Check for updates

Efficient processing of natural scenes in visual cortex

Tiberiu Tesileanu[1]*, Eugenio Piasini[2]* and
Vijay Balasubramanian[3,4]

[1]Center for Computational Neuroscience, Flatiron Institute, New York, NY, United States, [2]Scuola
Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy, [3]Department of Physics and
Astronomy, David Rittenhouse Laboratory, University of Pennsylvania, Philadelphia, PA,
United States, [4]Santa Fe Institute, Santa Fe, NM, United States

Neural circuits in the periphery of the visual, auditory, and olfactory systems are
believed to use limited resources efficiently to represent sensory information
by adapting to the statistical structure of the natural environment. This "efficient
coding" principle has been used to explain many aspects of early visual
circuits including the distribution of photoreceptors, the mosaic geometry and
center-surround structure of retinal receptive fields, the excess OFF pathways
relative to ON pathways, saccade statistics, and the structure of simple cell
receptive fields in V1. We know less about the extent to which such adaptations
may occur in deeper areas of cortex beyond V1. We thus review recent
developments showing that the perception of visual textures, which depends
on processing in V2 and beyond in mammals, is adapted in rats and humans
to the multi-point statistics of luminance in natural scenes. These results
suggest that central circuits in the visual brain are adapted for seeing key
aspects of natural scenes. We conclude by discussing how adaptation to
natural temporal statistics may aid in learning and representing visual objects,
and propose two challenges for the future: (1) explaining the distribution of
shape sensitivity in the ventral visual stream from the statistics of object shape
in natural images, and (2) explaining cell types of the vertebrate retina in terms
of feature detectors that are adapted to the spatio-temporal structures of
natural stimuli. We also discuss how new methods based on machine learning
may complement the normative, principles-based approach to theoretical
neuroscience.

1. Introduction: Sensory adaptation to natural environments

The sensory systems of animals face the challenge of using limited resources to
process very large and high dimensional sensory spaces. For example, the olfactory
systems of most animals use a few hundred to a thousand receptor types (Vosshall
et al., 2000; Zozulya et al., 2001; Zhang and Firestein, 2002) to encode a vast
number of mixtures of odorants drawn from the tens of thousands of possible volatile
molecules (Dunkel et al., 2009; Touhara and Vosshall, 2009; Mayhew et al., 2022) with

corresponding challenges for encoding and decoding odors (see Singh et al., 2021; Krishnamurthy et al., 2022) and references therein). The visual system faces the similarly acute problem of encoding the relevant information in continuously changing scenes composed of photons with frequencies that range continuously across the visual spectrum, light intensities spanning over 10 orders of magnitude (Tkačik et al., 2011), and a vast diversity of possible textures, shapes, objects, and kinds of motion. Overall, the challenge is that the visual input is extremely high dimensional and the human retina must manage the formidable task of encoding it with just three receptor types (red, green, and blue cones) during daytime, processed *via* the retinal network into about 20 types of visual feature detectors (retinal ganglion cells) that tile the visual space (Balasubramanian and Sterling, 2009).

The challenge of performing difficult tasks with limited resources and many constraints recurs across all the scales and domains of life. There is also a standard tactic used by living things to mitigate this challenge: evolutionary adaptation of systems to the environment or to the tasks required by the environmental niche, a notion that goes back to Darwin and his observation of the finches of the Galapagos (Darwin, 1859). Applied to sensory systems, this tactic requires adaptation of sensory algorithms, circuit architectures, and cell properties to the statistical structure of the environment on multiple timescales. Broadly, this suggests that the fixed architecture of sensory systems should be adapted over evolutionary times to the typical statistical structure of the environment, while plasticity supports fine tuning to the detailed differences that distinguish specific environments during the lifetime of individuals.

One powerful formulation of this idea is the *efficient coding hypothesis*. Different authors have adopted somewhat different formulations of this hypothesis, but we will take it to state that neural systems commit their limited resources to maximize the information relevant for behavior that they encode from the environment. To formulate this principle precisely we must define what we mean by "information," "relevant," and "behavior". However, in the sensory periphery, a standard approach is to simply assume that neural circuits do select behaviorally useful data from the environment (e.g., bright vs. dark local contrast extracted by retinal ON and OFF cells), and to ask instead how the circuits should be structured, and how computational resources should be allocated, to maximize the encoded information given biological constraints such as the number of available cells or the amount of ATP that the circuit can consume (see, e.g., Atick and Redlich, 1990; Ratliff et al., 2010 in retina, Teşileanu et al., 2019 in the olfactory epithelium, and Wei et al., 2015 in the entorhinal cortex). In information theoretic terms these approaches ask how peripheral sensory circuits should be organized to maximize the mutual information of their outputs with the environment, given constraints of noise, ATP consumption (Attwell and

Laughlin, 2001; Balasubramanian, 2021; Levy and Calvert, 2021), the number of available cells, and the like.

Thinking in this way, researchers have explained many aspects of early vision, e.g., nonlinearities in the fly visual system (Laughlin, 1981), center-surround receptive fields of neurons in the vertebrate retina (Atick and Redlich, 1990; van Hateren, 1992b; Vincent and Baddeley, 2003; Kuang et al., 2012; Pitkow and Meister, 2012; Simmons et al., 2013; Gupta et al., 2022), spike timing statistics (Fairhall et al., 2001), the preponderance of OFF cells over ON cells (Ratliff et al., 2010; Gjorgjieva et al., 2014), the mosaic organization of ganglion cells (Borghuis et al., 2008; Liu et al., 2009), the scarcity of blue cones and the large variability in numbers of red and green cones in humans (Garrigan et al., 2010), selection of predictive information by ganglion cells (Palmer et al., 2015; Salisbury and Palmer, 2016), and the expression of ion channels in insect photoreceptors (Weckström and Laughlin, 1995). Similar analyses suggest that the auditory (Schwartz and Simoncelli, 2001; Lewicki, 2002; Smith and Lewicki, 2006; Carlson et al., 2012) and olfactory (Teşileanu et al., 2019; Singh et al., 2021; Krishnamurthy et al., 2022) peripheries are also adapted to the statistical structure of the environment so that they use limited resources efficiently to represent sensory information (Sterling and Laughlin, 2015). While many of these analyses have focused on linear filtering properties, some have focused on the nonlinear separation of the visual stream into separate information channels like bright and dark spots or color channels (Garrigan et al., 2010; Ratliff et al., 2010; Gjorgjieva et al., 2014). It remains a challenge for the future to understand the complete repertoire of nonlinear visual features extracted by retinal ganglion cells (Gollisch and Meister, 2010) in these terms, a task that will likely require an extension of previous methods to include the temporal dynamics of natural scenes, the computational complexity of the required decoding network, and the mutual information between aspects of the visual stimulus and important behaviors.

Similar principles may also apply more centrally in the thalamus and in primary visual cortex, for example in asymmetries between ON and OFF responses (Komban et al., 2014; Kremkow et al., 2014) and the structure of receptive fields (Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998; Vinje and Gallant, 2000). In auditory cortex, contrast gain control has been shown to facilitate information transmission and help detecting signals against a noisy background (Rabinowitz et al., 2011; Angeloni et al., 2021). There is even evidence that the grid system in the entorhinal cortex acts as an efficient encoder of space (Wei et al., 2015). In this article we follow this line of thinking further, and describe recent studies that show that deeper layers of visual cortex in multiple species are adapted to the spatial and temporal statistics of natural scenes. In Section 2, we discuss the representation of visual textures in cortex, which occurs in areas V2 and above in humans. In Section 3, we discuss experiments

showing that the findings in humans also extend to rodents. In Section 4, we discuss a case where behavioral relevance is more narrowly defined, namely the representation of object identity in the visual ventral stream, and how representations that are invariant to identity-preserving transformations (e.g., changes in viewpoint) can be learnt from the temporal statistics of natural scenes. We conclude in Section 5 with a discussion of challenges for the future.

2. Adaptation to spatial statistics and texture perception

Above, we have discussed how the requirement of maximizing information transmission shapes the earliest visual layers. At the level of the retina and primary visual cortex (V1), neurons are often presumed to be mostly sensitive to simple, first- or second-order image statistics (e.g., Atick and Redlich, 1990; van Hateren, 1992b; Olshausen and Field, 1996; Borghuis et al., 2008; Pitkow and Meister, 2012; Simmons et al., 2013), although responses to complex spatio-temporal features are evidently also present in the phenomena like motion anticipation (Berry et al., 1999), lag-normalization (Trenholm et al., 2013), the omitted stimulus response (Schwartz et al., 2007), and nonlinear feature detection (Gollisch and Meister, 2010). As sensory data is further processed through the visual hierarchy, neurons develop selectivity for more complex visual elements such as shapes (Pasupathy and Connor, 1999; DiCarlo and Cox, 2007; Rust and DiCarlo, 2010) and textures (Landy and Graham, 2004; Freeman et al., 2013; Okazawa et al., 2015). Are these neurons are also tuned to maximize the efficient transfer of information? To approach this question, we first need a precise definition of what a visual texture is.

Intuitively, textures are images with "distinctive local features [...] arranged in a spatially extended fashion" (Victor et al., 2017); at a formal level, however, textures are best defined in terms of *ensembles* of images (Victor, 1994; Portilla and Simoncelli, 2000; Victor et al., 2017) because multiple different images can represent the same texture. Indeed, the statistical regularities of a texture patch, and not the precise spatial arrangement of light intensity, yield its perceptual quality. By grouping all images that are perceived as a single texture type into a statistical ensemble, we are effectively describing the statistical properties of the image that are important for defining that type of texture. However, despite the ensemble nature of textures, a single image is typically sufficient to identify a texture. It must thus be possible to infer statistical properties related to the whole ensemble from a single patch, provided the patch is large enough. This suggests that texture ensembles possess the property of *ergodicity*: spatial averaging coincides with ensemble averaging (Victor, 1994; Portilla and Simoncelli, 2000; Victor et al., 2017).

Within this framework, a specific type of texture is defined by a set of constraints on image statistics. This could be, for instance, fixing the average luminance, or the correlation between luminance values a certain distance apart. In general infinitely many such constraints may be needed to fully specify a texture ensemble, but a maximum-entropy approach is often used to pick out a unique ensemble that fixes only a finite number of statistics (Portilla and Simoncelli, 2000; Zhu et al., 2000; Victor et al., 2017). The entire set of textures that can be obtained by varying the values of a chosen class of image statistics can be arranged in a *texture space*, with each axis representing a statistic. These axes are generally not independent: they are subject to inter-dependencies and mutual constraints (Victor et al., 2017).

Several alternative texture parameterizations have been used in the literature (e.g., Victor and Conte, 1991; Portilla and Simoncelli, 2000; Zhu et al., 2000; Tkačik et al., 2010; Hermundstad et al., 2014; Teşileanu et al., 2020). Here we focus mainly on grayscale textures with a finite number of discrete luminance levels, and multi-point correlations restricted to small (2×2 or 3×3) neighborhoods (Victor and Conte, 1991; Tkačik et al., 2010; Hermundstad et al., 2014; Teşileanu et al., 2020).

To test efficient coding of such visual textures we need three ingredients. First, we need a method for analyzing natural images to measure the distribution of textures that is likely to be encountered by an animal. Second, we need a mathematical formalism for making testable predictions based on the natural distribution. And third, must create an experimental paradigm that allows us to test these predictions.

The first step is relatively straightforward given the formal definition of texture space described above. It amounts to selecting a dataset of natural images (e.g., van Hateren and van der Schaaf, 1998), splitting each image into patches assumed to have roughly constant texture[1], and then for each patch calculating the image statistics that define the chosen texture space. Next, each texture patch can be characterized by its coordinates in texture space, and typical efficient-coding calculations can be used to estimate the properties of an optimal filter that maximizes the transmitted information (van Hateren, 1992a; Hermundstad et al., 2014). Notably, the scale-invariance of natural images (Field, 1987; Stephens et al., 2013) implies that the outcome of this first step will not be strongly dependent on the resolution of the images (i.e., the size of a pixel); indeed, this prediction has been partially tested by Hermundstad et al. (2014), who reported results that were consistent independently of the scale of the initial block-averaging operation in their image-processing pipeline.

---

1 If patches are chosen randomly, the assumption of constant texture will sometimes be violated. This should however be relatively infrequent if the image patches are not too large.
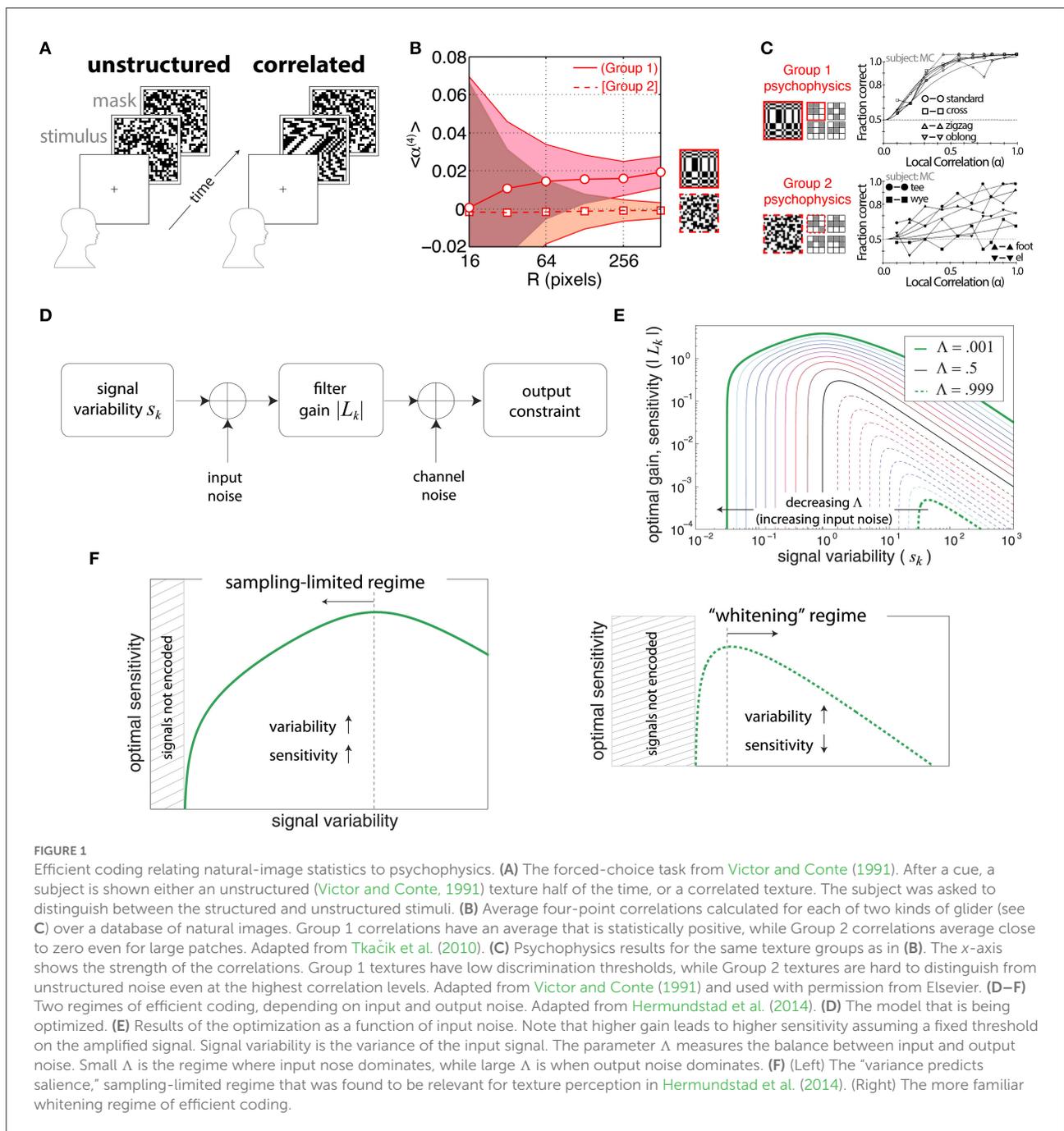
FIGURE 1
Efficient coding relating natural-image statistics to psychophysics. **(A)** The forced-choice task from Victor and Conte (1991). After a cue, a subject is shown either an unstructured (Victor and Conte, 1991) texture half of the time, or a correlated texture. The subject was asked to distinguish between the structured and unstructured stimuli. **(B)** Average four-point correlations calculated for each of two kinds of glider (see **C**) over a database of natural images. Group 1 correlations have an average that is statistically positive, while Group 2 correlations average close to zero even for large patches. Adapted from Tkačik et al. (2010). **(C)** Psychophysics results for the same texture groups as in **(B)**. The x-axis shows the strength of the correlations. Group 1 textures have low discrimination thresholds, while Group 2 textures are hard to distinguish from unstructured noise even at the highest correlation levels. Adapted from Victor and Conte (1991) and used with permission from Elsevier. **(D–F)** Two regimes of efficient coding, depending on input and output noise. Adapted from Hermundstad et al. (2014). **(D)** The model that is being optimized. **(E)** Results of the optimization as a function of input noise. Note that higher gain leads to higher sensitivity assuming a fixed threshold on the amplified signal. Signal variability is the variance of the input signal. The parameter Λ measures the balance between input and output noise. Small Λ is the regime where input nose dominates, while large Λ is when output noise dominates. **(F)** (Left) The "variance predicts salience," sampling-limited regime that was found to be relevant for texture perception in Hermundstad et al. (2014). (Right) The more familiar whitening regime of efficient coding.

A natural experimental approach for checking the predictions of efficient coding models is to measure the neural or behavioral responses of human or animal subjects to patches of known textures. The preferred method for obtaining texture patches is to generate them artificially, as this avoids biases introduced by contextual information that might be available in patches cropped from natural images (Julesz, 1962). This method also scales more easily to very large sample sizes, which are needed to analyze higher-order statistics. Texture

generation can be time consuming, but a variety of powerful algorithms have been developed for the class of textures that we are considering here (Victor and Conte, 2012; Piasini, 2021).

Victor and Conte (1991) used a forced-choice task to determine psychophysical thresholds for discriminating certain binary textures involving fourth-order correlations from unstructured textures (Figure 1A) with independent, identically distributed pixels. The fourth-order correlations were defined in terms of the relative positions of the four points involved

in the correlation, shown pictorially as a *glider* in Figure 1C. Tkačik et al. (2010) used the same gliders to analyze correlations in natural scenes, drawn from a database of pictures taken in the African savannah (Tkačik et al., 2011). They found that the most informative[2] natural-image correlations matched the texture dimensions that were more salient in the psychophysical trials. Conversely, the correlations that were least informative in the natural-image database were the ones that human subjects had difficulty distinguishing from unstructured noise (compare Figures 1B,C)[3].

Note that we are concerned with the *correlations* between pixels with relative positions determined by a glider. For example, if we want to consider a four-point correlation in which the four pixels are arranged in a certain way we represent it as a glider as in Figure 1. But drawing a glider in this way does not imply that the pixels in the glider must all have the same luminance, or indeed any other particular pattern of brightness. For example, a vertical two-point correlation could be significantly different from 0 even if a pattern consisting of two bright pixels arranged on top of each other never occurred in the image patch. This can for instance happen in a patch of alternating bright and dark pixels, which would lead to a highly negative vertical two-point correlation.

A later analysis (Hermundstad et al., 2014) focused on just one of the gliders identified as salient in Victor and Conte (1991) and Tkačik et al. (2010), the 2×2 glider shown with a solid red outline in Figure 1C, along with correlations between pairs and triples of pixels within this glider. In this study, a comparison between natural-image statistics and psychophysics revealed that textures that exhibited higher variance in natural scenes were also more perceptually salient—in other words, "variance predicts salience" (Hermundstad et al., 2014). This effect can be explained as a result of the presence of significant input (sampling) noise compared to output noise (Hermundstad et al., 2014; Figures 1D–F). We can understand why this occurs in the simple example of linear efficient coding models. In this case, salience is related to the gain factor associated with each stimulus—a higher gain factor leads to smaller detection thresholds. When the main source of noise occurs at the output, amplification increases the signal-to-noise ratio. Then, if the total output power is fixed, the most efficient encoding whitens the signal—this is the scenario that has been studied most extensively. Hermundstad et al. (2014) noted, however, that

when the *input* is corrupted by sampling noise, the signal-to-noise ratio can no longer be improved by amplification. In this regime, the best strategy is to de-emphasize the low-variance signals, which are corrupted by noise and thus not very informative, and instead use high gain factors for the high-variance, more informative signals. This highlights the importance of considering efficient-coding ideas that go beyond redundancy reduction by whitening.

In greater detail, the work from Hermundstad et al. (2014) used a four-alternative forced-choice (4AFC) design to analyze the sensitivity of human subjects in multiple directions in the space of binary textures with up to four-point correlations contained within 2×2 blocks (Figure 2A). The relevant texture space was 10-dimensional (Victor and Conte, 2012), and the psychophysical trials assayed all of these dimensions (Figure 2B), as well as pairs of dimensions (Figure 2C). In all these cases, the variance of the correlations in natural images was an excellent predictor of the perceptual salience of the corresponding textures in psychophysical trials.
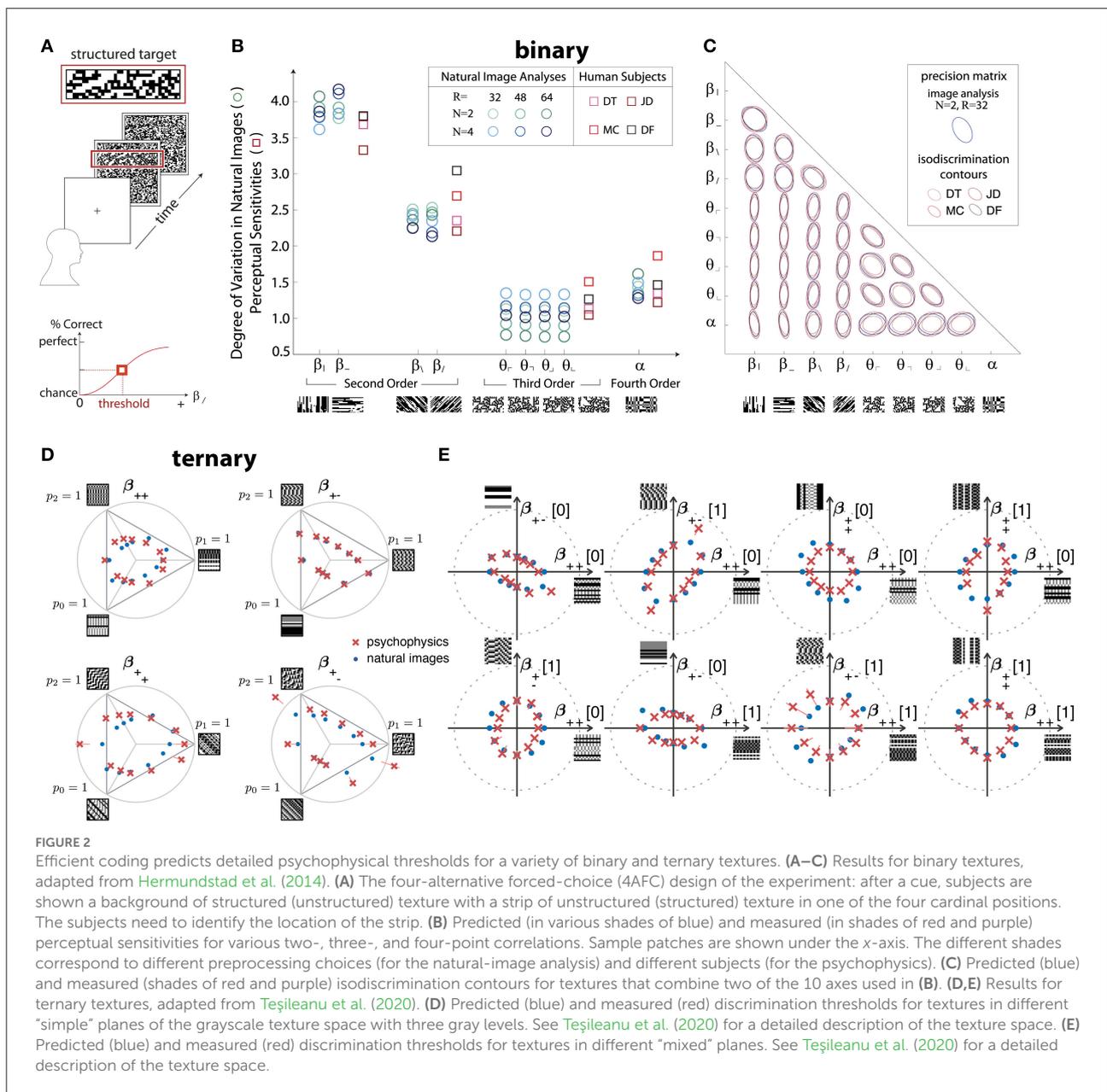
Going beyond binary luminance is challenging because, in this framework, the dimensionality of texture space for a four-pixel glider grows as the fourth power of the number of gray levels (Victor and Conte, 2012). In follow-up work, Teşileanu et al. (2020) introduced a new parameterization of the texture space for $G > 2$ gray levels and focused on textures with three gray levels (ternary textures). Using the same psychophysics paradigm as Hermundstad et al. (2014), they probed more than 300 different rays in the resulting 66-dimensional texture space. Their results found an excellent match between most predicted and observed discrimination thresholds, in agreement with the earlier results for binary textures (Figures 2D,E).

Interestingly, the results from Teşileanu et al. (2020) also showed glimpses of limitations of the efficient coding idea: the prediction errors in a few texture-space directions were much larger than in the other directions (Figure 2D). This effect appears to be related to symmetries that act differently on natural-image ensembles as compared to human texture sensitivity. While the precise meaning behind the mismatches is not yet clear, general considerations suggest that we should expect deviations from the simplest forms of efficient coding because of resources limitations, alternative criteria that the brain might optimize or balance, etc. Of course, resource limitations can be incorporated into a more general information maximization problem *via* additional constraints, and other considerations like the utility of information for behavior can likewise be addressed. We return to these more general formulations of efficient coding principle in the Discussion.

The methods described above use a simplified texture space—a small number of gray levels, correlations constrained to small neighborhoods—in order to facilitate texture generation. Correspondingly, the resulting textures capture only a small fraction of the complexity of natural textures. A different approach is to exhaustively search for the statistics that are

---

2 We call a certain multi-point correlation informative if knowing its value significantly reduces our uncertainty about the underlying luminance pattern *in the natural image ensemble that we are considering*.
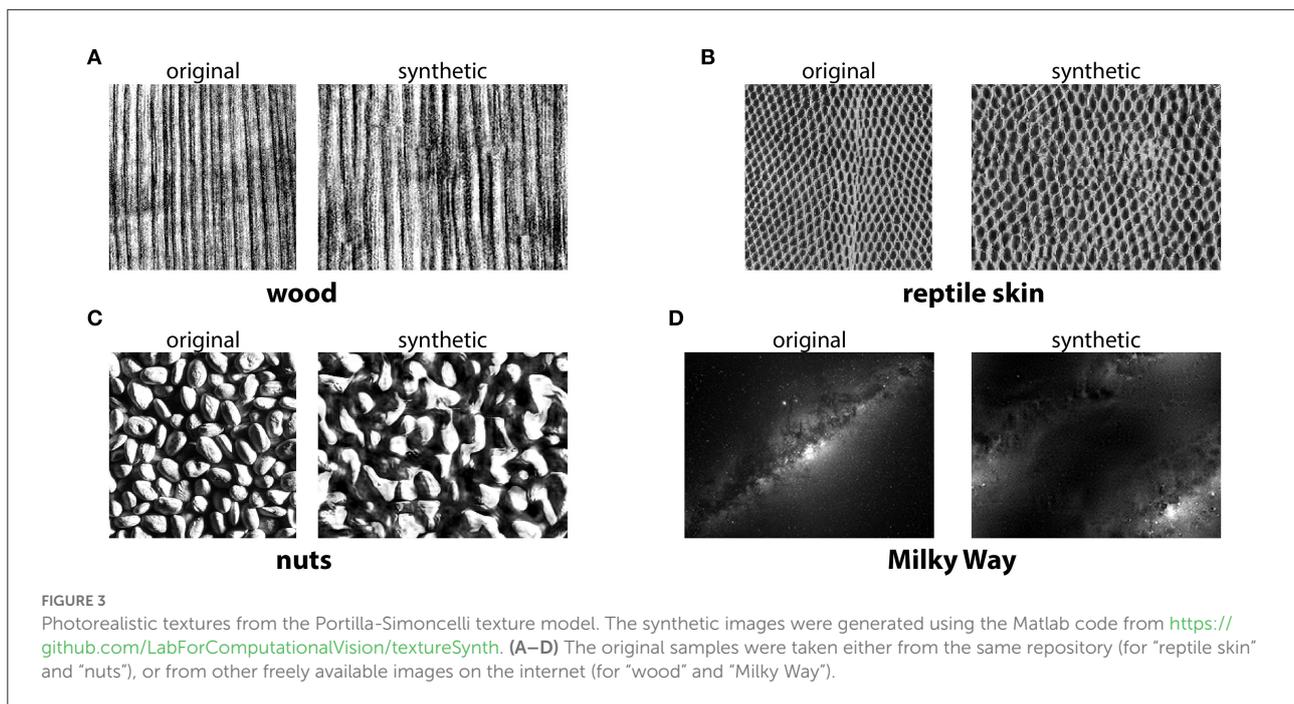
3 Note that the naming of texture group 2 differs between Victor and Conte (1991) and Tkačik et al. (2010): group 2 in the former refers to group 3 in the latter. Group 2 from Victor and Conte (1991) is not used in Tkačik et al. (2010).

FIGURE 2
Efficient coding predicts detailed psychophysical thresholds for a variety of binary and ternary textures. **(A–C)** Results for binary textures, adapted from Hermundstad et al. (2014). **(A)** The four-alternative forced-choice (4AFC) design of the experiment: after a cue, subjects are shown a background of structured (unstructured) texture with a strip of unstructured (structured) texture in one of the four cardinal positions. The subjects need to identify the location of the strip. **(B)** Predicted (in various shades of blue) and measured (in shades of red and purple) perceptual sensitivities for various two-, three-, and four-point correlations. Sample patches are shown under the x-axis. The different shades correspond to different preprocessing choices (for the natural-image analysis) and different subjects (for the psychophysics). **(C)** Predicted (blue) and measured (shades of red and purple) isodiscrimination contours for textures that combine two of the 10 axes used in **(B)**. **(D,E)** Results for ternary textures, adapted from Teşileanu et al. (2020). **(D)** Predicted (blue) and measured (red) discrimination thresholds for textures in different "simple" planes of the grayscale texture space with three gray levels. See Teşileanu et al. (2020) for a detailed description of the texture space. **(E)** Predicted (blue) and measured (red) discrimination thresholds for textures in different "mixed" planes. See Teşileanu et al. (2020) for a detailed description of the texture space.

needed to generate a texture that is as indistinguishable as possible from its natural counterpart (Victor et al., 2017). Very powerful methods in this direction include the multiscale techniques proposed by Portilla and Simoncelli (2000) and recent approaches based on deep learning (Gatys et al., 2015; Ustyuzhaninov et al., 2016; Ding et al., 2020; Park et al., 2020). See Figure 3 for an example of photorealistic textures.

There are two fundamental downsides of more realistic texture models (Victor et al., 2017). First, texture generation can be significantly more time consuming—though perhaps this is less of an issue with modern hardware. Second, texture space can

be much harder to describe and navigate: for instance, the model from Portilla and Simoncelli (2000) uses a parameterization in which many combinations of parameter values are not valid. This limitation can, however, be circumvented by reparameterizing the texture space in the vicinity of the manifold defined by naturally occurring textures (Lüdtke et al., 2015). Psychophysical studies using the textures from Portilla and Simoncelli (2000) suggest that peripheral vision is well-adapted to the distribution of such textures in natural scenes (Balas et al., 2009). A particularly striking illustration of this involves natural images that are modified using the texture model from Portilla and Simoncelli (2000) but look the same as the originals as long

FIGURE 3
Photorealistic textures from the Portilla-Simoncelli texture model. The synthetic images were generated using the Matlab code from https://github.com/LabForComputationalVision/textureSynth. **(A–D)** The original samples were taken either from the same repository (for "reptile skin" and "nuts"), or from other freely available images on the internet (for "wood" and "Milky Way").

as the changes occur only in the visual periphery (Freeman and Simoncelli, 2011).

Yet a different approach is to use the symmetries of a specific texture model to generate new patches starting from texture patches cropped from natural images (Gerhard et al., 2013). If the model is accurate, the generated patches should be indistinguishable from natural textures as far as human observers are concerned. Conversely, if human subjects can distinguish between the natural and generated textures, the model must be incomplete. While all the models investigated by Gerhard et al. (2013) were incomplete in this sense, the discrimination performance was high for models that poorly fit natural-scene textures and low for models that provided a better fit. This suggests that the better models—the ones that yielded generated textures that were hardest to distinguish from natural textures—are the ones that are better adapted to natural-scene statistics, in agreement with efficient-coding ideas.

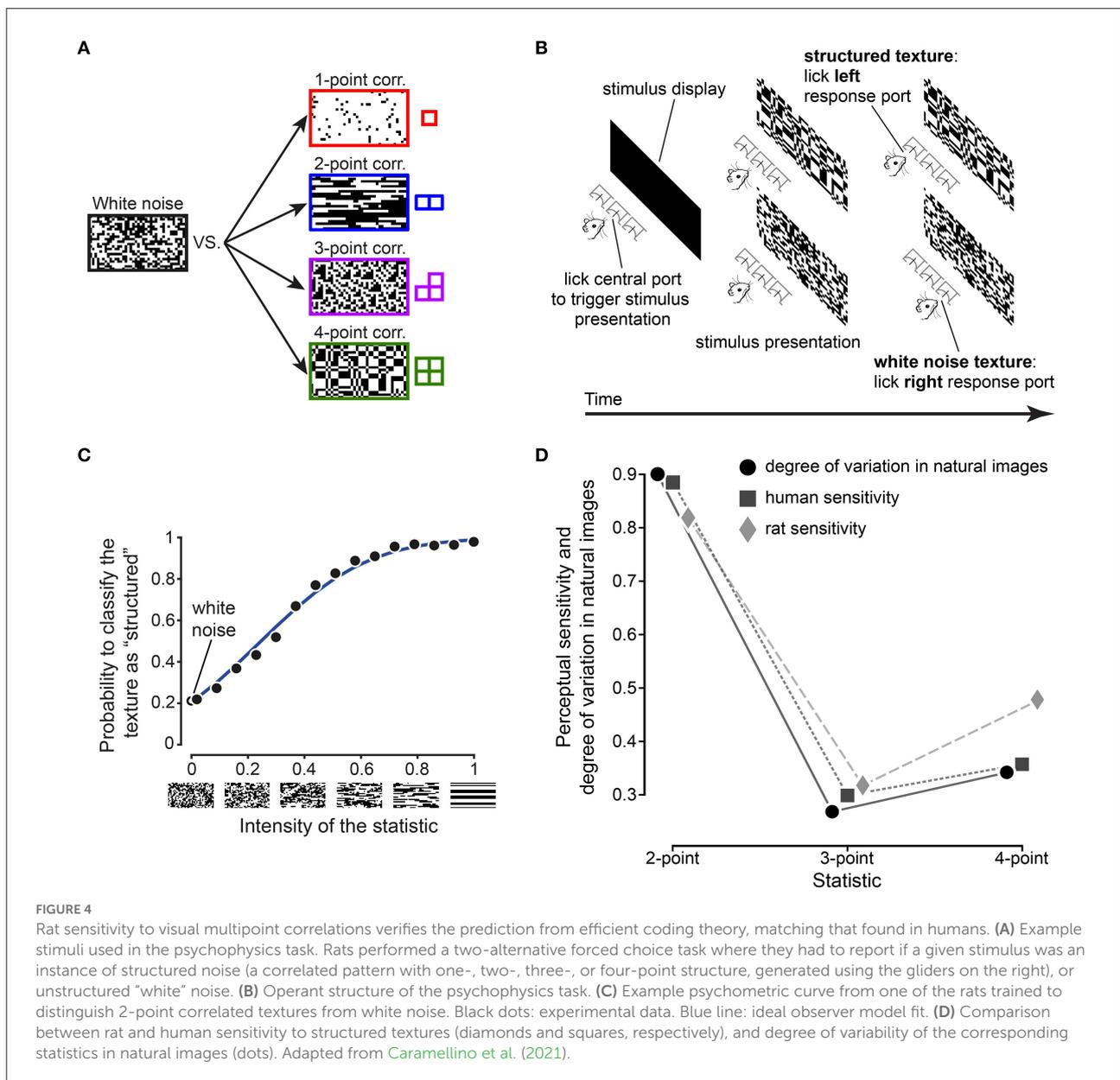## 3. Extending to other animal species

In the previous sections we have described several experiments that illustrate how human vision, and in particular texture perception, is adapted to the statistics of natural visual scenes. Despite the power and flexibility afforded by human psychophysics, this approach also has limitations. For instance, investigation into the neural mechanisms for texture representation in humans rely on techniques such as fMRI (Beason-Held et al., 1998), as invasive recordings are not possible in humans. Developmental studies in humans also have

very tight operational constraints (but see Gervain et al., 2021 for encouraging first steps), and of course causal manipulation of the developmental process is excluded.

In order to overcome these constraints, it is necessary to investigate perception and neural coding of visual textures in other animals.

Studies in macaque have investigated the encoding of multipoint correlations in visual cortex (Purpura et al., 1994; Yu et al., 2015), showing that representations of three- and four-point correlated patterns emerge prominently at the single cell level in V2 (Yu et al., 2015). However, dedicated psychophysical tests of the prediction of efficient coding theory are not available in these animals. In recent years, rodents have gained popularity as model systems for the study of vision. Rodents allow for running experiments with larger number of animals, and the anatomical layout of rodent visual cortex makes it easier to perform simultaneous recordings from distinct cortical areas contributing to the processing of visual stimuli (Glickfeld and Olsen, 2017; Tafazoli et al., 2017; Piasini et al., 2021). In particular, rats have been used successfully to investigate high-level processing involved in object recognition (Zoccolan et al., 2009; Tafazoli et al., 2017; Djurdjevic et al., 2018; Piasini et al., 2021). Moreover, unlike monkeys, rats are amenable to altered-rearing experiments, which were used to reveal how elementary coding features of primary visual cortex are adapted to the statistics of visual stimuli during development (Matteucci and Zoccolan, 2020). Overall then, rats are a convenient animal model for exploring efficient coding of visual textures.

The first step in such an investigation is to establish whether rodents (and rats in particular) exhibit the same pattern of

FIGURE 4
Rat sensitivity to visual multipoint correlations verifies the prediction from efficient coding theory, matching that found in humans. **(A)** Example stimuli used in the psychophysics task. Rats performed a two-alternative forced choice task where they had to report if a given stimulus was an instance of structured noise (a correlated pattern with one-, two-, three-, or four-point structure, generated using the gliders on the right), or unstructured "white" noise. **(B)** Operant structure of the psychophysics task. **(C)** Example psychometric curve from one of the rats trained to distinguish 2-point correlated textures from white noise. Black dots: experimental data. Blue line: ideal observer model fit. **(D)** Comparison between rat and human sensitivity to structured textures (diamonds and squares, respectively), and degree of variability of the corresponding statistics in natural images (dots). Adapted from Caramellino et al. (2021).

sensitivity to visual multipoint correlations as humans. In a recent study, Caramellino et al. (2021) designed a psychophysics task inspired by an experiment in humans (Victor and Conte, 2012; Hermundstad et al., 2014), and used it to probe rat sensitivity to visual textures. Briefly, rats were trained on a two-alternative forced choice task (2AFC), where a visual texture was presented on a monitor and the animal had to report if the texture was a sample of unstructured "white" noise (each pixel black or white with equal probability, independently from its neighbors), or if it was a sample from a maximum-entropy distribution with a nonzero level of one of four multipoint correlations (Figures 4A,B). A separate group of rats was trained for each type of correlation (one-, two-, three-, and four-point).

The rats' behavioral performance was interpreted with the help of an ideal observer model (Figure 4C) tailored to the task design, which allowed estimation of the animal's perceptual sensitivity. This estimate matched the sensitivity measured in humans in the work described above (Hermundstad et al., 2014) as well as the degree of variability of multipoint correlations in natural images (Figure 4D). These results therefore show that texture perception in rats is adapted to the statistics of natural stimuli, in a way that closely matches both the prediction of efficient coding theory and analogous results in humans.

The results discussed here have opened the way to studying how higher-order visual properties are efficiently encoded in the brain. Rapid progress is now being made through the

study of rodent vision. In a very recent preprint, Bolaños et al. (2022) present data linking texture perception and neural representation in mice, arguing that—compatibly with Yu et al. (2015) in macaque—useful representations of visual textures emerge in area LM (the rodent analog of V2). Importantly, the geometry of such representation seems predictive of the animals' behavioral performance in a visual discrimination task (texture vs. non-structured visual stimuli), and analogous results are reported for artificial neural networks. Together with Caramellino et al. (2021), this work is representative of a trend that uses texture perception as a tool to investigate vision in a broad set of species, including birds (Gervain et al., 2021) and even invertebrates (Reiter and Laurent, 2020).

4. Adaptation to temporal statistics and object recognition

The efficient coding principle as described above says that neural circuits are adapted to maximize the behaviorally relevant information they transmit, subject to appropriate constraints. While much of the research using this idea has focused on spatially organized information, adaptation to the temporal structure of stimuli as they arrive at the retina is also essential. This can occur on multiple timescales, from evolutionary adaptations to short-term sensory plasticity. In fact, some authors have proposed that long-range pairwise correlations in the luminance of natural scenes are not mainly removed in the retinal response by center-surround spatial receptive fields as usually presumed (Barlow et al., 1961; Atick and Redlich, 1990; van Hateren, 1992b; Barlow, 2001; Balasubramanian and Sterling, 2009), but rather by fixational eye movements and the resulting temporal effects on the retinal image (Kuang et al., 2012). More generally, the selection of information that is behaviorally relevant for an animal certainly involves temporal statistics, and may also involve state-dependent processes such as the influence of prior expectations, or feedback from the central brain. Indeed, already in the retina, the phenomena of motion anticipation and lag normalization (Berry et al., 1999; Trenholm et al., 2013) involve circuits that predict and use expected temporal regularities to adjust or enhance responses, and many of the nonlinear response features of retinal ganglion cells described in Gollisch and Meister (2010) also involve such effects. To ask whether these aspects of retinal coding can be understood in terms of efficient coding, we would ultimately need to identify how much utility they contribute to behavior, and not just how much information they convey. Perhaps the best-studied example of reformatting sensory information to make behaviorally-relevant quantities accessible occurs centrally, in the neural circuits of the visual cortex that represent object identity. Numerous theoretical works have proposed that useful object representations can be

learned without supervision by adapting to statistical regularities of natural stimuli. We will now review these ideas and the associated challenges of accounting for state-dependent neural processing.

4.1. Unsupervised learning of invariant object representations

An important evolutionary advantage conferred by vision (Striedter and Northcutt, 2020) is the ability to detect and identify the objects in a visual scene. Accordingly, animals perform this task remarkably well by the standards of modern computer vision techniques. The visual ventral stream is a hierarchy of areas in the visual cortex of mammals that, to a first approximation, process object identity and encode aspects of the visual world in increasing order of abstraction—starting from simple features like edges in V1, to neurons in area IT (in primates) which are selective for object identity, and whose response is largely invariant to changes in location, point of view, illumination, and contrast (DiCarlo et al., 2012; Bao et al., 2020). More generally, such *transformation-invariant representations* have been recognized as a fundamental building block of an efficient object recognition system that can learn new categories from a limited number of examples (Anselmi and Poggio, 2014).

But how are these representations built by the brain? Behavioral (Cox et al., 2005) and electrophysiological (Li and DiCarlo, 2008; Matteucci and Zoccolan, 2020) evidence suggests that they are not hard-wired by evolution, but are at least partially the result of adaptation to the contingent spatio-temporal statistics of the natural world. Specifically, it has been proposed that the visual system learns to "discount" changes in size, pose, illumination etc. by exploiting the *temporal persistence* of objects encountered in visual scenes.

Indeed, the identity of the objects composing a scene typically possesses some degree of persistence from one moment to the next, while the low-level details of the impression those objects leave on a sensory system might change from moment to moment following changes in their position and configuration, or in the environment (e.g., illumination). Hence, to a first approximation, it may be reasonable for a sensory area to assume that the input patterns received from sensory transduction at two successive instants represent the same objects, modulo a set of transformations that do not affect object identity. This is the class of transformations that we wish neural representations to be invariant to, if the goal is to perform object recognition.
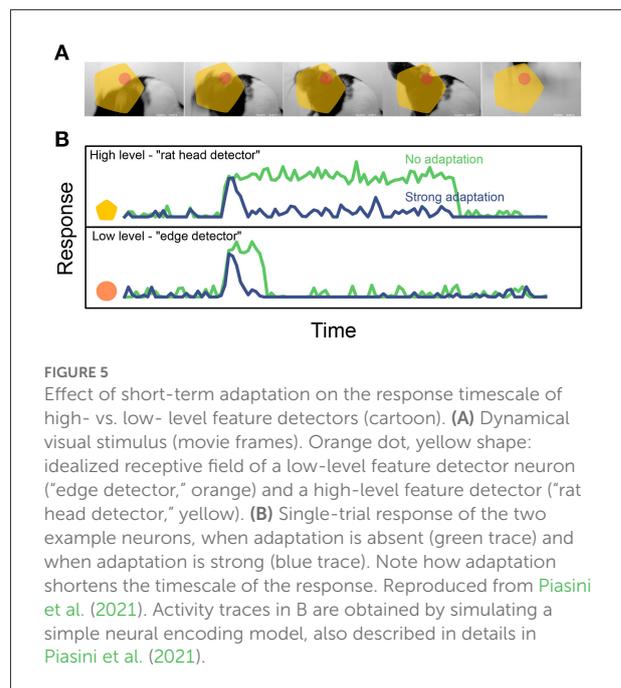
Different theories have been proposed for the mechanism of such adaptation. Földiák (1991) introduced a local synaptic learning rule which could lead to translation-invariant responses in a schematic model of complex cells in visual cortex. This line of inquiry was expanded in later years, leading to sophisticated models of representation learning across multiple areas of visual

cortex (see for instance Wallis and Rolls, 1997; Masquelier et al., 2007). Another approach, at a higher abstraction level, is due to Wiskott and Sejnowski (2002), who introduced the Slow Feature Analysis (SFA) algorithm. For a data stream $X = \{x_0, x_1, \ldots, x_{T-1}, x_T\}$, where $x_t \in \mathbb{R}^n$ is the content of the stream at time t, SFA finds a number of scalar features of the data $f^k = f^k(x)$ with maximal "slowness," in the sense that $(df^k/dt)^2$ is minimized on average over the stream (under certain constraints that rule out trivial solutions). Notably, the features f^k are restricted to be instantaneous functions of the data—that is, $f_t^k = f^k(x_t)$, and do not depend explicitly on values of x at other times. The algorithm is therefore not allowed to resort to temporal filtering and must discover slowly-varying features that are computable from any "snapshot" of the data, for instance the identity of an object in a video stream that shows the object moving around in the visual field. SFA can be applied recursively to its own outputs, thus generating a representational hierarchy of increasing abstraction. Such a hierarchy could then form the backbone of a model of the visual system (Wiskott and Sejnowski, 2002; Körding et al., 2004; Einhäuser et al., 2005; Wyss et al., 2006; Franzius et al., 2007), or more generally of any object-recognition system, regardless of the sensory modality (DiTullio et al., 2022).

4.2. State-dependent processing

The models discussed above treat sensory information encoding as a feedforward cascade of stateless functions: at each point in time, the retinal input gets processed by a sequence of stages in a way that is independent of previous or later signals. In recent years, this approach to modeling the ventral visual stream has yielded impressive results, particularly by leveraging deep convolutional neural networks which allowed for accurate prediction (Yamins et al., 2014; Schrimpf et al., 2020) and even causal manipulation (Bashivan et al., 2019) of neuronal activity. Indeed, at least in rodents, convolutional neural networks have been shown to reformat visual information similarly to the ventral stream, according to notions of intrinsic dimensionality of population representations and of single cell-level distillation of elementary image features such as luminosity, contrast, edge orientation, and presence of corners (Muratore et al., 2022). However, phenomena such as short-term adaptation, or circuit features such as recurrent or feedback connectivity, can introduce state or history dependence in neural computations, supporting fundamentally different modes of cortical operation based on transient dynamics (Buonomano and Maass, 2009) or predictive processing (Rao and Ballard, 1999; Keller and Mrsic-Flogel, 2018).

More generally, we call any information processing mode *state-dependent* if the output of a circuit at time t depends not only on the input at that time, but also on previous values of the input and of the output itself. Experimental

FIGURE 5
Effect of short-term adaptation on the response timescale of high- vs. low- level feature detectors (cartoon). **(A)** Dynamical visual stimulus (movie frames). Orange dot, yellow shape: idealized receptive field of a low-level feature detector neuron ("edge detector," orange) and a high-level feature detector ("rat head detector," yellow). **(B)** Single-trial response of the two example neurons, when adaptation is absent (green trace) and when adaptation is strong (blue trace). Note how adaptation shortens the timescale of the response. Reproduced from Piasini et al. (2021). Activity traces in B are obtained by simulating a simple neural encoding model, also described in details in Piasini et al. (2021).

evidence points to the widespread existence of state-dependent processing and of circuit features that can support it in visual cortex. For instance, neural coding of visual stimuli depends on the behavioral context (Niell and Stryker, 2008; Khan and Hofer, 2018), highlighting the existence of feedback connections projecting from other parts of the brain; cortico-cortical or cortico-thalamic feedback is also compatible with experimental observations (Lamme et al., 1998; Issa et al., 2018; Marques et al., 2018). The effects of short-term adaptation can be seen in the reduction of the responses to repeated stimuli or in those to continuous versus transient stimuli, two phenomena that may increase in intensity along the ventral stream (Grill-Spector et al., 2006; Kohn, 2007; Kaliukhovich et al., 2013; Webster, 2015; Stigliani et al., 2019; Fritsche et al., 2020). It is interesting to note that these ideas are now also starting to influence the design of artificial systems, e.g., recent deep neural network architectures that extend classic convolutional networks with the addition of recurrent and adaptive elements (Tang et al., 2018; Kar et al., 2019; Kreiman and Serre, 2020; van Bergen and Kriegeskorte, 2020; Vinken et al., 2020).

Whatever their functional roles, state-dependent mechanisms can have dramatic effects on the temporal dynamics of neural codes. For instance, imagine a visual stimulus containing an object undergoing identity-preserving transformations, such as the rat moving its head in Figure 5. In absence of state-dependent processing, a neuron that is selective for the presence of a rat head will fire continuously as long as the rat head is within the field of view. On the other hand a different neuron, selective for a low-level image feature such as the presence of an oriented edge within a small region,

would only fire briefly whenever the correct stimulus enters its receptive field. As discussed above, this should result in a *slower* encoding for the high-level feature than for the lower-level one (Figure 5B, green). However, if state-dependence is now added in the form of a simple short-term adaptation mechanism, the difference in timescale between these hypothetical neurons can be dramatically reduced, as both units switch to encoding feature onset/offset rather than feature presence (Figure 5B, blue). In this simplified example, there is a tension between metabolic and functional efficiency: adaptation decreases energy consumption by reducing the total activity (using a crude form of predictive coding, where the prediction at each point in time is that the stimulus will persist in the current state), but makes it harder for the system to build invariant representations of the stimuli, which are advantageous on functional grounds.

Another way in which state-dependent processing can affect the timescale of neural codes is by acting on *intrinsic* timescales. Intrinsic timescales describe the temporal extent over which fluctuations around the average response to a given stimulus are correlated (whenever it is necessary to distinguish them from the timescales of a stimulus' neural representation, we will call the latter *response* timescales). Operationally, multiple definitions of intrinsic timescales are possible, but for concreteness we will use the following Piasini et al. (2021). If $X_t = \{x_k^{(t)}\}_{k=1}^K$ is the activity of one neuron at time t recorded over K identical repetitions of the experiment (trials), the *intrinsic correlation* at time lag Δ is the average correlation coefficient between the activity at time X_t and $X_{t+\Delta}$:

$$C(\Delta) = \frac{1}{T-\Delta} \sum_{t=1}^{T-\Delta} \frac{\mathrm{Cov}\left[X_t, X_{t+\Delta}\right]}{\sqrt{\mathrm{Var}\left[X_t\right] \mathrm{Var}\left[X_{t+\Delta}\right]}},$$

where T is the duration of the recording, such that $1 \leq t \leq T$. The intrinsic timescale is then a measure of the characteristic time over which C decays as Δ grows. In this sense, the intrinsic timescale captures the temporal range over which "temporal noise correlations" are present. Intrinsic timescales are thought to increase along cortical hierarchies (Murray et al., 2014; Runyan et al., 2017; Wang, 2022), possibly reflecting an increase in the importance of certain classes of state-dependent processes, such as temporal integration or more complex dynamics emerging from recurrent connectivity (Chaudhuri et al., 2015; Piasini et al., 2021).

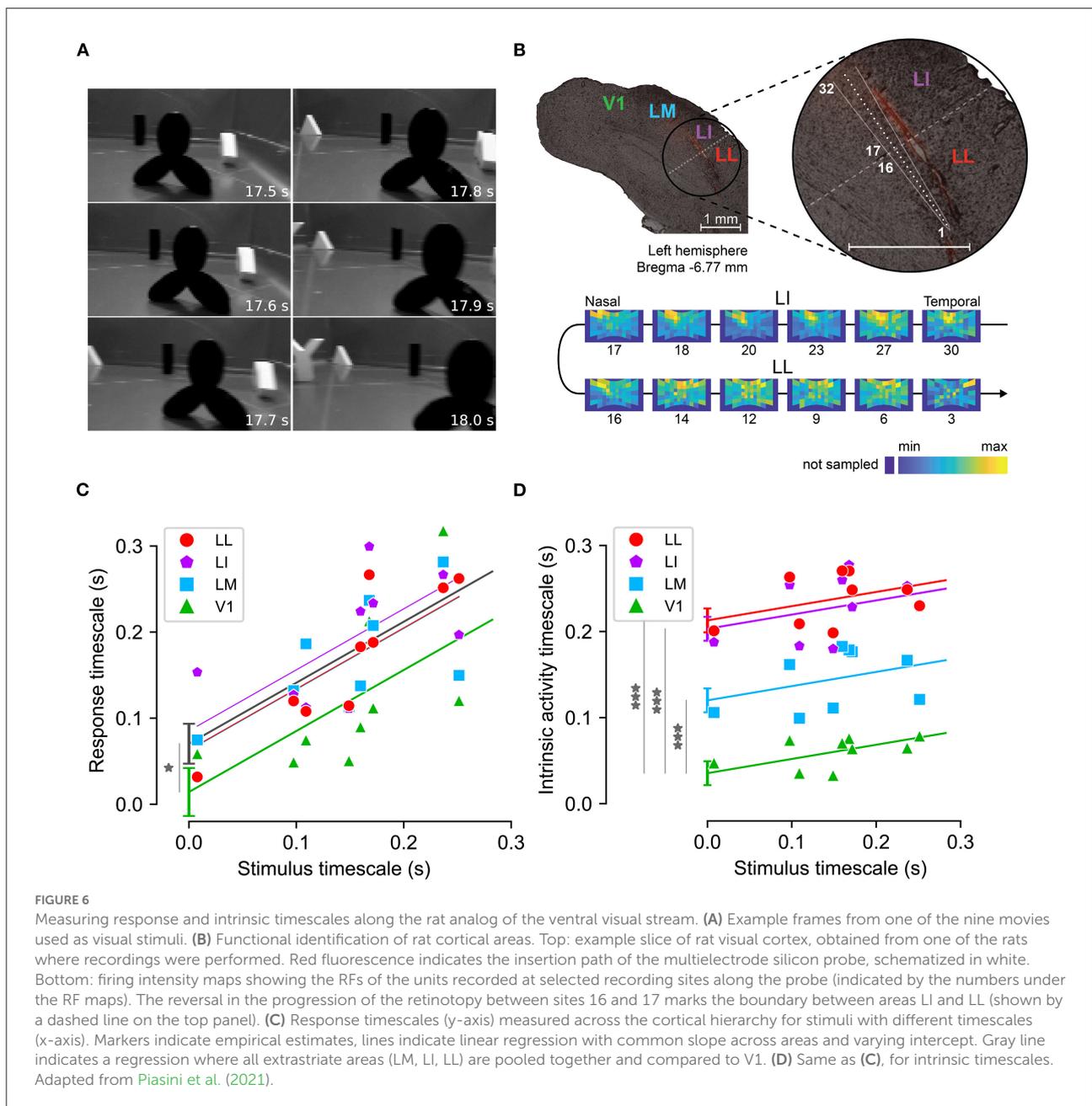## 4.3. Response and intrinsic timescales increase along the visual cortical pathway

While the classic view of invariant representations in the ventral stream suggests a straightforward increase of neural timescales along the hierarchy, the discussion above highlights that the existence of state-dependent mechanisms implies a more complex picture. To gain some insight into this matter,

Piasini et al. (2021) performed an empirical study of the timescales of visual cortical representations of dynamic stimuli (Figure 6).

Piasini et al. (2021) recorded the activity elicited by the presentation of dynamic visual stimuli (movies) in four areas of rat visual cortex, which form the rodent analog of the visual ventral stream (Vermaercke et al., 2014; Glickfeld and Olsen, 2017; Tafazoli et al., 2017; Vinken et al., 2017; Kaliukhovich and Op de Beeck, 2018). Analysis of the data showed that response timescales depended strongly on the timescale of the stimulus, and were significantly larger in extrastriate cortex than in V1 (Figure 6C). This suggests that adaptation and other state-dependent processing mechanisms do not prevent an increase of slowness of the neural code along the hierarchy, a correlate of increasing invariance to identity-preserving transformations. Moreover, analysis of intrinsic timescales revealed a weaker dependence on the timescale of the stimuli (as expected), but a very strong increase along the cortical hierarchy, compatibly with previous results in monkey (Murray et al., 2014) and in behaving mice (Runyan et al., 2017) suggesting that the importance of state-dependent and adaptive processes is also increasing along the cortical hierarchy. These results were confirmed also by re-analysing previously published data recorded in mouse (Siegle et al., 2021) and in awake rat (Vinken et al., 2016).

5. Challenges for the future

We would not expect the efficient coding hypothesis to hold throughout the brain in the elementary form of maximizing information transmission. One fundamental reason for this is that there are statistical and resource limitations. Even if we just consider texture encoding, note that as the complexity of the textures under consideration increases, the dimensionality of the corresponding texture space quickly grows. This leads, on the one hand, to difficulty in sampling the relevant natural-image ensemble: even if two texture dimensions differ in terms of their statistics in natural scenes, the number of samples required to characterize the difference may simply be too large for any realistic organism to achieve the required adaptation. Moreover, even if we solved the statistical-sampling problem, the actual improvements in information transfer obtained by adapting to very high-dimensional texture spaces might not be worth the brain circuitry required to perform the required encoding. Put differently, efficient coding might be limited by logistics. It is of course possible in principle to include resource or sampling limitations as additional constraints in our optimization problem. This would ensure that the efficient coding solution we find obeys those limitations. In practice, however, we might not have a quantitative understanding of some of these limitations. In this case the best we can do is employ the most accurate model we have, but keep in mind

FIGURE 6

Measuring response and intrinsic timescales along the rat analog of the ventral visual stream. **(A)** Example frames from one of the nine movies used as visual stimuli. **(B)** Functional identification of rat cortical areas. Top: example slice of rat visual cortex, obtained from one of the rats where recordings were performed. Red fluorescence indicates the insertion path of the multielectrode silicon probe, schematized in white. Bottom: firing intensity maps showing the RFs of the units recorded at selected recording sites along the probe (indicated by the numbers under the RF maps). The reversal in the progression of the retinotopy between sites 16 and 17 marks the boundary between areas LI and LL (shown by a dashed line on the top panel). **(C)** Response timescales (y-axis) measured across the cortical hierarchy for stimuli with different timescales (x-axis). Markers indicate empirical estimates, lines indicate linear regression with common slope across areas and varying intercept. Gray line indicates a regression where all extrastriate areas (LM, LI, LL) are pooled together and compared to V1. **(D)** Same as **(C)**, for intrinsic timescales. Adapted from Piasini et al. (2021).

that its predictions will be affected by the incomplete set of constraints used in the model.

Apart from logistics, another factor in potential conflict with efficient coding is the fact that animals encode information not with the end goal of information transmission, but in order to perform behaviors that are helpful for their survival. While in many cases, efficiently encoding information will be in alignment with this objective, in other cases different considerations might be more important. For instance, a very efficient encoding that is extremely hard to decode (or turn into helpful behavior) might not be very useful. We therefore

see another important way in which we expect simplistic information-optimizing efficient coding to only be part of the story: organisms need to fulfill other roles apart from encoding information. Indeed, it would be useful to develop a formalism which combines coding efficiency, computational realizability and effectiveness, and, critically, behavioral goals in a more complete normative theory of neural circuit organization.

Despite these caveats, all sensory systems must adapt in some way to the statistics of their natural inputs in order to perform well. Below we describe three promising directions for the future.

5.1. Explaining shape sensitivity from object shape distributions

The success of the efficient coding approach to texture processing in the brain suggests another question: can we explain the distribution of cells responsive to different kinds of shapes in the ventral visual pathway? Studies have shown that individual cells in V4 are responsive to fragments with different shapes and curvatures (Pasupathy and Connor, 1999, 2001, 2002). Likewise cells in IT (and its rodent analog) are selective for specific objects, again with different numbers of edges, corners, and curvatures (Tanaka, 1996; Hung et al., 2005; Rust and DiCarlo, 2010). The degree of invariance to image transformations also increases with depth in the visual pathway (Riesenhuber and Poggio, 1999; DiCarlo and Cox, 2007; Rust and DiCarlo, 2010; DiCarlo et al., 2012; Tafazoli et al., 2017). Perhaps the statistics of these responses are adapted to the distribution of shapes and shape fragments in natural scenes. Studies have demonstrated distinctive shape and curvature statistics in natural images, and have connected these statistics to visual perception (Geisler et al., 2001; Geisler and Perry, 2009). It would be interesting to directly relate these sorts of visual scene statistics to the neural circuits in the ventral visual pathway.

5.2. Explaining cell type distributions in parallel information pathways

Another outstanding challenge is to explain the structure of the parallel pathways that appear in many parts of the brain, where an information stream is processed by multiple cell types, each selective for part of the stream, and then transmitted through parallel fibers in a nerve tract to other parts of the brain (Perge et al., 2012). It is possible that efficient coding can provide a theory of such decompositions (Balasubramanian, 2015). The retina in particular is a tempting target for such an analysis as it presents a classic feedforward neural network architecture that winnows down the vast amount of data in the incident photons into an Ethernet cable's worth of information transmitted by about 20 parallel ganglion cell channels (Koch et al., 2006; Perge et al., 2009; Balasubramanian, 2015). To take such an approach, it will likely be important to include constraints associated with the *decoder*—i.e., the region of the brain that must use limited computational resources to rapidly read the information encoded in the incident parallel pathway. Indeed, recent work (Gjorgjieva et al., 2019) suggests that functional diversity in sensory neurons can be understood by balancing the mutual information between stimuli and responses against the error incurred by computationally constrained decoders. It would be interesting to understand if the repertoire of nonlinear feature detectors in

the retina (Gollisch and Meister, 2010) can be understood in this way.

5.3. Methods based on machine learning complementing more normative approaches

Finally, the advent of deep learning may provide an interesting new approach to understanding the logic of neural circuits deeper in the brain, where the guiding principle is that circuits beyond the sensory periphery must self-organize through local learning rules to achieve whatever tasks are behaviorally necessary. Indeed, some authors have suggested that the hierarchy of visual cortical areas should be understood in analogy with the layers of a deep network (Yamins et al., 2014; Yamins and DiCarlo, 2016; Schrimpf et al., 2020; Muratore et al., 2022; Nayebi et al., 2022). Of course these circuits ultimately operate on inputs drawn from the natural world, and hence should adapt through learning to both scene statistics and the target task. This approach has shed light on the presence of grid cells in the entorhinal cortex (Banino et al., 2018; Cueva and Wei, 2018; Sorscher et al., 2019; Cueva et al., 2020) and on the repertoire of retinal ganglion cells (McIntosh et al., 2016). This perspective also raises the intriguing possibility that circuits in the brain are not just organized to encode information efficiently, but also to learn efficiently (Teşileanu et al., 2017).

Author contributions

TT, EP, and VB conceived of and performed the research described in this paper. All authors contributed to the conception of the article and the writing. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Angeloni, C. F., Młynarski, W., Piasini, E., Williams, A. M., Wood, K. C., Garami, L., et al. (2021). Cortical efficient coding dynamics shape behavioral performance. *bioRxiv [Preprint]*. doi: 10.1101/2021.08.11.455845

Anselmi, F., and Poggio, T. (2014). *Representation Learning in Sensory Cortex: A Theory*. Technical Report 26, CBMM.

Atick, J. J., and Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Comput*. 2, 308–320. doi: 10.1162/neco.1990.2.3.308

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab*. 21, 1133–1145. doi: 10.1097/00004647-200110000-00001

Balas, B., Nakano, L., and Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis*. 9, 1–18. doi: 10.1167/9.12.13

Balasubramanian, V. (2015). Heterogeneity and efficiency in the brain. *Proc. IEEE* 103, 1346–1358. doi: 10.1109/JPROC.2015.2447016

Balasubramanian, V. (2021). Brain power. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2107022118. doi: 10.1073/pnas.2107022118

Balasubramanian, V., and Sterling, P. (2009). Receptive fields and functional architecture in the retina. *J. Physiol*. 587, 2753–2767. doi: 10.1113/jphysiol.2009.170704

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., et al. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature* 557, 429–433. doi: 10.1038/s41586-018-0102-6

Bao, P., She, L., McGill, M., and Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature* 583, 103–108. doi: 10.1038/s41586-020-2350-5

Barlow, H. (2001). Redundancy reduction revisited. *Network*. 12, 241. doi: 10.1080/net.12.3.241.253

Barlow, H. B., et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Commun*.

Bashivan, P., Kar, K., and DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science* 364:eaav9436. doi: 10.1126/science.aav9436

Beason-Held, L. L., Purpura, K. P., Krasuski, J. S., Maisog, J. M., Daly, E. M., Mangot, D. J., et al. (1998). Cortical regions involved in visual texture perception: a fMRI study. *Cogn. Brain Res*. 7, 111–118. doi: 10.1016/S0926-6410(98)00015-9

Bell, A. J., and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vis. Res*. 37, 3327–3338. doi: 10.1016/S0042-6989(97)00121-1

Berry, M. J., Brivanlou, I. H., Jordan, T. A., and Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature* 398, 334–338. doi: 10.1038/18678

Bolaños, os, F., Orlandi, J. G., Aoki, R., Jagadeesh, A. V., Gardner, J. L., and Benucci, A. (2022). Efficient coding of natural images in the mouse visual cortex. *bioRxiv [Preprint]*. doi: 10.1101/2022.09.14.507893

Borghuis, B. G., Ratliff, C. P., Smith, R. G., Sterling, P., and Balasubramanian, V. (2008). Design of a neuronal array. *J. Neurosci*. 28, 3178–3189. doi: 10.1523/JNEUROSCI.5259-07.2008

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci*. 10, 113–125. doi: 10.1038/nrn2558

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

Carlson, N. L., Ming, V. L., and DeWeese, M. R. (2012). Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. *PLoS Comput. Biol*. 8:e1002594. doi: 10.1371/journal.pcbi.1002594

Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H., and Wang, X.-J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* 88, 419–431. doi: 10.1016/j.neuron.2015.09.008

Cox, D. D., Meier, P., Oertelt, N., and DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nat. Neurosci*. 8, 1145–1147. doi: 10.1038/nn1519

Cueva, C. J., Wang, P. Y., Chin, M., and Wei, X.-X. (2020). Emergence of functional and structural properties of the head direction system by optimization of recurrent neural networks. *arXiv:1912.10189*.

Cueva, C. J., and Wei, X.-X. (2018). Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. *arXiv [Preprint]. arXiv: 1803.07770*. Available online at: https://arxiv.org/pdf/1803.07770.pdf

Darwin, C. and Kebler, L. (1859). *On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life*. London: J. Murray.

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci*. 11, 333–341. doi: 10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Ding, K., Ma, K., Wang, S., and Simoncelli, E. P. (2020). Image quality assessment: unifying structure and texture similarity. *IEEE Trans. Pattern Anal. Mach. Intell*. 44, 267–2581. doi: 10.1109/TPAMI.2020.3045810

DiTullio, R. W., Piasini, E., Chaudhari, P., Balasubramanian, V., and Cohen, Y. E. (2022). Time as a supervisor: temporal regularity and auditory object learning. *bioRxiv [Preprint]*. doi: 10.1101/2022.11.10.515986

Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H., and Zoccolan, D. (2018). Accuracy of rats in discriminating visual objects is explained by the complexity of their perceptual strategy. *Curr. Biol*. 28, 1005–1015.e5. doi: 10.1016/j.cub.2018.02.037

Dunkel, M., Schmidt, U., Struck, S., Berger, L., Gruening, B., Hossbach, J., et al. (2009). Superscent?a database of flavors and scents. *Nucleic Acids Res*. 37(Suppl. 1), D291?-D294. doi: 10.1093/nar/gkn695

Einhäuser, W., Hipp, J., Eggert, J., Körner, E., and König, P. (2005). Learning viewpoint invariant object representations using a temporal coherence principle. *Biol. Cybern*. 93, 79–90. doi: 10.1007/s00422-005-0585-8

Fairhall, A. L., Lewen, G. D., Bialek, W., and de Ruyter van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature* 412, 787–792. doi: 10.1038/35090500

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394. doi: 10.1364/JOSAA.4.002379

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Comput*. 3, 194–200. doi: 10.1162/neco.1991.3.2.194

Franzius, M., Sprekeler, H., and Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput. Biol*. 3:e166. doi: 10.1371/journal.pcbi.0030166

Freeman, J., and Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nat. Neurosci*. 14, 1195–1201. doi: 10.1038/nn.2889

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

Fritsche, M., Lawrence, S. J. D., and de Lange, F. P. (2020). Temporal tuning of repetition suppression across the visual cortex. *J. Neurophysiol*. 123, 224–233. doi: 10.1152/jn.00582.2019

Garrigan, P., Ratliff, C. P., Klein, J. M., Sterling, P., Brainard, D. H., and Balasubramanian, V. (2010). Design of a trichromatic cone array. *PLoS Comput. Biol*. 6:e1000677. doi: 10.1371/journal.pcbi.1000677

Gatys, L., Ecker, A. S., and Bethge, M. (2015). "Texture synthesis using convolutional neural networks," in *Advances in Neural Information Processing Systems, Vol. 28*. doi: 10.1109/CVPR.2016.265

Geisler, W. S., and Perry, J. S. (2009). Contour statistics in natural images: grouping across occlusions. *Vis. Neurosci*. 26, 109–121. doi: 10.1017/S0952523808080875

Geisler, W. S., Perry, J. S., Super, B., and Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vis. Res.* 41, 711–724. doi: 10.1016/S0042-6989(00)00277-7

Gerhard, H. E., Wichmann, F. A., and Bethge, M. (2013). How sensitive is the human visual system to the local statistics of natural images? *PLoS Comput. Biol.* 9:e1002873. doi: 10.1371/journal.pcbi.1002873

Gervain, J., Nallet, C., Vallortigara, G., Zanon, M., Lemaire, B., Caramellino, R., et al. (2021). "The efficient coding of visual textures in rats, chicks and human infants," in *Presented at the 29th Kanizsa Symposium* (Padova). Available online at: https://dpg.unipd.it/en/percup/kanizsa-lecture

Gjorgjieva, J., Meister, M., and Sompolinsky, H. (2019). Functional diversity among sensory neurons from efficient coding principles. *PLoS Comput. Biol.* 15:e1007476. doi: 10.1371/journal.pcbi.1007476

Gjorgjieva, J., Sompolinsky, H., and Meister, M. (2014). Benefits of pathway splitting in sensory coding. *J. Neurosci.* 34, 12127–12144. doi: 10.1523/JNEUROSCI.1032-14.2014

Glickfeld, L. L., and Olsen, S. R. (2017). Higher-order areas of the mouse visual cortex. *Annu. Rev. Vis. Sci.* 3, 251–273. doi: 10.1146/annurev-vision-102016-061331

Gollisch, T., and Meister, M. (2010). Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* 65, 150–164. doi: 10.1016/j.neuron.2009.12.009

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23. doi: 10.1016/j.tics.2005.11.006

Gupta, D., Mlynarski, W., Sumser, A., Symonova, O., Svaton, J., and Joesch, M. (2022). Panoramic visual statistics shape retina-wide organization of receptive fields. *bioRxiv [Preprint]*. doi: 10.1101/2022.01.11.475815

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., and Tkačik, G. (2014). Variance predicts salience in central sensory processing. *eLife* 3:e03722. doi: 10.7554/eLife.03722

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866. doi: 10.1126/science.1117593

Issa, E. B., Cadieu, C. F., and DiCarlo, J. J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *eLife* 7:e42870. doi: 10.7554/eLife.42870

Julesz, B. (1962). Visual pattern discrimination. *IRE Trans. Inform. Theory* 8, 84–92. doi: 10.1109/TIT.1962.1057698

Kaliukhovich, D. A., De Baene, W., and Vogels, R. (2013). Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. *J. Cogn. Neurosci.* 25, 777–789. doi: 10.1162/jocn_a_00355

Kaliukhovich, D. A., and Op de Beeck, H. (2018). Hierarchical stimulus processing in rodent primary and lateral visual cortex as assessed through neuronal selectivity and repetition suppression. *J. Neurophysiol.* 120, 926–941. doi: 10.1152/jn.00673.2017

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., and DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* 22, 974–983. doi: 10.1038/s41593-019-0392-5

Keller, G. B., and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435. doi: 10.1016/j.neuron.2018.10.003

Khan, A. G., and Hofer, S. B. (2018). Contextual signals in visual cortex. *Curr. Opin. Neurobiol.* 52, 131–138. doi: 10.1016/j.conb.2018.05.003

Koch, K., McLean, J., Segev, R., Freed, M. A., Berry, M. J. II, Balasubramanian, V., et al. (2006). How much the eye tells the brain. *Curr. Biol.* 16, 1428–1434. doi: 10.1016/j.cub.2006.05.056

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *J. Neurophysiol.* 97, 3155–3164. doi: 10.1152/jn.00086.2007

Komban, S. J., Kremkow, J., Jin, J., Wang, Y., Lashgari, R., Li, X., et al. (2014). Neuronal and perceptual differences in the temporal processing of darks and lights. *Neuron* 82, 224–234. doi: 10.1016/j.neuron.2014.02.020

Körding, K. P., Kayser, C., Einhäuser, W., and König, P. (2004). How are complex cell properties adapted to the statistics of natural stimuli? *J. Neurophysiol.* 91, 206–212. doi: 10.1152/jn.00149.2003

Kreiman, G., and Serre, T. (2020). Beyond the feedforward sweep: feedback computations in the visual cortex. *Ann. N. Y. Acad. Sci.* 1464, 222–241. doi: 10.1111/nyas.14320

Kremkow, J., Jin, J., Komban, S. J., Wang, Y., Lashgari, R., Li, X., et al. (2014). Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3170–3175. doi: 10.1073/pnas.1310442111

Krishnamurthy, K., Hermundstad, A. M., Mora, T., Walczak, A. M., and Balasubramanian, V. (2022). Disorder and the neural representation of complex odors. *Front. Comput. Neurosci.* 16:917786. doi: 10.3389/fncom.2022.917786

Kuang, X., Poletti, M., Victor, J. D., and Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Curr. Biol.* 22, 510–514. doi: 10.1016/j.cub.2012.01.050

Lamme, V. A., Supèr, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* 8, 529–?535. doi: 10.1016/S0959-4388(98)80042-1

Landy, M. S., and Graham, N. (2004). "Visual perception of texture," in *The Visual Neurosciences*, eds L. M. Chalupa and J. S. Werner (Cambridge, MA: MIT Press), 1106–1118. doi: 10.7551/mitpress/7131.003.0084

Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitsch. Naturforsch.* 36, 910–912. doi: 10.1515/znc-1981-9-1040

Levy, W. B., and Calvert, V. G. (2021). Communication consumes 35 times more energy than computation in the human cortex, but both costs are needed to predict synapse number. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2008173118. doi: 10.1073/pnas.2008173118

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nat. Neurosci.* 5, 356–363. doi: 10.1038/nn831

Li, N., and DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321, 1502–1507. doi: 10.1126/science.1160028

Liu, Y. S., Stevens, C. F., and Sharpee, T. O. (2009). Predictable irregularities in retinal receptive fields. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16499–16504. doi: 10.1073/pnas.0908926106

Lüdtke, N., Das, D., Theis, L., and Bethge, M. (2015). A generative model of natural texture surrogates. *arXiv:1505.07672*. doi: 10.48550/arXiv.1505.07672

Marques, T., Nguyen, J., Fioreze, G., and Petreanu, L. (2018). The functional organization of cortical feedback inputs to primary visual cortex. *Nat. Neurosci.* 21, 757–764. doi: 10.1038/s41593-018-0135-z

Masquelier, T., Serre, T., Thorpe, S., and Poggio, T. (2007). *Learning Complex Cell Invariance From Natural Videos: A Plausibility Proof*. Technical report, CSAIL. doi: 10.21236/ADA477541

Matteucci, G., and Zoccolan, D. (2020). Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. *Sci. Adv.* 6:eaba3742. doi: 10.1126/sciadv.aba3742

Mayhew, E. J., Arayata, C. J., Gerkin, R. C., Lee, B. K., Magill, J. M., Snyder, L. L., et al. (2022). Transport features predict if a molecule is odorous. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2116576119. doi: 10.1073/pnas.2116576119

McIntosh, L., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S. (2016). "Deep learning models of the retinal response to natural scenes," in *Advances in Neural Information Processing Systems, Vol. 29*.

Muratore, P., Tafazoli, S., Piasini, E., Laio, A., and Zoccolan, D. (2022). "Prune and distill: similar reformatting of image information along rat visual cortex and deep neural network," in *Advances in Neural Information Processing Systems, Vol. 35*.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

Nayebi, A., Sagastuy-Brena, J., Bear, D. M., Kar, K., Kubilius, J., Ganguli, S., et al. (2022). Recurrent connections in the primate ventral visual stream mediate a trade-off between task performance and network size during core object recognition. *Neural Comput.* 34, 1652–1675. doi: 10.1162/neco_a_01506

Niell, C. M., and Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* 28, 7520–7536. doi: 10.1523/JNEUROSCI.0623-08.2008

Okazawa, G., Tajima, S., and Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque v4. *Proc. Natl. Acad. Sci. U.S.A.* 112, E351–?E360. doi: 10.1073/pnas.1415146112

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

Palmer, S. E., Marre, O., Berry, M. J., and Bialek, W. (2015). Predictive information in a sensory population. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6908–6913. doi: 10.1073/pnas.1506855112

Park, T., Zhu, J. -Y., Wang, O., Lu, J., Shechtman, E., Efros, A. A., et al. (2020). Swapping autoencoder for deep image manipulation. *arXiv [Preprint]. arXiv: 2007.00653*. Available online at: https://arxiv.org/pdf/2007.00653.pdf

Pasupathy, A., and Connor, C. E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.* 82, 2490–2502. doi: 10.1152/jn.1999.82.5.2490

Pasupathy, A., and Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86, 2505–2519. doi: 10.1152/jn.2001.86.5.2505

Pasupathy, A., and Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/972

Perge, J. A., Koch, K., Miller, R., Sterling, P., and Balasubramanian, V. (2009). How the optic nerve allocates space, energy capacity, and information. *J. Neurosci.* 29, 7917–7928. doi: 10.1523/JNEUROSCI.5200-08.2009

Perge, J. A., Niven, J. E., Mugnaini, E., Balasubramanian, V., and Sterling, P. (2012). Why do axons differ in caliber? *J. Neurosci.* 32, 626–638. doi: 10.1523/JNEUROSCI.4254-11.2012

Piasini, E. (2021). *Metex - A Software Package for Maximum Entopy TEXtures*. doi: 10.5281/zenodo.5561807

Piasini, E., Soltuzu, L., Muratore, P., Caramellino, R., Vinken, K., Op de Beeck, H., et al. (2021). Temporal stability of stimulus representation increases along rodent visual cortical hierarchies. *Nat. Commun.* 12:4448. doi: 10.1038/s41467-021-24456-3

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

Portilla, J., and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* 40, 49–71. doi: 10.1023/A:1026553619983

Purpura, K. P., Victor, J. D., and Katz, E. (1994). Striate cortex extracts higher-order spatial correlations from visual textures. *Proc. Natl. Acad. Sci. U.S.A.* 91, 8482–8486. doi: 10.1073/pnas.91.18.8482

Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., and King, A. J. (2011). Contrast gain control in auditory cortex. *Neuron* 70, 1178–1191. doi: 10.1016/j.neuron.2011.04.030

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

Ratliff, C. P., Borghuis, B. G., Kao, Y.-H., Sterling, P., and Balasubramanian, V. (2010). Retina is structured to process an excess of darkness in natural scenes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17368–17373. doi: 10.1073/pnas.1005846107

Reiter, S., and Laurent, G. (2020). Visual perception and cuttlefish camouflage. *Curr. Opin. Neurobiol.* 60, 47–54. doi: 10.1016/j.conb.2019.10.010

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature* 548, 92–96. doi: 10.1038/nature23020

Rust, N. C., and DiCarlo, J. J. (2010). Selectivity and tolerance ("invariance??") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* 30, 12978–12995. doi: 10.1523/JNEUROSCI.0179-10.2010

Salisbury, J. M., and Palmer, S. E. (2016). Optimal prediction in the retina and natural motion statistics. *J. Stat. Phys.* 162, 1309–1323. doi: 10.1007/s10955-015-1439-y

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., et al. (2020). Brain-score: which artificial neural network for object recognition is most brain-like? *bioRxiv [Preprint]*. doi: 10.1101/407007

Schwartz, G., Harris, R., Shrom, D., and Berry, M. J. (2007). Detection and prediction of periodic patterns by the retina. *Nat. Neurosci.* 10, 552–554. doi: 10.1038/nn1887

Schwartz, O., and Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nat. Neurosci.* 4, 819–825. doi: 10.1038/90526

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* 592, 86–92. doi: 10.1038/s41586-020-03171-x

Simmons, K. D., Prentice, J. S., Tkačik, G., Homann, J., Yee, H. K., Palmer, S. E., et al. (2013). Transformation of stimulus correlations by the retina. *PLoS Comput. Biol.* 9:e1003344. doi: 10.1371/journal.pcbi.1003344

Singh, V., Tchernookov, M., and Balasubramanian, V. (2021). What the odor is not: estimation by elimination. *Phys. Rev. E* 104:024415. doi: 10.1103/PhysRevE.104.024415

Smith, E. C., and Lewicki, M. S. (2006). Efficient auditory coding. *Nature* 439, 978–982. doi: 10.1038/nature04485

Sorscher, B., Mel, G., Ganguli, S., and Ocko, S. (2019). "A unified theory for the origin of grid cells through the lens of pattern formation," in *Advances in Neural Information Processing Systems, Vol. 32*.

Stephens, G. J., Mora, T., Tkačik, G., and Bialek, W. (2013). Statistical thermodynamics of natural images. *Phys. Rev. Lett.* 110:018701. doi: 10.1103/PhysRevLett.110.018701

Sterling, P., and Laughlin, S. (2015). *Principles of Neural Design*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262028707.001.0001

Stigliani, A., Jeska, B., and Grill-Spector, K. (2019). Differential sustained and transient temporal processing across visual streams. *PLoS Comput. Biol.* 15:e1007011. doi: 10.1371/journal.pcbi.1007011

Striedter, G. F., and Northcutt, R. G. (2020). *Brains Through Time: A Natural History of Vertebrates*. New York, NY: Oxford University Press. doi: 10.1093/oso/9780195125689.001.0001

Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F. B., Vanzella, W., Riggi, M., et al. (2017). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* 6:e22794. doi: 10.7554/eLife.22794

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139. doi: 10.1146/annurev.ne.19.030196.000545

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., et al. (2018). Recurrent computations for visual pattern completion. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8835–8840. doi: 10.1073/pnas.1719397115

Teşileanu, T., Cocco, S., Monasson, R., and Balasubramanian, V. (2019). Adaptation of olfactory receptor abundances for efficient coding. *eLife* 8:e39279. doi: 10.7554/eLife.39279

Teşileanu, T., Conte, M. M., Briguglio, J. J., Hermundstad, A. M., Victor, J. D., and Balasubramanian, V. (2020). Efficient coding of natural scene statistics predicts discrimination thresholds for grayscale textures. *eLife* 9:e54347. doi: 10.7554/eLife.54347

Teşileanu, T., Ölveczky, B., and Balasubramanian, V. (2017). Rules and mechanisms for efficient two-stage learning in neural circuits. *eLife* 6:e20944. doi: 10.7554/eLife.20944

Tkačik, G., Garrigan, P., Ratliff, C. P., Milčinski, G., Klein, J. M., Seyfarth, L. H., et al. (2011). Natural images from the birthplace of the human eye. *PLoS ONE*, 6, e20409. doi: 10.1371/journal.pone.0020409

Tkačik, G., Prentice, J. S., Victor, J. D., and Balasubramanian, V. (2010). Local statistics in natural scenes predict the saliency of synthetic textures. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18149–18154. doi: 10.1073/pnas.0914916107

Touhara, K., and Vosshall, L. B. (2009). Sensing odorants and pheromones with chemosensory receptors. *Annu. Rev. Physiol.* 71, 307–332. doi: 10.1146/annurev.physiol.010908.163209

Trenholm, S., Schwab, D. J., Balasubramanian, V., and Awatramani, G. B. (2013). Lag normalization in an electrically coupled neural network. *Nat. Neurosci.* 16, 154–156. doi: 10.1038/nn.3308

Ustyuzhaninov, I., Brendel, W., Gatys, L. A., and Bethge, M. (2016). Texture synthesis using shallow convolutional networks with random filters. *arXiv [Preprint]. arXiv: 1606.00021*. Available online at: https://arxiv.org/pdf/1606.00021.pdf

van Bergen, R. S., and Kriegeskorte, N. (2020). Going in circles is the way forward: the role of recurrence in visual inference. *Curr. Opin. Neurobiol.* 65, 176–193. doi: 10.1016/j.conb.2020.11.009

van Hateren, J. H. (1992a). A theory of maximizing sensory information. *Biol. Cybern.* 68, 23–29. doi: 10.1007/BF00203134

van Hateren, J. H. (1992b). Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J. Compar. Physiol. A* 171, 157–170. doi: 10.1007/BF00188924

van Hateren, J. H., and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. B* 265, 359–366. doi: 10.1098/rspb.1998.0303

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

Victor, J. D. (1994). Images, statistics, and textures: implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: comment. *J. Opt. Soc. Am. A* 11:1680. doi: 10.1364/JOSAA.11.001680

Victor, J. D., and Conte, M. M. (1991). Spatial organization of nonlinear interactions in form perception. *Vis. Res.* 31, 1457–1488. doi: 10.1016/0042-6989(91)90125-O

Victor, J. D., and Conte, M. M. (2012). Local image statistics: maximum-entropy constructions and perceptual salience. *J. Opt. Soc. Am. A* 29:1313. doi: 10.1364/JOSAA.29.001313

Victor, J. D., Conte, M. M., and Chubb, C. F. (2017). Textures as probes of visual processing. *Annu. Rev. Vis. Sci.* 3, 275–296. doi: 10.1146/annurev-vision-102016-061316

Vincent, B. T., and Baddeley, R. J. (2003). Synaptic energy efficiency in retinal processing. *Vis. Res.* 43, 1285–1292. doi: 10.1016/S0042-6989(03)00096-8

Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287, 1273–1276. doi: 10.1126/science.287.5456.1273

Vinken, K., Boix, X., and Kreiman, G. (2020). Incorporating intrinsic suppression in deep neural networks captures dynamics of adaptation in neurophysiology and perception. *Sci. Adv.* 6:eabd4205. doi: 10.1126/sciadv.abd4205

Vinken, K., Van den Bergh, G., Vermaercke, B., and Op de Beeck, H. P. (2016). Neural representations of natural and scrambled movies progressively change from rat striate to temporal cortex. *Cereb. Cortex* 26, 3310–3322. doi: 10.1093/cercor/bhw111

Vinken, K., Vogels, R., and Op de Beeck, H. (2017). recent visual experience shapes visual processing in rats through stimulus-specific adaptation and response enhancement. *Curr. Biol.* 27, 914–919. doi: 10.1016/j.cub.2017.02.024

Vosshall, L. B., Wong, A. M., and Axel, R. (2000). An olfactory sensory map in the fly brain. *Cell* 102, 147–159. doi: 10.1016/S0092-8674(00)00021-0

Wallis, G., and Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194. doi: 10.1016/S0301-0082(96)00054-8

Wang, X.-J. (2022). Theory of the multiregional neocortex: large-scale neural dynamics and distributed cognition. *Annu. Rev. Neurosci.* 45, 533–560. doi: 10.1146/annurev-neuro-110920-035434

Webster, M. A. (2015). Visual adaptation. *Annu. Rev. Vis. Sci.* 1, 547–567. doi: 10.1146/annurev-vision-082114-035509

Weckström, M., and Laughlin, S. B. (1995). Visual ecology and voltage-gated ion channels in insect photoreceptors. *Trends Neurosci.* 18, 17–21. doi: 10.1016/0166-2236(95)93945-T

Wei, X.-X., Prentice, J., and Balasubramanian, V. (2015). A principle of economy predicts the functional architecture of grid cells. *eLife* 4:e08362. doi: 10.7554/eLife.08362

Wiskott, L., and Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* 14, 715–770. doi: 10.1162/089976602317318938

Wyss, R., König, P., and Verschure, P. F. M. J. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS Biol.* 4:e120. doi: 10.1371/journal.pbio.0040120

Yamins, D. L., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8619–8624. doi: 10.1073/pnas.1403112111

Yu, Y., Schmid, A. M., and Victor, J. D. (2015). Visual processing of informative multipoint correlations arises primarily in V2. *eLife* 4:e06604. doi: 10.7554/eLife.06604

Zhang, X., and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* 5, 124–133. doi: 10.1038/nn800

Zhu, S. C., Liu, X. W., and Wu, Y. N. (2000). Exploring texture ensembles by efficient Markov chain Monte Carlo-toward a "trichromacy" theory of texture. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 554–569. doi: 10.1109/34.862195

Zoccolan, D., Oertelt, N., DiCarlo, J. J., and Cox, D. D. (2009). A rodent model for the study of invariant visual object recognition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8748–8753. doi: 10.1073/pnas.0811583106

Zozulya, S., Echeverri, F., and Nguyen, T. (2001). The human olfactory receptor repertoire. *Genome Biol.* 2, 1–12. doi: 10.1186/gb-2001-2-6-research0018