}$$

We can write:

$$\begin{aligned} p(\text{report noise}|s) &= \int_a^{x^*} \frac{1}{\sigma} \frac{\phi\left(\frac{x-s}{\sigma}\right)}{\Phi\left(\frac{b-s}{\sigma}\right) - \Phi\left(\frac{a-s}{\sigma}\right)} \mathrm{d}x \\ &= \frac{\Phi\left(\frac{x^*(a,\sigma)-s}{\sigma}\right) - \Phi\left(\frac{a-s}{\sigma}\right)}{\Phi\left(\frac{b-s}{\sigma}\right) - \Phi\left(\frac{a-s}{\sigma}\right)} \\ &= \frac{\Phi\left(\frac{x^*(a,\sigma)-s}{\sigma}\right) - \Phi\left(\frac{-1-s}{\sigma}\right)}{\Phi\left(\frac{1-s}{\sigma}\right) - \Phi\left(\frac{-1-s}{\sigma}\right)} \end{aligned} \tag{6}$$

where in the last passage we have substituted *a* = −1 and *b* = 1.

Example: single-level discrimination case

To give an intuitive interpretation of the results above, consider the case where *K* = 1, so the possible values of the statistic are only two, namely 0 and s_1. In this case:

$$\begin{aligned} D^{(1)} &= \alpha + \ln \frac{\Phi\left(\frac{b-s_1}{\sigma}\right) - \Phi\left(\frac{a-s_1}{\sigma}\right)}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)} - \frac{2xs_1 - s_1^2}{2\sigma^2} \\ &= \alpha + \beta - \frac{2xs_1 - s_1^2}{2\sigma^2} \end{aligned}$$

Where

$$\beta := \ln \frac{\Phi\left(\frac{b-s_1}{\sigma}\right) - \Phi\left(\frac{a-s_1}{\sigma}\right)}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)}$$

So that we can write *x** in closed form:

$$x^{*(1)} = D^{(1)-1}(0) = \frac{s_1}{2} + \frac{\sigma^2}{s_1}(\alpha + \beta)$$

which can be read as saying that the decision boundary is halfway between 0 and s_1, plus a term that depends on the prior bias and the effect of the boundaries of the domain of *x* (but involves the sensitivity too, represented by *σ*).

Simplifying things even further, if we remove the domain boundaries (by setting $a \to -\infty$ and $b \to +\infty$), we have that $\beta \to 0$. In this case, by plugging the expression above in Equation 6 we obtain:

$$p(\text{report noise}|s) = \Phi\left[\frac{\sigma}{s_1}\alpha - \frac{(s - s_1/2)}{\sigma}\right] \qquad (7)$$

and therefore we recover a simple cumulative Normal form for the psychometric function. By looking at Equation 7 it is clear how the prior bias α introduces a horizontal shift in the psychometric curve, and σ controls the slope (but also affects the horizontal location when $\alpha \neq 0$).

Fitting the ideal observer model to the experimental data

Independently for each rat, we infer a value of α and σ by maximising the likelihood of the data under the model above. More in detail, for a given rat and a given statistic value s (including 0), we call N_s the number of times the rat reported "noise", and T_s the total number of trials. For a given fixed value of α and σ, under the ideal observer mode the likelihood of N_s will be given by a Binomial probability distribution for T_s trials and probability of success given by the probability of reporting noise in Equation 6:

$$p_s(N_s|\alpha,\sigma) = \binom{T_s}{N_s} p(\text{rep. noise}|s,\alpha,\beta)^{N_s}\left(1 - p(\text{rep. noise}|s,\alpha,\beta)\right)^{T_s - N_s}$$

Assuming that the data for the different values of s is conditionally independent given α and σ, the total log likelihood for the data of the given rat is simply the sum of the log likelihoods for the individual values of N_s:

$$\ln p(\{N_{s_k}\}_{k=1}^{K}|\alpha,\sigma) = \sum_{k=1}^{K} \ln p_{s_k}(N_k|\alpha,\sigma)$$

We find numerically the values of α and σ that maximise this likelihood, using Matlab's mle function with initial condition $\alpha = 0.1$, $\sigma = 0.4$. Note that evaluating the likelihood for any given value of α and σ requires finding $x*$, defined as the zero of Equation 2. We do this numerically by using Matlab's fzero function with initial condition $x = 0$.

Comparing the estimated sensitivity in rats to sensitivity in humans and variability in natural images

To compare quantitatively our sensitivity estimates in rat to those in humans and to the variance of the statistics in natural images reported in (Hermundstad et al. 2014), we computed the degree of correspondence, as defined in Hermundstad et al. 2014, between these sets of numbers. Briefly, define $s_r = (1/\sigma_\beta , 1/\sigma_\theta , 1/\sigma_\alpha)$ as the array containing the rat sensitivities for the three statistics that were tested both here and by Hermundstad et al. 2014 ($\beta = \beta_-$ and $\theta = \theta_\lrcorner$ in the notation used by Hermundstad et al. 2014), s_h as the array containing the corresponding values for humans, and v as that containing the standard deviations of the distribution of the corresponding statistics in natural images. For our comparisons, we use the values of v reported by Hermundstad et al. 2014 for the image analysis defined by the parameters $N = 2$ and $R = 32$ (i.e., the analysis used for the numbers reported in the table in Figure 3C in their paper). The degree of correspondence between any two of these arrays is their cosine dissimilarity:

$$c(\text{rat}, \text{human}) = \frac{s_r \cdot s_h}{\|s_r\| \cdot \|s_h\|}$$
$$c(\text{rat}, \text{images}) = \frac{s_r \cdot v}{\|s_r\| \cdot \|v\|}$$
$$c(\text{human}, \text{images}) = \frac{s_h \cdot v}{\|s_h\| \cdot \|v\|}$$

The degree of correspondence is limited by construction to values between 0 and 1, with 1 indicating a perfect correspondence up to a scaling factor. Hermundstad et al. 2014 report values of 0.987 to 0.999 for *c* (human, images), averaging over all texture coordinates and depending on the details of the analysis.

To assess statistical significance of our values of c, we compare our estimated values with the null probability distribution of the cosine dissimilarity of two unit vectors sampled randomly in the positive orthant of the 3-dimensional Euclidean space. If such vectors are described, in spherical coordinates, as

$$v_1 = (\theta_1, \phi_1) \quad , \quad v_2 = (\theta_2, \phi_2)$$

with $0 \leq \theta 1, \theta 2, \varphi_1, \varphi_2 \leq \pi/2$, the cosine of the angle they form with each other is:

$$c_{\text{null}} = v_1 \cdot v_2 = \cos(\theta_1)\cos(\theta_2) + \cos(\phi_1 - \phi_2)\sin(\theta_1)sin(\theta_2)$$

The p-values reported in the text for *c*(rat, human) and *c*(rat, images) are computed by sampling 10^7 values of c_{null} , and assessing the fraction of samples with values larger than the empirical estimates.

References

Alemi-Neissi, A., F. B. Rosselli, and D. Zoccolan. 2013. “Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition.” Journal of Neuroscience 33, no. 14: 5939–5956. doi:10.1523/jneurosci.3629-12.2013.

Atick, Joseph J., and A. Norman Redlich. 1990. “Towards a Theory of Early Visual Processing.” Neural Computation 2, no. 3: 308–320. doi:10.1162/neco.1990.2.3.308.

Caramellino, Riccardo, Eugenio Piasini, Andrea Buccellato, Anna Carboncino, Vijay Balasubra- manian, and Davide Zoccolan. 2021. Data from "Rat sensitivity to multipoint statistics is predicted by efficient coding of natural scenes". doi:10.5281/zenodo.4762567.

Carlson, Nicole L., Vivienne L. Ming, and Michael Robert DeWeese. 2012. “Sparse Codes for Speech Predict Spectrotemporal Receptive Fields in the Inferior Colliculus.” Ed. by Tim Behrens. PLoS Computational Biology 8, no. 7: e1002594. doi:10.1371/journal.pcbi. 1002594.

DiCarlo, James J., Davide Zoccolan, and Nicole C. Rust. 2012. “How Does the Brain Solve Visual Object Recognition?” Neuron 73, no. 3: 415–434. doi:10.1016/j.neuron.2012. 01.010.

Djurdjevic, Vladimir, Alessio Ansuini, Daniele Bertolini, Jakob H. Macke, and Davide Zoccolan. 2018. “Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy.” Current biology 28 (7): 1005–1015.e5. ISSN: 1879-0445. doi:10.1016/j.cub.2018.02.037.

Fairhall, Adrienne L., Geoffrey D. Lewen, William Bialek, and Robert R. de Ruyter van Steveninck. 2001. “Efficiency and ambiguity in an adaptive neural code.” Nature 412, no. 6849: 787–792. doi:10.1038/35090500.

Fleming, Stephen M., Laurence T. Maloney, and Nathaniel D. Daw. 2013. “The Irrationality of Categorical Perception.” Journal of Neuroscience 33, no. 49 : 19060–19070. doi:10.1523/ jneurosci.1263-13.2013.

Geisler, Wilson S. 2011. “Contributions of ideal observer theory to vision research.” Vision Research 51, no. 7: 771–781. doi:10.1016/j.visres.2010.09.027.

Glickfeld, Lindsey L., and Shawn R. Olsen. 2017. “Higher-Order Areas of the Mouse Visual Cortex.” Annual Review of Vision Science 3, no. 1: 251–273. doi:10.1146/annurev-vision-102016-061331.

Glickfeld, Lindsey L, R Clay Reid, and Mark L Andermann. 2014. “A mouse model of higher visual cortical function.” Current Opinion in Neurobiology 24: 28–33. doi:10.1016/j. conb.2013.08.009.

Hermundstad, Ann M, John J Briguglio, Mary M Conte, Jonathan D Victor, Vijay Balasubrama- nian, and Gašper Tkacˇik. 2014. “Variance predicts salience in central sensory processing.” eLife 3. doi:10.7554/elife.03722.

Huberman, Andrew D., and Cristopher M. Niell. 2011. “What can mice tell us about how vision works?” Trends in Neurosciences 34, no. 9: 464–473. doi:10.1016/j.tins.2011.07. 002.

Hunt, Jonathan J., Peter Dayan, and Geoffrey J. Goodhill. 2013. “Sparse Coding Can Predict Primary Visual Cortex Receptive Field Changes Induced by Abnormal Visual Input.” Ed. by Matthias Bethge. PLoS Computational Biology 9, no. 5: e1003005. doi:10.1371/ journal.pcbi.1003005.

Kaliukhovich, Dzmitry A., and Hans Op de Beeck. 2018. “Hierarchical stimulus processing in rodent primary and lateral visual cortex as assessed through neuronal selectivity and repetition suppression.” Journal of Neurophysiology 120, no. 3: 926–941. doi:10.1152/ jn.00673.2017.

Katzner, Steffen, and Sarah Weigelt. 2013. “Visual cortical networks: of mice and men.” Current Opinion in Neurobiology 23, no. 2: 202–206. doi:10.1016/j.conb.2013.01.019.

Keller, Jason, Hans Strasburger, Daniel T Cerutti, and Bernhard A Sabel. 2000. “Assessing spatial vision — automated measurement of the contrast-sensitivity function in the hooded rat.” Journal of Neuroscience Methods 97, no. 2: 103–110. doi:10.1016/s0165- 0270(00)00173-4.

Keyser, Roxane De, Christophe Bossens, Jonas Kubilius, and Hans P. Op de Beeck. 2015. “Cue-invariant shape recognition in rats as tested with second-order contours.” Journal of Vision 15, no. 15: 14. doi:10.1167/15.15.14.

Kourtzi, Zoe, and Charles E. Connor. 2011. “Neural Representations for Object Perception: Structure, Category, and Adaptive Coding.” Annual Review of Neuroscience 34, no. 1: 45–67. doi:10.1146/annurev-neuro-060909-153218.

Laughlin, S. 1981. “A simple coding procedure enhances a neuron’s information capacity.” Zeitschrift fur Naturforschung. Section C, Biosciences 36 (9-10): 910–912. ppublish. ISSN: 0341-0382.

Lehky, Sidney R, and Keiji Tanaka. 2016. “Neural representation for object recognition in inferotemporal cortex.” Current Opinion in Neurobiology 37: 23–35. doi:10.1016/j. conb.2015.12.001.

Liu, Kefei, and Haishan Yao. 2014. “Contrast-dependent OFF-dominance in cat primary visual cortex facilitates discrimination of stimuli with natural contrast statistics.” European Journal of Neuroscience 39, no. 12: 2060–2070. doi:10.1111/ejn.12567.

Matteucci, Giulio, Rosilari Bellacosa Marotti, Margherita Riggi, Federica B. Rosselli, and Davide Zoccolan. 2019. “Nonlinear processing of shape information in rat lateral extrastriate cortex.” The Journal of Neuroscience: 1938–18. doi:10.1523/jneurosci.1938-18. 2018.

Matteucci, Giulio, and Davide Zoccolan. 2020. “Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells.” Science Advances 6, no. 22: eaba3742. doi:10.1126/sciadv.aba3742.

Młynarski, Wiktor F., and Ann M. Hermundstad. 2021. “Efficient and adaptive sensory codes.” Nature Neuroscience. doi:10.1038/s41593-021-00846-0.

Nassi, Jonathan J., and Edward M. Callaway. 2009. “Parallel processing strategies of the primate visual system.” Nature Reviews Neuroscience 10, no. 5: 360–372. doi:10.1038/nrn2619.

Niell, Cristopher M., and Massimo Scanziani. 2021. “How Cortical Circuits Implement Cortical Computations: Mouse Visual Cortex as a Model.” Annual Review of Neuroscience 44, no. 1. doi:10.1146/annurev-neuro-102320-085825.

Olshausen, Bruno A., and David J. Field. 1996. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images.” Nature 381, no. 6583: 607–609. doi:10.1038/381607a0.

Orban, Guy A. 2008. “Higher Order Visual Processing in Macaque Extrastriate Cortex.” Physio- logical Reviews 88, no. 1: 59–89. doi:10.1152/physrev.00008.2007.

Piasini, Eugenio. 2021. metex — Maximum Entropy TEXtures. Version 1.0.0. doi:10.5281/ zenodo.4753198. https://pypi.org/project/metex.

Piasini, Eugenio, Liviu Soltuzu, Paolo Muratore, Riccardo Caramellino, Kasper Vinken, Hans Op de Beeck, Vijay Balasubramanian, and Davide Zoccolan. 2021. “Temporal

stability of stimulus representation increases along rodent visual cortical hierarchies." In press, Nature Communications.

Pitkow, Xaq, and Markus Meister. 2012. "Decorrelation and efficient coding by retinal ganglion cells." Nature Neuroscience 15, no. 4: 628–635. doi:10.1038/nn.3064.

Prusky, G. T., P. W. West, and R. M. Douglas. 2000. "Behavioral assessment of visual acuity in mice and rats." Vision research 40 (16): 2201–2209. ISSN: 0042-6989. doi:10.1016/s0042- 6989(00)00081-x.

Prusky, Glen T., K. Troy Harker, Robert M. Douglas, and Ian Q. Whishaw. 2002. "Variation in visual acuity within pigmented, and between pigmented and albino rat strains." Behavioural brain research 136 (2): 339–348. ISSN: 0166-4328. doi:10.1016/s0166-4328(02)00126-2.

Purpura, K. P., J. D. Victor, and E. Katz. 1994. "Striate cortex extracts higher-order spatial correlations from visual textures." Proceedings of the National Academy of Sciences 91, no. 18: 8482–8486. doi:10.1073/pnas.91.18.8482.

Reinagel, P. 2015. "Using rats for vision research." Neuroscience 296: 75–79. doi:10.1016/ j.neuroscience.2014.12.025.

Schwartz, Odelia, and Eero P. Simoncelli. 2001. "Natural signal statistics and sensory gain control." Nature Neuroscience 4, no. 8: 819–825. doi:10.1038/90526.

Smith, Evan C., and Michael S. Lewicki. 2006. "Efficient auditory coding." Nature 439, no. 7079: 978–982. doi:10.1038/nature04485.

Sterling, Peter, and Simon Laughlin. 2015. Principles of Neural Design. MIT Press.

Tafazoli, Sina, Houman Safaai, Gioia De Franceschi, Federica Bianca Rosselli, Walter Vanzella, Margherita Riggi, Federica Buffolo, Stefano Panzeri, and Davide Zoccolan. 2017. "Emer- gence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex." eLife 6. doi:10.7554/elife.22794.

Tesileanu, Tiberiu, Simona Cocco, Rémi Monasson, and Vijay Balasubramanian. 2019. "Adaptation of olfactory receptor abundances for efficient coding." eLife 8. doi:10.7554/elife. 39279.

Tesileanu, Tiberiu, Mary M Conte, John J Briguglio, Ann M Hermundstad, Jonathan D Victor, and Vijay Balasubramanian. 2020. "Efficient coding of natural scene statistics predicts discrimination thresholds for grayscale textures." eLife 9. doi:10.7554/elife.54347.

Tkacik, G., J. S. Prentice, J. D. Victor, and V. Balasubramanian. 2010. "Local statistics in natural scenes predict the saliency of synthetic textures." Proceedings of the National Academy of Sciences 107, no. 42: 18149–18154. doi:10.1073/pnas.0914916107.

Vascon, Sebastiano, Ylenia Parin, Eis Annavini, Mattia D'Andola, Davide Zoccolan, and Marcello Pelillo. 2018. "Characterization of Visual Object Representations in Rat Primary Visual Cortex." In Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

Vermaercke, Ben, Florian J. Gerich, Ellen Ytebrouck, Lutgarde Arckens, Hans P. Op de Beeck, and Gert Van den Bergh. 2014. “Functional specialization in rat occipital and temporal visual cortex.” Journal of Neurophysiology 112, no. 8: 1963–1983. doi:10.1152/jn.00737. 2013.

Vermaercke, Ben, and Hans P. Op de Beeck. 2012. “A Multivariate Approach Reveals the Behavioral Templates Underlying Visual Discrimination in Rats.” Current Biology 22, no. 1: 50–55. doi:10.1016/j.cub.2011.11.041.

Victor, Jonathan D., and Mary M. Conte. 2012. “Local image statistics: maximum-entropy constructions and perceptual salience.” Journal of the Optical Society of America. A, Optics, image science, and vision 29 (7): 1313–1345. ISSN: 1520-8532. doi:10.1364/JOSAA.29.001313.

Vinken, Kasper, Rufin Vogels, and Hans Op de Beeck. 2017. “Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement.” Current Biology 27, no. 6: 914–919. doi:10.1016/j.cub.2017.02.024.

White, Leonard E., and David Fitzpatrick. 2007. “Vision and Cortical Map Development.” Neuron 56, no. 2: 327–338. doi:10.1016/j.neuron.2007.10.011.

Williams, Brice, Joseph Del Rosario, Tomaso Muzzu, Kayla Peelman, Stefano Coletta, Edyta K. Bichler, Anderson Speed, Lisa Meyer-Baese, Aman B. Saleem, and Bilal Haider. 2021. “Spatial modulation of dark versus bright stimulus responses in the mouse visual system.” Current Biology 31, no. 18: 4172–4179.e6. doi:10.1016/j.cub.2021.06.094.

Xing, D., C.-I. Yeh, and R. M. Shapley. 2010. “Generation of Black-Dominant Responses in V1 Cortex.” Journal of Neuroscience 30, no. 40: 13504–13512. doi:10.1523/jneurosci. 2473-10.2010.

Yeh, C.-I, D. Xing, and R. M. Shapley. 2009. “"Black" Responses Dominate Macaque Primary Visual Cortex V1.” Journal of Neuroscience 29, no. 38: 11753–11760. doi:10.1523/ jneurosci.1991-09.2009.

Yu, Yunguo, Anita M Schmid, and Jonathan D Victor. 2015. “Visual processing of informative multipoint correlations arises primarily in V2.” eLife 4. doi:10.7554/elife.06604.

Zoccolan, Davide. 2015. “Invariant visual object recognition and shape processing in rats.” Behavioural brain research 285: 10–33. ISSN: 1872-7549. doi:10.1016/j.bbr.2014. 12.053.

Chapter III

Coding of multipoint statistics in rat visual cortex

Abstract

Efficient processing of sensory data requires adapting the neuronal encoding strategy to the statistics of natural stimuli. One area where this concept has been applied is the perceptual saliency of textures: multiple works have shown that humans are most sensitive to textures containing multipoint correlations that vary the most across natural images. The neuronal mechanisms underlying such adaptation are not understood but investigating them in depth requires invasive experiments that are impossible in humans. I have shown in the previous chapter that rats can be trained to discriminate between binary textures containing unstructured random noise and textures defined by single- to 4-point correlations among nearby pixels. We observed a decrease in sensitivity from 2- to 4-point correlations and a further decrease from 4- to 3-point, fully reproducing results obtained on humans. Building on this result, we performed extracellular recordings of neural activity from primary visual cortex (V1) and the extrastriate area LL of anaesthetized rats passively viewing textures movies containing multi-point correlations up to 4 points or unstructured noise. We established that both V1 and LL populations of units accurately encode 1- and 2-point correlations. Regarding 3- and 4-point correlations, instead, we found poor discriminability from noise in both areas with slightly better performances in V1. We were also able to demonstrate that the representation of the correlations was more invariant in LL, providing further proof of the hierarchical organization of rat visual cortex and the similarity of the progression of rat lateral extrastriate areas with the primate ventral pathway.

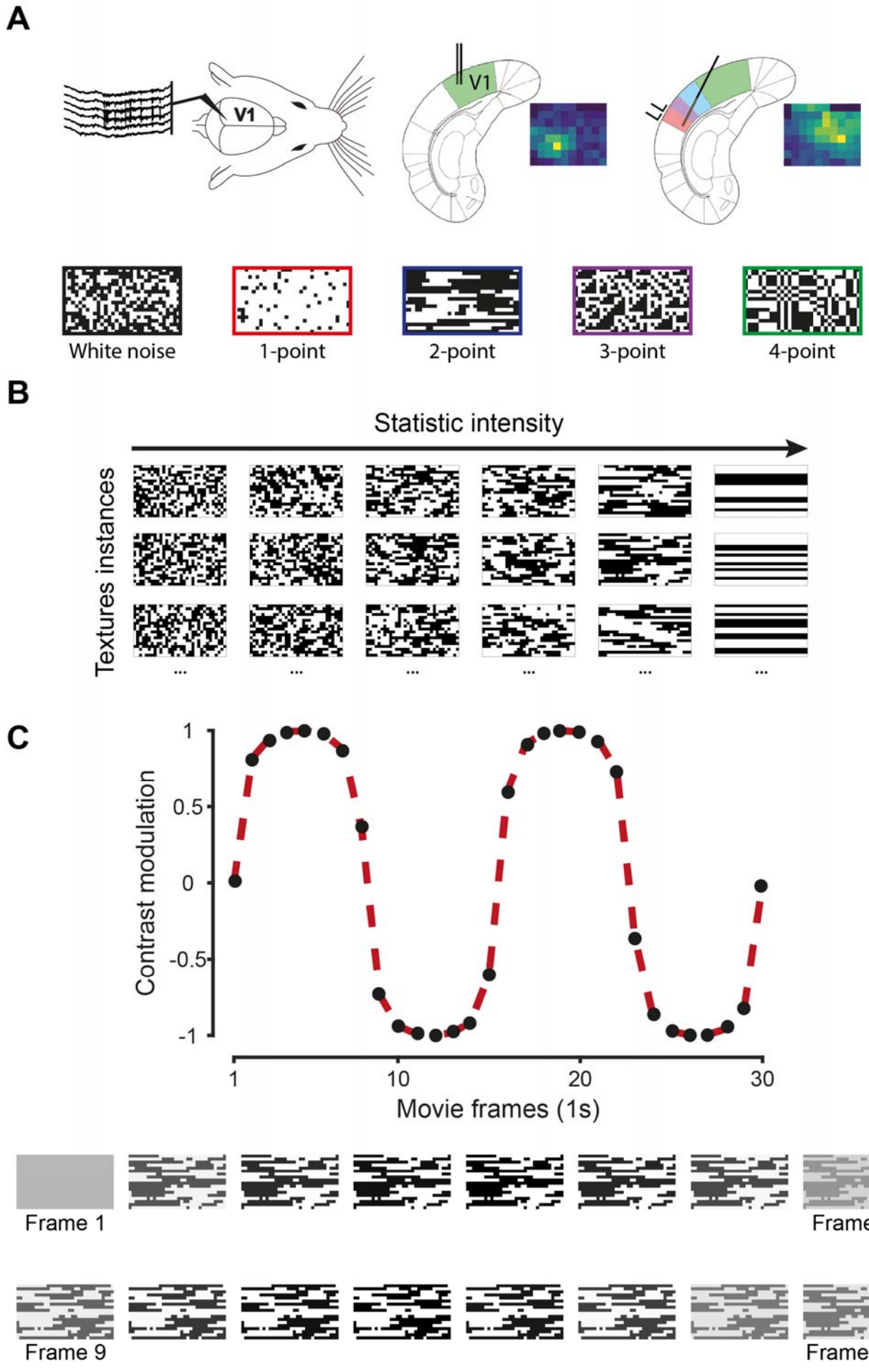
Introduction

Efficient processing of sensory data in the brain is widely believed to be based on adaptation to the statistical features present in natural scenes. In the visual domain the efficient coding principle (Barlow, 1961; Attneave, 1954) has historically been able to explain multiple computations that happen in the periphery such as the encoding of luminance intensity distributions or lateral inhibition in the retina (Laughlin, 1981; Srinivasan, 1982). Recently, it was shown that this normative principle can be extended to the cortical stages of visual processing, where it can account for the differential allocation of resources needed to represent higher order statistics contained in natural images (Hermundstad et al., 2014; Tesileanu et al., 2020; Tkačik et al., 2010). It was demonstrated that human sensitivity to synthetic visual textures defined by local multipoint correlations depends on the variability of such correlations in natural scenes. As I have explained in the previous chapter, we further extended this result, by demonstrating the same phenomenon in rats (Caramellino et al., 2021). This was achieved by training rats to discriminate between binary textures containing unstructured random noise and textures defined by single- to four-point correlations among nearby pixels. We observed a decrease in sensitivity from 2- to 4-point correlations and a further decrease from 4- to 3-point, fully

reproducing the results obtained in humans. This indicates that: 1) neuronal populations in rat visual cortex exist that are able to encode visual textures defined by multipoint correlations; and 2) the discriminatory power of such representations must conform the ranking of perceptual sensitivities overserved at the behavioral level. This, in turn, raises the questions of where, along the rat visual processing hierarchy, such neuronal populations emerge.

Previous results from single cell recordings in anaesthetized macaque monkeys provided partially contrasting evidence. An early study on the subject found units selective for 4-point correlations in V1 (Purpura et al., 1994). Later work suggested, instead, that only in V2 neurons responses to textures containing 3- and 4-point correlations was significantly different from the response to noise textures (Yu et al., 2015). Similar findings were obtained in studies using naturalistic textures containing different set of high order statistics: these works, despite the difference in the features examined, showed that encoding of high order statistics starts in V2 and increases along the ventral stream up to V4 (freeman et al., 2013; Ziemba et al., 2016, Okazawa et al., 2017). Population recording from monkey's ventral stream during the presentation of naturalistic textures, instead, displayed a higher-than-expected sensitivity for higher order statistics already in V1, probably due to feedback projections from higher level areas (Ziemba et al., 2019). Taken together, these results suggest that the processing of image correlations in the monkey's ventral stream is still not perfectly understood.

Monkey have always been the standard model to study advanced properties of vision, thanks to the fact that their visual system is capable of complex computations that closely match human ones (Orban et al., 2004; Rajalingham et al., 2015; DiCarlo et al., 2012; Lehky and Tanaka, 2016; N. C. Rust & DiCarlo, 2010). Still, the primate brain is very complex to study (Felleman & Van Essen, 1991) and many invasive techniques are often impractical. Rodents, instead, are an ideal model to study the adaptation to natural statistics found in the environment, as they allow a wide array of causal manipulations – e.g., optogenetic or controlled rearing. Rats in particular are able to perform complex computations such as the advanced shape processing required for invariant object recognition (Alemi-Neissi et al., 2013; De Keyser et al., 2015; Djurdjevic et al., 2018; Vermaercke and Op de Beeck, 2012). Furthermore, rat extrastriate areas have been shown to possess some of the key properties found in the primate ventral pathway (Matteucci et al., 2019; Piasini et al., 2021; Tafazoli et al., 2017; Vermaercke et al., 2014; Vinken et al., 2017), like the ability to support invariant visual object recognition. Finally, it



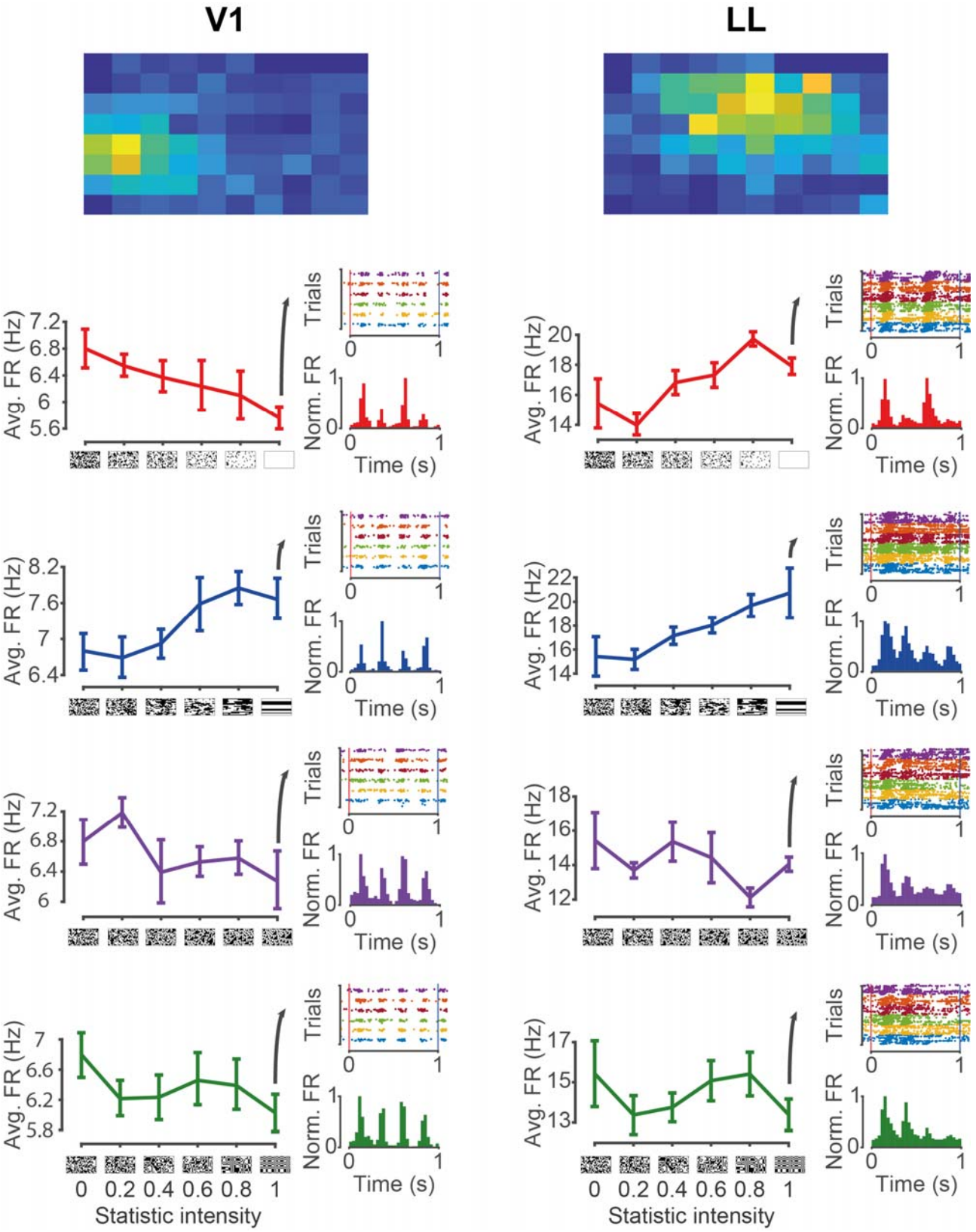
has already been demonstrated that it's possible to rear rats in altered visual environments

Figure 1. Experiment design. (**A**) Electrophysiological recordings of neuronal activity were performed, under anaesthesia, in two areas of rat's visual cortex: V1 (left), and the extrastriate area LL (right). Recording were performed using a 64 channels 2-shanks electrode. The array was inserted perpendicularly to the brain surface when aiming at V1, and with an angle of 20° relative to the perpendicular line in sessions targeting LL. During each recording animals were presented textures movies containing white noise (black border) or one of 4 multipoint correlations: 1-point (red border), 2-point (blue border), 3-point (purple border), 4-point (green border). (**B**) Specifically, in the protocol we included 6 noise textures, and 6 different texture instances for each of 5 levels of intensity (0.2, 0.4, 0.6, 0.8, 1) of the correlations (top). Each texture was shown 30 times (trials) during the protocol. (**C**) During the presentation, textures underwent a contrast reversal cycle modulated by a 2 Hz smoothed square wave (bottom, first cycle). The resulting texture movie lasted 1s and was presented at 30 Hz (30 frames).

to test the impact of adaptation to visual input statistic on the development of the visual system (Matteucci & Zoccolan, 2020).

To determine the neural correlates of the sensitivity to multipoint correlations in rats, we performed extracellular population recordings of neural activity from rat primary visual cortex (V1) and from extrastriate area LL, which is considered one of the highest stages of the rat analogue of the primate ventral stream. Naïve rats passively viewed sequences of movies, lasting 1 s each, in which a given texture underwent contrast reversal. We then analysed the responses of neuronal populations in both brain areas to these stimuli, in order to ascertain the extent to which the resulting neuronal representations supported discrimination of textures from white noise. Our results, obtained through a decoding analysis comparing responses to textures defined by one intensity level of a correlation against noise, showed high classification performances for 1- and 2-point correlation in both areas, while discriminability of 3- and 4-point correlations from noise was present, to a much smaller extent, only in V1. Still, we were able to confirm that LL has more invariant representation of the statistics compared to V1, providing further proof of the hierarchical organization of rats' putative ventral stream.

Results



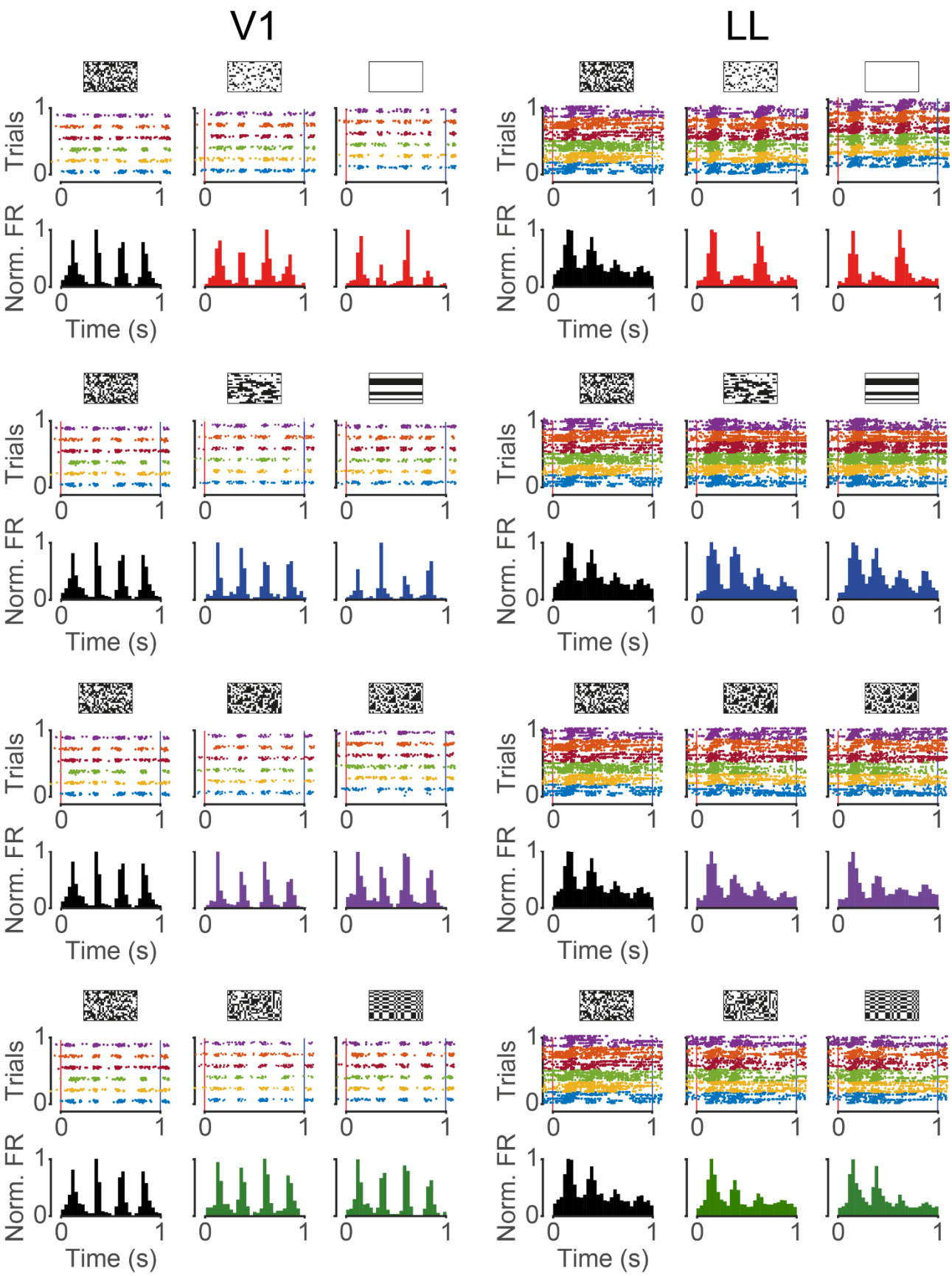
We performed multi-channel extracellular recordings of neuronal activity in anaesthetized rats during the presentation of visual textures defined by multipoint correlations ranging from 1 to 4 points (Figure 1A), i.e., the same used for the behavioral study of the previous chapter. A total of 15 recording sessions was carried out from two areas of rat's visual cortex: the

Figure 2. Example of recorded units in V1 and LL. A representative V1 unit (left) is compared to a unit recorded in LL (right). We can see that the two areas differ in receptive field sizes, with the LL unit possessing a larger receptive field than the unit from V1. Below the receptive field maps, we have the tuning curves for the four statistics (from top to bottom: 1-, 2-, 3-, 4-point correlations). These are obtained by computing, for each intensity level, the average across trials and instances of the total number of spikes recorded during the entire duration of a texture movie. Working with movies the total firing rate doesn't take into account the differences in the response pattern produced by the contrast reversal procedure. These can be appreciated by looking at the raster plots (presented here for, the highest intensity level, on the top right of each tuning curve). Raster plots depict the precise time course of the response in the trials of the 6 textures sharing the same statistic intensity (displayed in different colors). The information is then summarized by the cumulative (across texture instances), normalized psth (presented below each raster). To build a psth we binarized our presentation window in 33 ms time bins, and, then, computed the average across trials of the responses in each bin.

primary visual cortex V1 (5 sessions – 431 multiunits), and the laterolateral extrastriate area LL (10 sessions – 288 multiunits), which sits at highest stages of the rat analogue of the ventral stream.

The main stimulus set consisted of binary maximum entropy textures, defined by correlations from 1- to 4-point, and white noise textures. The latter were generated by assigning each pixel to black or white with equal probability, resulting in a complete lack of spatial correlations. Structured textures, instead, were built using a published software library (Piasini, 2021) that implements the method developed in (Victor and Conte, 2012). Maximum entropy textures allow a fine tuning of their statistical structure, leading to images that are as close as possible to with noise, except in the intensity of the desired multipoint correlation. Importantly, for every combination of statistic and intensity level (e.g., 2-point statistic at 0.4 intensity), we produced 6 different instances of textures with those properties (the same was done for the white noise patterns; Figure 1B). During presentation (1 s) both structured and unstructured textures underwent contrast reversal to maximise the response evoked by the stimulus while reducing the effect of adaptation (See Figure 1C). This procedure is convenient because it does not introduce any of the confounds that could emerge, due to the orientation of some patterns, if we used, for the same goal, drifting images.

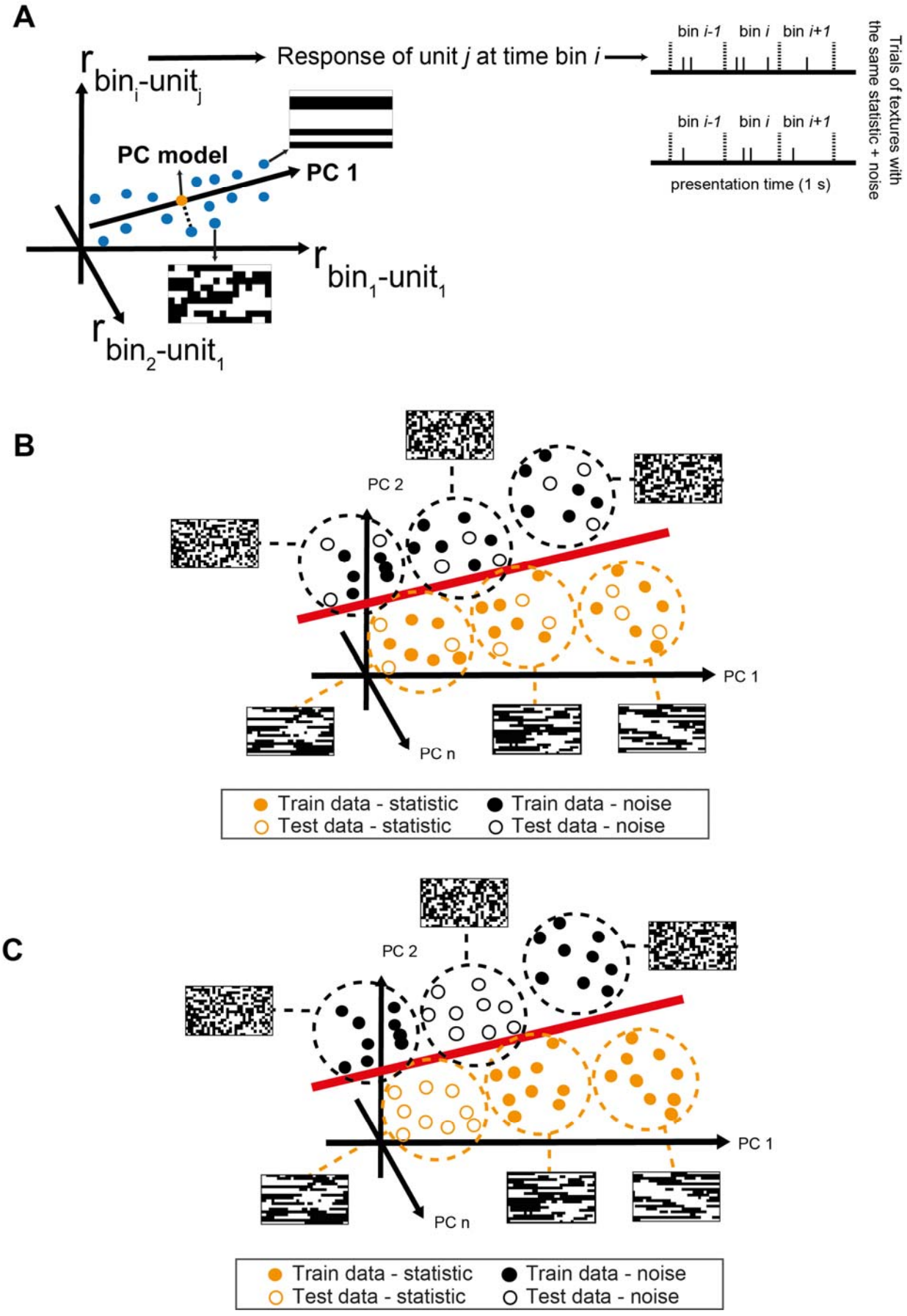
Figure 2 shows two example units, one recorded in V1 and the other recorded in the extrastriate area LL. For each unit we built tuning curves by measuring the average response during the entire presentation of textures containing the different levels of intensity of the four statistics. Across the two areas we both found units that displayed an increase in their firing rate for growing intensity of certain statistics, and units that showed a reduction in the firing rate for higher intensities of one or more statistics. Furthermore, some units displayed more complex and harder to interpret response patterns, that weren't fully understandable from the average firing rate only. Indeed, working with dynamic stimuli, the average firing rate didn't provide complete information on the sensitivity to the statistics but was a useful tool to perform an initial qualitative exploration of the recorded units. A more in depth analysis of the firing patterns in time of the two units presented in Figure 2 can be seen in Figure 3: the responses to the four statistics and noise differ not only in intensity, but, in some cases, also in the modulation of the firing rate induced by the contrast reversal.



These differences may be relevant in understanding how multipoint correlations are processed in rat visual cortex, so we decided to take the time dimension into account in our analysis.

Figure 3. Time course of single units' responses. The activity, during stimulus presentation, of the units introduced in Figure 2 is displayed here to highlight the effect of the contrast modulation on the different statistics and their intensities. The raster plots and the corresponding PSTHs are presented for white noise textures and 2 levels of intensity of each of statistics (0.6 and 1). Responses are modulated by the contrast reversal procedure, but the period and intensity of this modulation depends, in some cases, on the type of multipoint correlation and its intensity. The two areas differ substantially in term of time course of the responses. In V1, responses to noise and all statistics, except 1-point correlation, are modulated by each cycle of the contrast reversal. In LL the modulation with the second cycle of the contrast reversal is only present for certain statistics – e.g. high intensity level of 2-point correlations.

To understand where the different multipoint correlations were processed, we assessed the ability of subpopulations of units, recorded from V1 or LL, to distinguish the different types of structured textures from white noise. To this aim, for each area, we randomly selected 10 sub-populations of 100 units. We then considered each presentation (trial) of a given stimulus as a vector whose components were the spike counts of every unit in each of the 33ms time bins corresponding to the full presentation of the stimulus (30 time bins). So, our response vectors had a dimensionality equal to the product between the number of time bins and the number of units. Given that this dimensionality far exceeds the number of total trials we had, we decided to apply PCA (Pearson, 1901) on the responses to all the presentations of textures with the same statistic (independently from the intensity and including noise). This procedure was aimed at reducing the number of redundant dimensions and the noise in the representational space (Figure 4A). We then fed to an SVM classifier the dimensionality reduced response vectors for a single intensity of a statistic and for the white noise, and we train the classifier in a binary task (statistic vs. noise). We run the classification procedure in two different ways: "without invariance" and "with invariance". In the analysis "without invariance" we performed a random division of the whole dataset (i.e., the population response vectors associated to every presentation of each texture instance) in three folds, then trained on two of them and used the third one to compute a cross-validated classification performance (Figure 4B). After repeating this process for every combination of the three folds, we computed the average of the performances obtained on each test set. In the analysis "with invariance", instead, we split the data in such a way that the classifier was trained using only the population vectors associated to the presentation of four of the six instances of texture and noise patterns, leaving the population vectors associated to the remaining two in the test set (to compute the cross-validated classification performance; Figure 4C). As before, the decoding was repeated for all the possible allocation of the textures in the training and test set, and the average performance was taken. The key difference between the two analyses was that, in the first case, the test set could include population vectors associated to textures' instances that were also present in the training set (although with response collected in different trials). In the invariant case instead, the test set was made by population vectors associated to textures' instances that were never "seen before" by the classifier (i.e., were not part of the training set). Therefore, succeeding in the invariant task required the underlying neuronal representation to encode the actual statistical content of the textures, independently from the specific appearance of the texture instances being processed.



We repeated both decoding analyses using different numbers of principal

Figure 4. Population analysis structure. (**A**) Units' responses to every trial of all the textures containing one statistic and white noise were projected in a lower dimensionality space using PCA. Each response vector corresponding to a stimulus trial had a number of components equal to the product between the number of units in the sub population (100) and the number of time bins (30), due to the need of taking into account the temporal course of the response. Through PCA the dimensionality was reduced by projecting the response vectors in the space defined by the first N principal components, with N = 2, 5, 10, 30, 100, 500. (**B**) The decoding analysis "without invariance" was performed by randomly dividing the transformed response vectors in 3 folds and using 2 of them as training (filled circles) set for the SVM classifier and one as test set (empty circles). The SVM classifier had to distinguish between trials coming from textures containing one specific intensity level of a statistic and white noise. The classification process was repeated considering every possible assignment of the 3 folds to training or test. This type of cross validation procedure meant that each fold could contain trials of any texture, so training and test both contained information on almost every stimulus. (**C**) The decoding analysis "with invariance" differed from the previous one because we manually determined the test set (empty circles) in order for it to contain all the trials belonging to 2 of the 6 textures belonging to each class (noise vs. statistic). The classification procedure was repeated considering all the possible combinations of the textures in training and test set. Differently from the analysis "without invariance", here, during the testing phase, we measured the ability of the classification model to generalize to trials of textures it didn't receive any information about during training.

components (PCs = 2, 5, 10, 30, 100, 500 and the control without applying PCA). The results for each statistic and brain area obtained from these decoding procedures can be seen in Figure 5A for the analysis without invariance, and in Figure 5B for the analysis with invariance. In these figures we reported, for every number of principal components, the average classification performance yielded by the 10 sampled sub-population of units, for each level of a statistic.

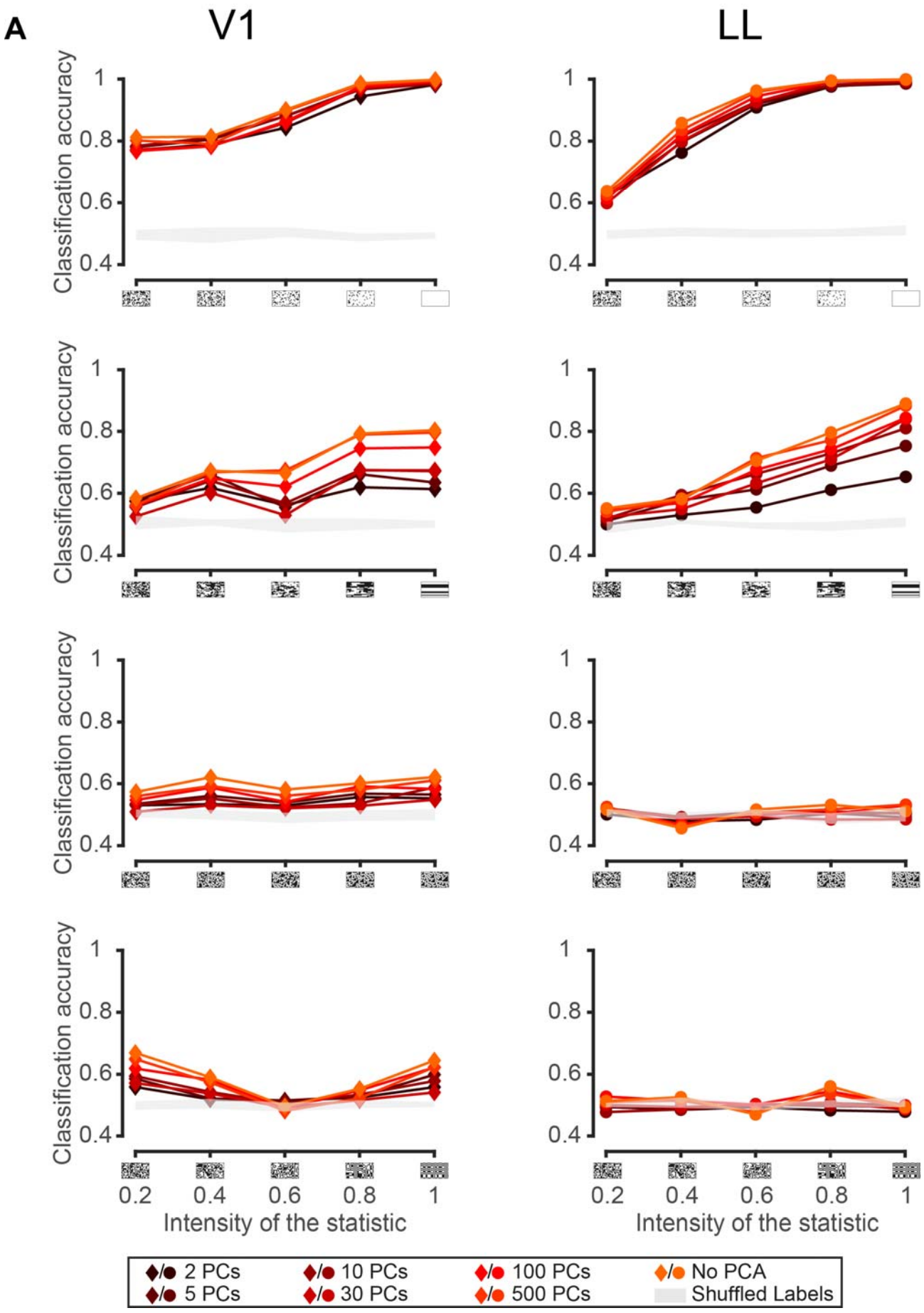
Considering the analysis without invariance first, we found that both areas were able to correctly discriminate from noise textures defined by 1-point correlations (Figure 5A, first row), with V1 populations being able to reliably classify even low intensities of the statistic independently from the number of PCs considered for the dimensionality reduction (average performance ~80% for all PCs values). This suggests that the luminance information that defines 1-point texture statistic is reliably encoded even in small neuronal populations, with V1 outperforming LL, as expected given its largest sensitivity to luminance (Tafazoli et al, eLife, 2017). Contrary to 1-point correlations, discriminability of the 2-point statistic (Figure 5A, second row) was influenced by the number of PCs in both areas. In V1 we found that for small numbers of PCs, the classification accuracy did not sensibly increase as a function of the intensity of the statistic, while a growth of performance for higher intensities was observed when a larger number of PCs was considered. By contrast, in LL, the classification accuracy linearly increased as a function of the intensity of the statistic for any number of PCs considered, but the slope was considerably smaller when using only 2 PCs. For 3- and 4-point correlations (Figure 5A, third and fourth rows), we found that average performances in V1 were above chance, at least for some intensities and when a large number of PCs was considered, but there was not a clear increase of the accuracy along the statistic intensity axis (chance performance was determined by running the classifier on population vectors with shuffled labels between structured and unstructured textures; see gray shading in

Figure 5A and 5B). In particular, in the case of 4-point correlations we found that the discriminability from noise peaked for the lowest and highest intensities. In LL, instead, we found that the classification accuracy for 3- and 4-point correlations was at chance for any number of PCs and intensity of the statistics.

The analysis with invariance (Figure 5B) yielded accuracy trends very similar to those just described, but with worse performances, especially in V1. For instance, the classification accuracy on 3- and 4-point correlations obtained for V1 was at chance for most statistic intensities and PCs values. The classification accuracy on 2-point correlations was also visibly impaired in V1, while the changes were less pronounced in LL.

To better understand the differences between the two areas, both in term of encoding of the statistics and invariance, we decided to compare them after averaging the classification accuracies obtained across the 5 levels of intensity (i.e., the five data points in each curve of Figures 5A and 5B), so as to have a single, summary performance value for each statistic and each number of PCs considered (Figure 6). In the case of 1-point correlations, both V1 and LL yielded high average performances (>80%) for any number of PCs, although V1 was superior to LL when invariance to texture instance was not required (compare solid curves). However, V1 performance slightly dropped when invariance was required (compare yellow and red curves), while no reduction was observed in LL (compare blue and cyan curves). In the case of 2-point correlations, decoding accuracy was higher in LL than V1 for any number of PCs greater than 2. Furthermore, V1 average accuracy displayed a considerable drop when invariance was required, while the performance afforded by LL populations remained virtually unchanged. In the case of 3- and 4-point correlations, average performances in V1 were higher than in LL for any number of PCs. Results from the invariant analysis on these correlations displayed a decrease in classification accuracy in V1, with performances close to chance level.

Noticing how different in the two areas was the robustness of the encoding of texture statistics to variation in texture appearance, we directly compared the differences between the average classification accuracies yielded by the analysis without and with invariance in V1 and LL (Figure 7, red vs. blue bars). Across all statistics, and most of the numbers of PCs considered, there was a significant decrease in average decoding accuracy for V1 when the analysis required invariance (one-sample t-test). By contrast, LL performances changed little when invariance was required, and in those cases where a significant decrease was observed, e.g. for 2-point correlations with 500 PCs, it was considerably lower than in V1. This result confirms the initial hypothesis that LL units are more invariant in their representation of textures defined by different multipoint correlations and, as such, better suited to support rat texture discrimination behavior reported in the previous chapter.


A
V1
LL
Classification accuracy
1
0.8
0.6
0.4
0.2
0.4
0.6
0.8
1
Intensity of the statistic
2 PCs
5 PCs
10 PCs
30 PCs
100 PCs
500 PCs
No PCA
Shuffled Labels

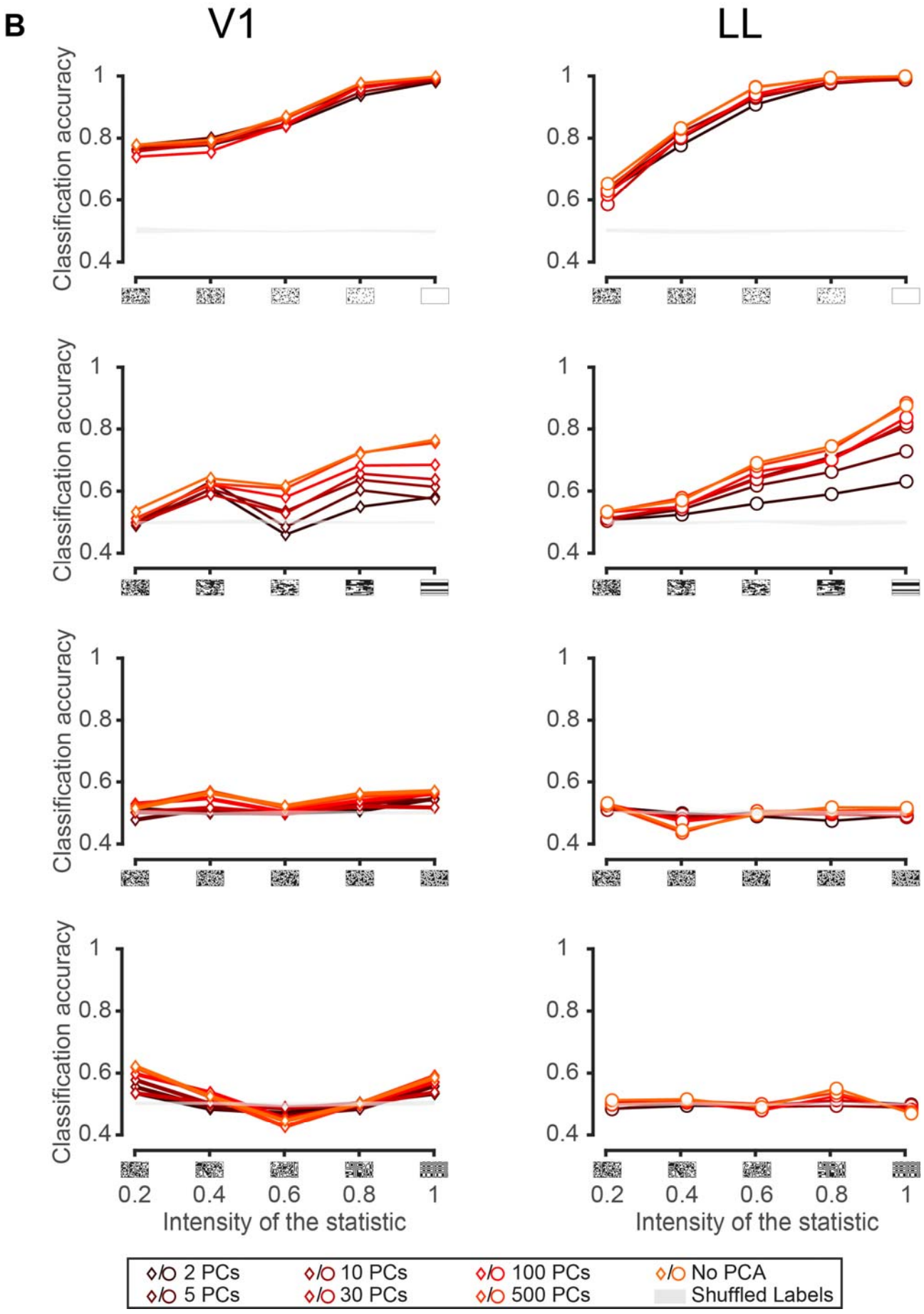

B
V1
LL
Classification accuracy
1
0.8
0.6
0.4
0.2
0.4
0.6
0.8
1
Intensity of the statistic
2 PCs
5 PCs
10 PCs
30 PCs
100 PCs
500 PCs
No PCA
Shuffled Labels

Figure 5. Classification performances for the two decoding analyses. (**A**) Classification accuracies were computed by pooling together the results of the decoding analysis without invariance performed on the activity of 10 subpopulations of 100 units. Each panel reports the performances obtained in an area (V1 left, LL right) on a specific multipoint correlation (from top to bottom: 1-, 2-, 3-, and 4-point correlations). For each intensity of these statistics, we estimated the accuracy obtained performing the classification (averaged across the 10 subpopulations) after projecting the response vectors in a space defined by a different number of principal components (PCs), identified here by the points color-code. The grey area corresponds to the performance obtained using shuffled labels when building the classifier. This was used as a reference for the chance level. The results obtained using shuffled labels for all numbers of PCs were grouped together in a single patch, as there was no substantial difference between them. (**B**) Classification accuracies obtained for the analysis with invariance, and reported as in A.

Discussion

In this study we investigated the processing of spatial correlations up to 4 point in rat primary visual cortex and exstrastriate area LL. The analyses we conducted showed that both areas can distinguish between noise textures and structured ones based on the overall luminance (1-point correlation). Considering higher-order statistics, populations of units in both areas were able to distinguish textures containing 2-point correlations, but in LL decoding performances were superior, especially when invariance was required (Figure 5B). Regarding 3-, and 4-point correlations, only in V1 we found an ability higher than chance to distinguish textures with these statistics from those containing white noise (Figure 6). Still, looking at single levels of intensity of the correlations we noticed that there wasn't a tuning of the decoding performance with the statistic intensity (Figure 5A). The principal difference we measured between the two visual areas was in their degree of invariance (see Figure 7). In V1 we found, for all correlations, a significant decrease in decoding accuracy when invariance was necessary to perform the classification task. By contrast, in LL classification performances were more stable between the non-invariant and invariant analyses, indicating that neuronal populations in this area encode information about the statistical content of a texture in a more invariant way. Previous studies have reported an increase in the invariance of object representations under identity preserving transformations from V1 to LL, in analogy to the primate ventral stream (Tafazoli et al., 2017). Our work extends these observations to the processing of textures defined by multipoint correlations of different order, providing additional evidence that rat lateral extrastriate areas behave as the functional homologue of the monkey ventral pathway (see further discussion below).

In our analysis, we performed population decoding analyses in a dimensionality reduced version of the neuronal representational space, as defined by different numbers of principal components. In fact, we wanted to measure the impact of dimensionality reduction on the classification accuracy to better understand the geometry of neural representations of the different statistics in the 2 areas (Bolanos et al., 2022). We found that, for 1-point correlations, 2 principal components were enough to reach stable performances with high accuracy in both areas (Figure 6). This means that this statistic is already discriminable from noise in a very low dimensional subspace, suggesting that luminance is an image feature that is robustly encoded by most neurons. This conclusion is consistent with the results reported by Tafazoli et al (2017), who showed how the firing of individual neurons (especially

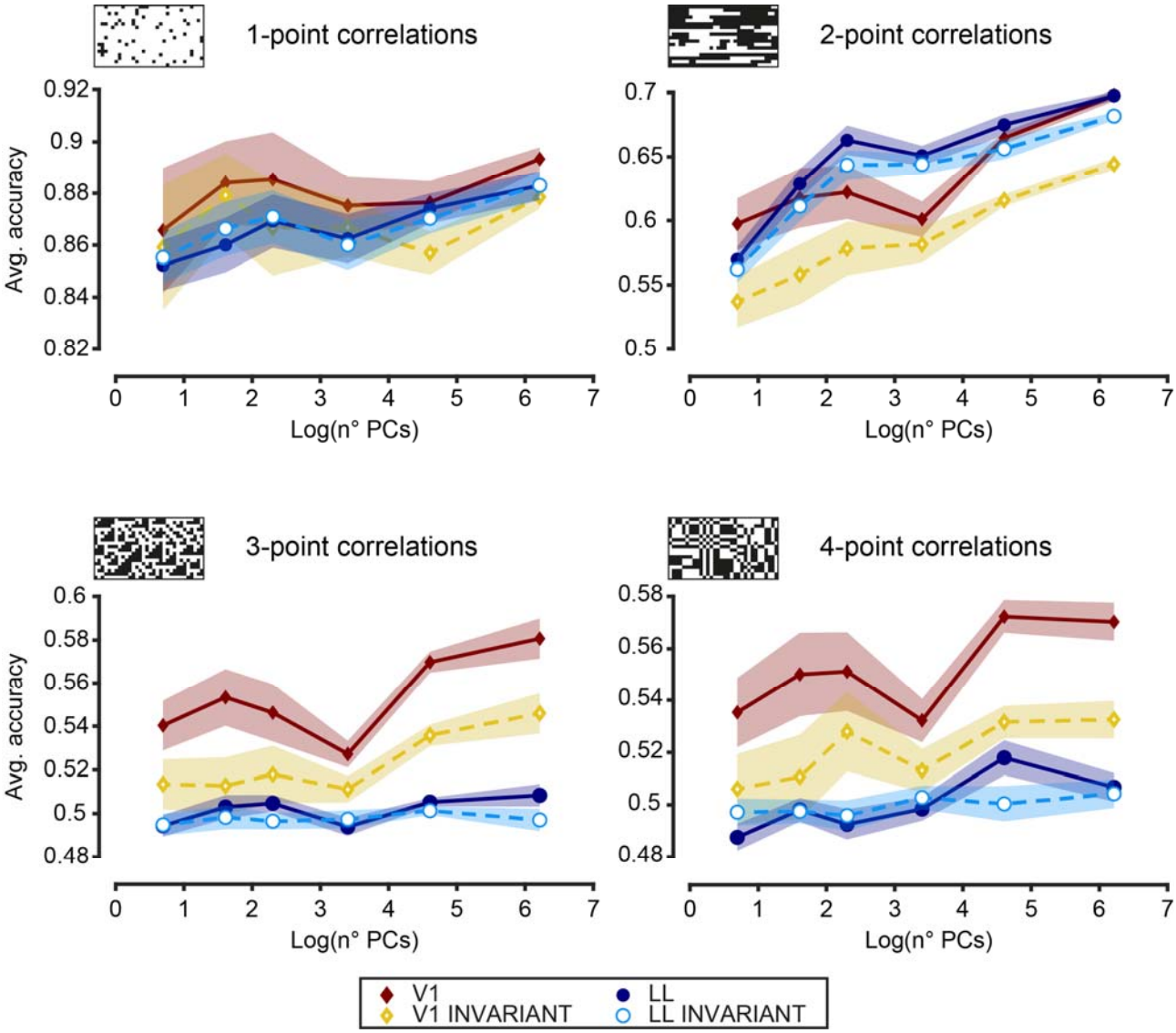


Figure 6. Average classification accuracy for multipoint correlations. Average classification accuracies across the levels of intensity of multipoint correlations were computed for each subpopulation (see Material and methods). The mean across subpopulations of these values with its standard error is presented here (data-points + shaded area). Each panel corresponds to a different correlation. Filled points connected by a solid line report the average accuracy for the analysis without invariance performed in a projection of the response space defined by a certain amount of PCs. Empty point connected by a dashed line are, instead, the results of the decoding analysis with invariance. Data-points shape codifies for the 2 different brain areas examined: diamonds for V1, dots for LL.

in V1) convey substantial information about stimulus luminance. In the case of 2-point correlations, the classification accuracy afforded by LL sharply grew when the principal components increased from 2 to 10, while the rise was much gentler when the PCs further increase up to 500. Interestingly, 5 principal components were enough to outperform V1. In V1, instead, we found a roughly constant increase in average performances up to 500 components (>90% explained variance), where, in the analysis without invariance, the average accuracy reached the level observed in LL. This indicates that 2-point correlations are represented in lower-dimensional linear subspaces in LL, as compared to V1, which is,

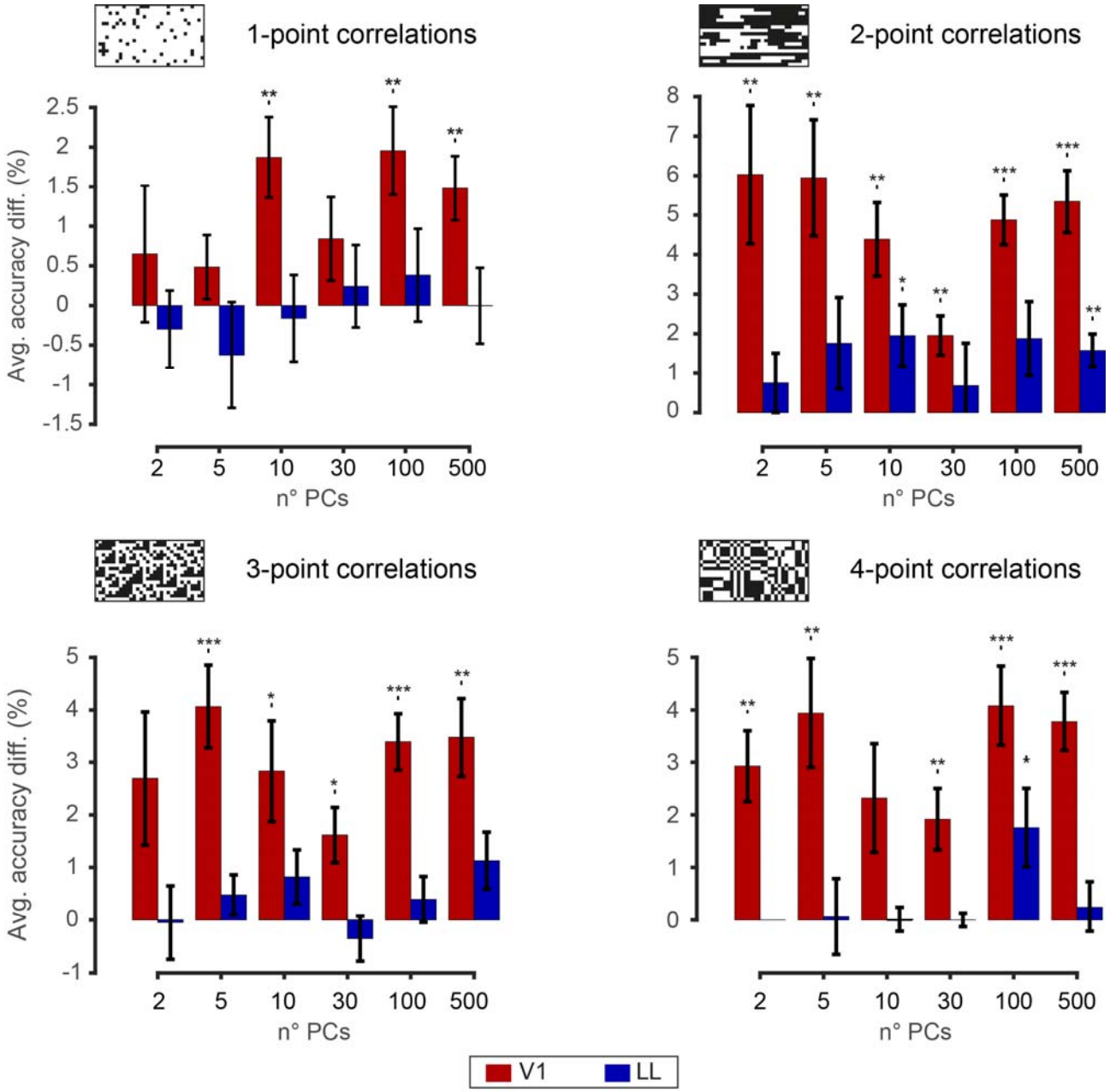


Figure 7. Differences in classification accuracy between decoding analyses. Differences in classification accuracy, between the non-invariant and invariant decoding analysis, were computed for each subpopulation (see Material and methods). The mean + SEM of this values is presented here. Each panel corresponds to a different multipoint correlation. Red bars report average differences in performance measured for V1 subpopulations, while blue bars correspond to the average differences for LL subpopulations. Differences are separately presented for every number of PCs used during the decoding process. Significance markers *, ** and *** indicate, respectively, $p < 0.05$, $p < 0.01$ and $p < 0.001$ for a one-sample t-test.

again, consistent with higher rank of LL in ventral processing (Tafazoli et al, 2017). For 3- and 4-point correlations we measured low accuracies for both areas independently from the

number of principal components. This implies that neuronal populations are very weakly tuned for these statistics in both areas.

In the previous chapter we have seen how behavioral sensitivities for multipoint correlations in rat match those measured for humans as well as the variability of the same correlations in natural images. The outcome of the decoding analysis performed in this chapter didn't fully clarify the origin of this sensitivity ranking completely. Sensitivity to 1- and 2-point correlations was high in both the examined areas confirming what was observed at a behavioral level. At the same time, we found poor discrimination from noise for both 3- and 4-point correlations. This poor discriminability of 4-point correlations was particularly surprising, as behavioral studies showed that sensitivity for this statistic was significantly higher than the one measured for 3-point correlations.

Furthermore, our results partially differ from what was observed in single cell recordings in macaque monkey's visual cortex. The seminal study by Purpura et al. (1994) showed that, in anaesthetized monkeys, certain V1 neurons were sensitive to isodipole textures containing 4-point correlations, analogue to our textures at the maximum intensity level. A later study, performed using isodipole textures with a statistical content corresponding to our 3- and 4-point correlations at maximum intensity, found that these statistics were discriminable from noise only starting from V2 (Yu et al., 2015). We found that V1 populations attained an average classification accuracy above chance on 3-, and 4-point correlations, while LL populations were unable to distinguish these statistics form noise effectively. The reason behind LL neurons lack of sensitivity for these statistics in our study is still unclear, but we hypothesize it could be due to the scale of the pixel that form the textures. In our experiments, we used textures with a pixel dimension close to anaesthetized rat visual acuity limit. This choice was made to guarantee that the fraction of pattern inside a unit's receptive field was representative enough of the entire texture. Still, this design could have negatively impacted the discriminability of statistics defined by more complex patterns, given the tendency of rat visual neurons to respond more strongly to low spatial frequencies, at least under anaesthesia (Matteucci and Zoccolan, personal communication). To test this hypothesis, we are planning to perform further recording sessions using textures made of larger pixels, i.e. at lower spatial frequency closer to the peak sensitivity of anaesthetized rats. Moreover, we are also currently performing an exploratory investigation on how deep neural networks encode maximum entropy textures defined by multipoint correlations as a function of the pixel's dimension.

When comparing the results of this chapter to those yielded by the behavioral experiments, it should be taken into account the largest invariance to the appearance of the texture instances we found in LL, as compared to V1. Similarly, to the case of object recognition (DiCarlo et al, 2012), also in a texture discrimination task as the one described in the previous chapter, tolerance to variation in the appearance the input stimuli is an essential requirement to succeed in the task. During the course of training, rats were presented with hundreds of different texture instances for each intensity level (see Chapter 2). This made it impossible for them to memorize every specific instance. To succeed in the task, they had to generalize to constantly new instances of the trained statistic. From a cortical point of view, this means that the underlying neuronal representation must be able to tolerate variations in the appearance of the features that define the textures – e.g., in the location or scale of the patterns of which the textures are made, such as the horizontal stripes produced by 2-point correlations, or the L or squared shapes produced by 3- and 4-

point correlations. In addition, since the input stimuli are not deterministic, the defying patterns are embedded in a background of noise, which, in analogy to the case of object recognition, requires some form of clutter tolerance in order to be processed – i.e., an ability to extract the target pattern from nuisances in the background (Dicarlo et al, 2012; Zoccolan et al, 2005, 2007). The fact that LL displays a virtually perfect invariance to all these transformations in the appearance of texture instances suggests that this area, way more than V1, is the one supporting the discrimination performance measured for the rats in the previous chapter, at least for 1- and 2-point statistics.

With regard to 2-point statistics, in our study we found an equal, or even higher, discriminability of these correlations from noise in LL than in V1 (see Figure 6). This result could seem surprising, as textures defined by 2-point correlations appear to be close to horizontal gratings at the higher levels of intensity of the statistic. If textures defined by 2-point correlations were fully analogue to oriented gratings, one would expect them to be maximally discriminable in the V1 representation, given that orientation selectivity decreases from V1 to LL (Matteucci et al, 2019). Still, it's important to remember that these textures are not the same as gratings. They don't possess a fixed spatial frequency and the orientated pattern is only partially maintained at low intensity of the statistic. Because of to these reasons, LL larger receptive fields (Tafazoli et al, 2017; Mattteucci et al, 2017) could allow to a better integration of the information about the statistical content of a texture, despite the lower orientation selectivity compared to V1. In addition, the larger invariance of LL neurons to a number of transformations (e.g., translation and scale; Tafazoli et al, 2017), and specifically to phase changes of oriented gratings (Matteucci et al, 2019), could also contribute to support the better encoding of "irregular", not fully deterministic oriented patters.

In summary, our findings provide robust evidence of the increase in invariant processing of visual patterns along rat putative ventral stream. The ability of LL neurons to encode 2-point correlations in a more invariant way, compared to V1 neurons, suggests that the extraction of local statistics, which requires to integrate information coming from multiple locations, happens in the higher stages of the rat's visual hierarchy. Still, for 3- and 4-point correlations we could only find weak tuning in populations of V1 neurons, which was virtually abolished when generalization to texture instances was required (i.e., invariant task), and of tuning in LL. Still, further research is needed to understand the reasons behind the lack of sensitivity for these statistics in the latter. Finding the precise neural basis behind the sensitivity for these two statistics, and the conditions under which this sensitivity emerges, could be fundamental to understand what processes influence the extraction of specific statistical properties of an image. Moreover, this is a necessary step towards providing a causal demonstration of efficient coding through altered-rearing experiments in rodents (Hunt et al., 2013; Matteucci and Zoccolan, 2020; White and Fitzpatrick, 2007).

Materials and Methods

Animal preparation and surgery

A total of 12 naïve Long Evans male rats (Charles River Laboratories) underwent extracellular recording either from the primary visual cortex or the laterolateral extrastriate area (LL). Animals weighted between 300g and 600g and had an age between 1 and 6

months. Each rat was initially anesthetized with an intraperitoneal (IP) injection of a solution of 0,3 mg/kg of medetomidine (Dormitor®, Orion Pharma) and 0,3 mg/kg of fentanyl (Fentanest®, Pfizer). The level of anaesthesia was monitored during the entire length of the procedure by checking the absence of hind paw reflex, as well as monitoring blood oxygenation, heart and respiratory rate through a pulse oximeter (Pulsesense-VET, Nonin). The anaesthesia level was maintained stable through a continuous IP infusion of the same anaesthetic solution used for the initial induction at a lower concentration (0.1 mg/kg/h Fentanyl and 0.1 g/kg/h Medetomidin), using a syringe pump system (NE-500; New Era Pump Systems). Animal Body temperature was kept constant at ~37° by a warming pad, due to the risk of anaesthesia induced hypothermia. A continuous flow of oxygen was delivered to the rat to prevent hypoxia. Eyes hydration was preserved by covering them with ophthalmic solution (Epigel®; Ceva Vetem).

After induction, and the subsequent loss of hind paw reflexes, the animal was placed in a stereotaxic apparatus (Narishige, SR-5R) with the skull parallel to the base of the device. An incision was then performed on the scalp over the left and posterior parietal bones, followed by a craniotomy over the target area in the left hemisphere (typically, a 2x2 mm window). Dura was removed to allow the insertion of the electrode. Throughout the experiment, the exposed brain was irrigated with saline solution to prevent drying. After the completion of the surgical procedure, and prior to probe insertion, the animal right eye (contralateral to the craniotomy) was immobilized using and eye ring, while the left eye was covered with a black tape patch. The stereotaxic apparatus was, then, placed on an elevated rotating platform. The platform was rotated until the right eye of the rat was directed towards the monitor keeping a distance of 30cm between animal's eye and display. When targeting V1, the ideal coordinates of penetration were ~6.5 mm posterior from bregma and ~3.5 mm left of the sagittal suture (i.e., AP 6.5, ML 3.5). 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) with an angle of insertion of 20° relative to the support of the stereotaxic apparatus.

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.

Extracellular recordings

Extracellular recordings were performed using a 2-shanks 64 channels silicon array with a 25 µm distance between recording sites (NeuroNexus Technologies - A2x32-5mm-25-200-177-A64). The electrode was grounded before insertion in the brain (by wiring it to the animal's skull skin), and then manually lowered using an oil hydraulic micromanipulator (Narishige, MO-10; insertion speed: ~ 5 µm/s) to the desired depth (~800-1000 µm when targeting V1, ~2500-3000 µm when targeting LL). As mentioned in the previous section, to reach LL the probe was inclined of 20° relative to the support of the stereotaxic apparatus (corresponding to 70° relative to the brain surface), whereas, for V1 recordings, it was inserted perpendicularly to the brain surface.

Extracellular signals were acquired using a TDT system 3 workstation with a sampling rate of ~24 kHz in combination with an RS4 data streamer (Tucker-Davis Technologies). Action potentials (spikes) were detected online by thresholding the band-pass filtered signal

(0.3-3 kHz) and were used for data monitoring and the estimation of receptive fields at each recording site (detailed procedure below).

Visual stimuli

Stimuli were presented full-field to the right eye of the rat at a distance of 30 cm on a 47 inches LCD monitor (SHARP PN-E471R), with 1920x1080 resolution, 60 Hz refresh rate, 9 ms response time, 700 cd/m^2 maximum brightness, 1200:1 contrast ratio, and spanning a visual field of ~120° azimuth and ~89° elevation.

During a recording sessions two types of stimulus sets were used: a 10-15 minutes long receptive field (RF) mapping protocol used to determine the quality of the neural activity and to precisely identify the visual area each channel was recording from (see details below); and a ~2 hours long main protocol containing maximum entropy textures, gratings, and moving bars, analogue to those used in the RF mapping protocol. The presentation of the stimuli was controlled through MATLAB® using the Psychtoolbox package (Kleiner, 2007), and in both sets stimuli were presented in a pseudorandom order. All stimuli were gamma-corrected before presentation.

The receptive field protocol consisted of 10° drifting oriented bars presented at 73 locations on the screen. The bars covered the central 110° horizontally and the central-upper 70° vertically. This choice was made to maximise the coverage of the screen in relation to rat's visual field. Bars had 4 different orientations (0º, 45º, 90º and 135º) and moved perpendicularly to their orientation from -5° to 5° and back again to -5° with respect to the starting location. A full receptive field mapping set contained 292 short movies (73 locations x 4 orientations), each with a duration of 0.363 s (11 frames at 30Hz) followed by a 0.25 s interval of black screen. The bars were white over a black background. During presentation of these stimuli, channels spiking responses were plotted as a function of screen position in real-time, in order to estimate the positions of the RFs across the recording sites. Specifically, area LL was identified by the third "rightwards" (i.e., from nasal to temporal) reversal of the retinotopy at a recording depth of ~2.5 mm dorso-ventral (Tafazoli, 2017). The same analysis was also repeated offline to associate recorded units after spike sorting to a channel RF (each unit was associated to the most represented channel across its spikes).

Once the final recording location was determined through the RF mapping procedure, the main protocol was presented. The protocol contained 3 types of stimuli:

1. Texture movies. We used as main stimuli maximum entropy textures generated using the methods described by Victor and Conte (Victor & Conte, 2012). To this end, we implemented a standalone library and software package that we have since made publicly available as free software (Piasini, 2021). In the experiment, we used white noise textures as well as textures with 5 positive intensities (0.2, 0.4, 0.6, 0.8, 1) of four different multipoint correlations from 1- up to 4-point, the same used in Caramellino et al. (2021). It should be noted that, with the exception of the extreme value of the 1-point correlation statistic, where an intensity equal to 1 corresponds to a fully white image, the intensity level of a given statistic does not specify deterministically the resulting texture image. In the protocol, for any intensity level of

each statistic, 6 random instances of the textures were included. Each texture had a size of 32x18 pixels and occupied the entire monitor (full- field stimuli). The pixels had a dimension of 6.25° each. This size corresponds to a spatial frequency of 0.08 c/d (cycles/degree). Textures had a resolution of 320x180 pixels and were up-scaled to 1920x1080 pixels before being displayed. During presentation (1 s total) each texture underwent a contrast reversal procedure. Specifically, the contrast was modulated by a smoothed square wave: $y = \frac{1}{atan(1/\delta)} atan(sin(\frac{2\pi tf}{\delta}))$; where δ is the smoothing factor (δ = 0.2), *f* is the wave frequency (*f* = 2 Hz), and *t* is the "time" (0 $\leq$ *t* $\leq$ 1). Resulting movies were formed by 30 frames (30 Hz presentation frequency) where the texture started from mid gray, then reached max contrast, returned to mid gray, had its contrast reversed, returned again to mid gray, and the process was repeated (due to 2 Hz frequency). Frames were sampled linearly along across contrast values, leading to more frames close to contrast peaks.

2. Grating movies. The protocol contained static gratings with 6 possible orientations (0°, 30°, 60°, 90°, 120°, 150°) and 4 phases (-π/2, -π/6, π/6, π/2) at a spatial frequency of 0.08 c/d. These stimuli had a resolution of 960x540 un were up-scaled to 1920x1080 before presentation. Each grating was presented for 1 s and underwent the same contrast reversal procedure as texture stimuli. The resulting grating movie was therefore formed by 30 frames (presented with a frequency of 30Hz).
3. Drifting Bars. The protocol also contained drifting bars identical to those used during the receptive field mapping procedure.

All stimuli included in the main protocol were presented interleaved in a pseudorandom order that differed in each session. The full protocol was formed by 30 blocks each containing (in a different order) all textures, gratings, and 1/6th of the bars.

Spike-sorting procedure and area assignment

The raw activity recorded by each channel of our probe in a session was analysed offline to obtain single and multi-unit responses. We used Kilosort3 (Pachitariu, 2016) to perform spike extraction and the automatic clustering of the activity. We decided to do not perform any manual curation of the resulting clustering as the final result was satisfying and we didn't want to include any personal bias in the procedure (results of the clustering procedure were visually analysed using Phy template-gui). In all analysis we used both single and multi-units.

The units recorded in sessions targeting LL were assigned to the exstrastriate area to which they belonged based on the reversal of the retinotopy (Matteucci et al., 2019; Tafazoli et al., 2017). To perform this assignment, we decided to use the channel activity during the initial RF mapping protocol as it lead to RFs that were better defined compared to those obtained using the units' responses to the drifting bars included in the main protocol. Each unit was associated to the channel where most of its activity was recorded. We discarded units corresponding to channels with undefined RFs as it was impossible to track the retinotopy. This procedure allowed us to identify 288 units in LL and 35 units in LI (laterointermediate area). Units assigned to LI were not used in any analysis.

Determination of units' inactivity windows

During recording sessions and spike-sorting we noticed that units' activity often did not span the full length of the registration. There were periods of complete inactivity not associated to the stimuli on screen, but mostly due to:

- Electrode drifting: electrode micro-movement causes some units to be detected later in the recording or to stop being detected early.
- Units death: Units that may have been damaged by the electrode insertion (or other external sources such as anaesthesia) can die before the end of the session.
- Activity bursts: During a recording session it may happen to have a highly synchronised activity burst across all channels of the probe followed by several minutes of extremely low activity unrelated to the stimuli. These bursts cause sorted units to have lapses in their activity profile with almost any spike.

Considering that the periods of inactivity were not related to the stimuli we were administering to the rat, we decided to exclude them from the data analysis. To identify the precise time intervals where a unit was not responding due to stimulus-unrelated factors, we first computed the unit average response over the entire duration of the recording. Then, we divided the activity in 60 s time bins. We defined a time bin, and all the corresponding stimuli trials, to be "inactive" if the unit's activity in the previous bin was higher than the average while the activity at the actual bin and at least the next 3 ones (4 minutes total) was lower than half the average activity. We considered an inactive interval to end when the activity returned above half the average activity.

We established half the average activity as a threshold because in presence of prolonged periods of inactivity the neuron average activity was lower than its baseline under normal circumstances (the remaining of the recording). Therefore, when activity went below this threshold value for a long period it could only be due to a burst or the death of the unit. Visual inspection of the results confirmed that this criterion was effective in determining the inactivity periods.

We wanted the low/no activity intervals to be at least several minutes long because we didn't want to mistake stimulus-related inactivity with this stimulus-unrelated one.

Population decoding analysis – Data pre-processing and base structure

The population decoding analysis consisted of training SVM classifiers (Cortes & Vapnik, 1995) to distinguish between trials containing a specific level of intensity of one of the multipoint correlations and white noise trials. Given that the number of units between the two areas in exam was considerably different (431 units in V1 vs 288 units in LL), we needed to perform a population sub-sampling in order to obtain comparable results. Therefore, we built 10 sub-populations of 100 units for both areas. The decoding procedure was run separately on each subpopulation. Before performing the decoding, for each statistic (including noise textures), we removed from the dataset all units that were inactive (as described above) for 10 or more trials of a single stimulus. We then determined the remaining unit with the highest number of inactive trials and, after randomly shuffling each unit's trials order, we equalized the number of trials across units by removing those in excess.

Working with dynamic stimuli, due to the contrast reversal, we had to consider, for each trial of a certain stimulus, the response of each unit of a sub-population in every time bin we divided the textures presentation (30 time bins, 33 ms each). Our classifier, therefore, would have to operate in a space with dimensionality equal to the number of units in the sub-population (100) multiplied by the number of time bins (30). This number of dimensions was considerably higher than the number of trials of the stimuli leading to many redundant features. In order to decrease this redundancy and partially de-noise textures representations, before classification, we performed dimensionality reduction on our data using PCA (Pearson, 1901). PCA was applied to all trials coming from textures sharing the same statistic with the inclusion of white noise (noise textures were the same for all statistics). After applying PCA, we projected our data matrix in the principal components (PCs) space defined by the first N principal components, with N = 2, 5, 10, 30, 100, 500 PCs depending on the analysis.

Population decoding analysis – decoder "without invariance"

The SVM classifier was fed the responses (in PC space of units' activations) in the trials of the 6 movies corresponding to a certain level of one multipoint correlations, and those of the 6 textures movies containing white noise. Therefore, each trial corresponded to a point in the features space and the classifier had to separate data-points corresponding to a statistic intensity greater than 0 from those corresponding to noise. The accuracy of the classifier was assessed by computing the average ratio of correct classifications through a 3-fold cross validation process. To determine the chance level associated to each classifier we also performed the same decoding procedure with shuffled labels between the two classes (statistic vs noise).

Population decoding analysis – decoder "with invariance"

We performed a second type of decoding analysis, structured as the one "without invariance". Here, instead of using randomly selected cross-validation folds containing trials coming from any texture, we specifically selected a held-out set comprised of all the trials belonging to 2 textures for each class. In order to balance the procedure, that could be biased by the choice of specific combinations of training/test sets, we repeated the decoding for each possible pairing of textures in the training and test set. The final classification performance, for a specific comparison between textures containing a certain intensity of a statistic and white noise, is obtained by averaging the classification across all the combinations of training and test sets.

The main characteristic of the invariant decoding analysis is that, to achieve a good accuracy, the classification model has to possess a certain degree of invariance for changes among textures containing the same intensity of a statistic. The same invariance is not required in the previous decoding analysis as the cross-validation splits are randomly selected, and, therefore, each of them contains trials coming from every texture, so the model is not tested on trials coming from completely new stimuli.

Both decoding analyses were performed using the responses of the same 10 subpopulations of units in the 2 examined areas. This was done to guarantee that the results obtained for the two types of analysis were comparable.

Average classification accuracies and differences between analyses

Considering a single area and number of PCs, each of the classifiers we built resulted in a certain level of accuracy for the discriminability of a specific level of intensity of a statistic from noise. As mentioned above, every classification was run 10 times using different subpopulations of 100 units, so for a level of intensity of a multipoint correlation we had ten classification accuracy measures. We then computed the average classification accuracies (seen in Figure 5) by averaging these 10 estimates.

In order to extract a single quantification of the discriminability from noise of each statistic, we also computed an average classification accuracy across levels of intensity (Figure 6). This value was obtained by first averaging the accuracy obtained for each correlation intensity for a single population and then computing the mean + SEM across the 10 subpopulations. An analogue procedure was carried out when computing the differences between the accuracy obtained in the “non-invariant” decoding process with those obtained in the invariant one. Working with the same subpopulations in both analyses, we were able to compute coupled differences of performance across statistic intensities, and then obtain an average difference for each subpopulation (Figure 7). We run a one-sample two-tails t-test to verify if the average of a distribution of differences for a certain combination of statistic, area and number of PCs was significantly different from 0. This analysis was performed to test if there was a significant difference in classification accuracy between the two decoding analyses.

References

Alemi-Neissi, A., Rosselli, F., & Zoccolan, D. (2013, April 3). Multifeatural Shape Processing in Rats Engaged in Invariant Visual Object Recognition. *Journal of Neuroscience, 33*, 5939-5956. doi:10.1523/jneurosci.3629-12.2013

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183–193. doi:10.1037/h0054663

Barlow, H. B. (1961, September). Possible Principles Underlying the Transformations of Sensory Messages. In *Sensory Communication* (p. 216–234). doi:10.7551/mitpress/9780262518420.003.0013

Bolaños, F., Orlandi, J., Aoki, R., Jagadeesh, A., Gardner, J., & Benucci, A. (2022, September 17). Efficient coding of natural images in the mouse visual cortex. *bioRxiv*. doi:10.1101/2022.09.14.507893

Caramellino, R., Piasini, E., Buccellato, A., Carboncino, A., Balasubramanian, V., & Zoccolan, D. (2021, December). Rat sensitivity to multipoint statistics is predicted by efficient coding of natural scenes. (S. E. Palmer, & T. E. Behrens, A cura di) *eLife, 10*, e72081. doi:10.7554/eLife.72081

Cortes, C., & Vapnik, V. (1995, September). Support-vector networks. *Machine Learning, 20*, 273-297. doi:10.1007/bf00994018

De Keyser, R., Bossens, C., Kubilius, J., & Op De Beeck, H. (2015, November 25). Cue-invariant shape recognition in rats as tested with second-order contours. *Journal of Vision, 15*, 14. doi:10.1167/15.15.14

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012, February). How Does the Brain Solve Visual Object Recognition? *Neuron, 73*, 415–434. doi:10.1016/j.neuron.2012.01.010

Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H., & Zoccolan, D. (2018). Accuracy of Rats in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual Strategy. *Current Biology, 28*, 1005-1015.e5. doi:https://doi.org/10.1016/j.cub.2018.02.037

Felleman, D. J., & Essen, D. C. (1991, January). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1*, 1–47. doi:10.1093/cercor/1.1.1

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013, May). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience, 16*, 974–981. doi:10.1038/nn.3402

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., & Tkačik, G. (2014, November). Variance predicts salience in central sensory processing. *eLife, 3*. doi:10.7554/elife.03722

Hunt, J., Dayan, P., & Goodhill, G. (2013, May 9). Sparse Coding Can Predict Primary Visual Cortex Receptive Field Changes Induced by Abnormal Visual Input. *PLoS Computational Biology, 9*, e1003005. doi:10.1371/journal.pcbi.1003005

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox. *What's new in Psychtoolbox*.

Laughlin, S. (1981, October). A Simple Coding Procedure Enhances a Neuron\textquotesingles Information Capacity. *Zeitschrift für Naturforschung C, 36*, 910–912. doi:10.1515/znc-1981-9-1040

Lehky, S. R., & Tanaka, K. (2016). Neural representation for object recognition in inferotemporal cortex. *Current Opinion in Neurobiology, 37*, 23-35. doi:https://doi.org/10.1016/j.conb.2015.12.001

Matteucci, G., & Zoccolan, D. (2020, May). Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. *Science Advances, 6*. doi:10.1126/sciadv.aba3742

Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B., & & Zoccolan, D. (2019). Nonlinear processing of shape information in rat lateral extrastriate cortex. *Journal of Neuroscience, 39(9)*, 1649-1670.

Okazawa, G., Tajima, S., & Komatsu, H. (2016, September). Gradual Development of Visual Texture-Selective Properties Between Macaque Areas V2 and V4. *Cerebral Cortex*. doi:10.1093/cercor/bhw282

Orban, G. A., Essen, D. V., & Vanduffel, W. (2004, July). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences, 8*, 315–324. doi:10.1016/j.tics.2004.05.009

Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M., & Harris, K. D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (A cura di), *Advances in Neural Information Processing Systems. 29.* Curran Associates, Inc. Tratto da https://proceedings.neurips.cc/paper/2016/file/1145a30ff80745b56fb0cecf65305017-Paper.pdf

Pearson, K. (1901). Principal components analysis. The London, Edinburgh, and Dublin Philosophical Magazine and. *Journal of Science, 6*, 559.

Purpura, K., Victor, J., & Katz, E. (1994, August 30). Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences, 91*, 8482-8486. doi:10.1073/pnas.91.18.8482

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and Tolerance (\textquotedblleftInvariance\textquotedblright) Both Increase as Visual Information Propagates from Cortical Area V4 to IT. *Journal of Neuroscience, 30*, 12978–12995. doi:10.1523/JNEUROSCI.0179-10.2010

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982, November). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 216*, 427–459. doi:10.1098/rspb.1982.0085

Tafazoli, S., Safaai, H., Franceschi, G. D., Rosselli, F. B., Vanzella, W., Riggi, M., . . . Zoccolan, D. (2017, April). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife, 6*. doi:10.7554/elife.22794

Tesileanu, T., Conte, M. M., Briguglio, J. J., Hermundstad, A. M., Victor, J. D., & Balasubramanian, V. (2020, August). Efficient coding of natural scene statistics predicts discrimination thresholds for grayscale textures. *eLife, 9*. doi:10.7554/elife.54347

Tkacik, G., Prentice, J., Victor, J., & Balasubramanian, V. (2011, June). Local statistics in natural scenes predict the saliency of synthetic textures. *Nature Precedings*. doi:10.1038/npre.2011.6009.1

Vermaercke, B., & Op de Beeck, H. (2012). A Multivariate Approach Reveals the Behavioral Templates Underlying Visual Discrimination in Rats. *Current Biology, 22*, 50-55. doi:https://doi.org/10.1016/j.cub.2011.11.041

Vermaercke, B., Gerich, F. J., Ytebrouck, E., Arckens, L., de Beeck, H. P., & den Bergh, G. V. (2014, October). Functional specialization in rat occipital and temporal visual cortex. *Journal of Neurophysiology, 112*, 1963–1983. doi:10.1152/jn.00737.2013

Victor, J. D., & Conte, M. M. (2012, June). Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America A, 29*, 1313. doi:10.1364/josaa.29.001313

Vinken, K., Vogels, R., & de Beeck, H. O. (2017). Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. *Current Biology, 27*, 914-919. doi:https://doi.org/10.1016/j.cub.2017.02.024

White, L., & Fitzpatrick, D. (2007, October). Vision and Cortical Map Development. *Neuron, 56*, 327-338. doi:10.1016/j.neuron.2007.10.011

Yu, Y., Schmid, A. M., & Victor, J. D. (2015, April). Visual processing of informative multipoint correlations arises primarily in V2. *eLife, 4*. doi:10.7554/elife.06604

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016, May). Selectivity and tolerance for visual texture in macaque V2. *Proceedings of the National Academy of Sciences, 113*. doi:10.1073/pnas.1510847113

Ziemba, C. M., Perez, R. K., Pai, J., Hallum, L. E., Shooner, C., Kelly, J. G., & Movshon, J. A. (2019, July). Laminar differences in responses to naturalistic texture in macaque V1 and V2. doi:10.1101/710426

Zoccolan, D. (2005, September). Multiple Object Response Normalization in Monkey Inferotemporal Cortex. *Journal of Neuroscience, 25*, 8150–8164. doi:10.1523/jneurosci.2058-05.2005

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007, November). Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex. *Journal of Neuroscience, 27*, 12292–12307. doi:10.1523/jneurosci.1897-07.2007

Chapter IV

Conclusions

The research described in this thesis aimed at investigating rat ability to process binary (black and white) visual textures defined by local correlations. These image features constitute the cornerstone of the higher order statistical structure in natural scenes. In our behavioral study (Project I; chapter 2) I provided strong evidence that rats and humans are adapted to process multipoint correlations up to 4 points in the same way. This result testifies the fundamental role of efficient coding as a normative principle at the cortical stages of visual processing (Hermundstad et al., 2014; Tkačik et al., 2010).

Our investigation for the neural correlates behind the processing of multipoint correlations (Project II; chapter 3) was not able to precisely identify where the most complex correlations, 3- and 4-point ones, are processed. Still, we found a clear difference in invariance between V1 and LL in terms of encoding 1- and 2-point image statistics. This further confirm the existence of a hierarchical shape processing stream in rat (Tafazoli et al., 2017), analogue to primate ventral stream.

The lack of discriminability for 3- and 4-point correlations in LL, which sits at the higher stages rat ventral pathway, was not in line with our hypothesis. Further analysis on single neurons responses might allow us to understand the reasons for the low discriminability of these correlations from noise. This could be due to the highly dimensionality of the feature space in population decoding analysis, combined with the noise of such high-dimensional representations. These two elements together could make extracting useful information for the classification task more difficult. Alternatively, the scarce sensitivity for these correlations might be a consequence of a complete absence of units that are tuned to them. Furthermore, in the first project presented, we have seen that rats trained on 3- and 4-point correlations required a longer time to learn the discrimination task, as compared to those trained on lower order correlations. This could imply that, in rats, a certain degree of training is necessary to make neural representations of higher order statistics discriminable from noise. Future investigation of the neural activity of trained rats might provide useful information on the effect of training on the representations of multipoint correlations in the different areas of rat's visual cortex. In the meantime, we are going to perform additional recordings in anaesthetized naive rats using textures with reduced spatial frequency as pixel dimension could strongly influence the ability of neurons to extract information on higher order correlations. Finally, one should consider that at least one additional downstream, higher-order stage beyond LL exists along the rat ventral pathway – an area named TO that has been explored in a few studies (Vermaercke et al., 2014; Vinken et al., 2016). One possibility is that 3- and 4-point correlations are more explicitly encoded in this area.

Overall, our results were able to provide strong attestation of the role of efficient coding in shaping visual perception across species. Having found matching sensitivities between rats and humans to textures defined by local multipoint correlations confirms the importance of these statistics in the processing of natural scenes. Through these results, we were able to attest the validity of rodents as a viable animal model to study to impact that

adaptation to the statistical features of the environment has on cortical visual processing. This finding opens the way to the possibility of performing rearing experiments in an altered visual environment (Matteucci & Zoccolan, 2020), which could provide a causal proof of efficient coding influence on the development of the visual system. The neural recordings we performed displayed a significant increase of invariance in LL compared to V1. This result corroborates the analogy between rats exstrastriate areas and the primate ventral stream and supports the possibility of using visual textures to probe fundamental properties of visual processing in rodents. The outcome of this research could also be relevant to the field of artificial intelligence. The fact that both humans and rodents share similar adaptation to local correlations suggest that these features are particularly important for the development of visual perception. Therefore, implementing mechanisms tuned for the extraction of such statistical features could be crucial to improve CNNs ability to learn in an unsupervised manner.

In conclusion, our results: 1) provide a demonstration for the importance of efficient coding in sculpting perception across species; 2) further reinforce the role of the rat as an animal model to study cortical visual processing; and 3) open a path towards a direct causal study of the adaptation of the visual system to the statistical features of the natural environment.

References

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., & Tkačik, G. (2014, November). Variance predicts salience in central sensory processing. *eLife, 3*. doi:10.7554/elife.03722

Matteucci, G., & Zoccolan, D. (2020, May). Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. *Science Advances, 6*. doi:10.1126/sciadv.aba3742

Tafazoli, S., Safaai, H., Franceschi, G. D., Rosselli, F. B., Vanzella, W., Riggi, M., . . . Zoccolan, D. (2017, April). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife, 6*. doi:10.7554/elife.22794

Tkačik, G., Prentice, J. S., Victor, J. D., & Balasubramanian, V. (2010, October). Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences, 107*, 18149–18154. doi:10.1073/pnas.0914916107

Vermaercke, B., Gerich, F. J., Ytebrouck, E., Arckens, L., de Beeck, H. P., & den Bergh, G. V. (2014, October). Functional specialization in rat occipital and temporal visual cortex. *Journal of Neurophysiology, 112*, 1963–1983. doi:10.1152/jn.00737.2013

Vinken, K., den Bergh, G. V., Vermaercke, B., & de Beeck, H. P. (2016, May). Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. *Cerebral Cortex, 26*, 3310–3322. doi:10.1093/cercor/bhw111

Acknowledgements

I would like to express my gratitude to all the people that have helped me during my PhD. First, I want to thank my supervisor Prof. Davide Zoccolan for his invaluable guidance and constant support through these 4 years. I'm also deeply indebted to my teachers Liviu Soltuzu, Alessandro Di Filippo, Natalia Grion and Giulio Matteucci for teaching me everything I know about data analysis and how to perform an experiment. Without them I could have never completed this project. I would also like to thank all the members of my lab, and especially my follow PhD students Angelina, Mattia and Paolo for being great companions in this journey, always available for support and advice. I would also like to show my gratitude the people that collaborated with me on the projects presented in this thesis: Anna Carboncino, Eugenio Piasini and the fantastic master students I had the honour to supervise during their thesis: Andrea Buccellato, Valeriya Zelenkova and Daria Ricci, great scientists and incredible persons.

Very special thanks to Margherita Riggi, who not only thought me a lot about working with animals, but was an extraordinary friend, always patient, encouraging and supportive whenever I needed.

Many thanks to Sogand Talebi and Silene Fornasaro, I will always keep dear memory of the incredibly fun time we spent together in Trieste. In particular, I have to thank Silene for also giving me invaluable support and always cheering me up during the last stressing months of the PhD.

I thank my friends Simone and Alessio for all the great time we had together. Our gaming nights have always been the best way to relax after a hard day of work.

Finally, of course, I want to express my very profound gratitude to my family for supporting me in every choice I made and for always being there when I needed.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.