


# Optimal Prediction in the Retina and Natural Motion Statistics

Jared M. Salisbury<sup>1,2</sup> · Stephanie E. Palmer<sup>2</sup> 

Received: 1 July 2015 / Accepted: 15 December 2015 / Published online: 11 January 2016  
© Springer Science+Business Media New York 2016

**Abstract** Almost all behaviors involve making predictions. Whether an organism is trying to catch prey, avoid predators, or simply move through a complex environment, the organism uses the data it collects through its senses to guide its actions by extracting from these data information about the future state of the world. A key aspect of the prediction problem is that not all features of the past sensory input have predictive power, and representing all features of the external sensory world is prohibitively costly both due to space and metabolic constraints. This leads to the hypothesis that neural systems are optimized for prediction. Here we describe theoretical and computational efforts to define and quantify the efficient representation of the predictive information by the brain. Another important feature of the prediction problem is that the physics of the world is diverse enough to contain a wide range of possible statistical ensembles, yet not all inputs are probable. Thus, the brain might not be a generalized predictive machine; it might have evolved to specifically solve the prediction problems most common in the natural environment. This paper summarizes recent results on predictive coding and optimal predictive information in the retina and suggests approaches for quantifying prediction in response to natural motion. Basic statistics of natural movies reveal that general patterns of spatiotemporal correlation are present across a wide range of scenes, though individual differences in motion type may be important for optimal processing of motion in a given ecological niche.

**Keywords** Prediction · Neural computation · Natural statistics · Retina · Motion processing

---

✉ Stephanie E. Palmer  
sepalmer@uchicago.edu

<sup>1</sup> Graduate Program in Computational Neuroscience, Chicago, IL, USA

<sup>2</sup> Department of Organismal Biology and Anatomy, University of Chicago, 1027 E 57th St., Chicago, IL 60637, USA

## 1 Introduction

What are the predictable components of the input to an animal's visual system in its natural environment? While the characteristics of static images have been explored in large image repositories [3, 9, 27, 40, 46, 55, 57], less is known or measured in the temporal domain [15, 27]. One interesting feature of scaling in static images is the power law distribution of spatial variation in local contrast [27, 46]. This scaling implies that natural images are scale free, displaying the same basic structure on all length scales. Power law behavior in the frequency distribution of temporal fluctuations in total scene luminance have also been observed in a variety of natural contexts, and scenes display slightly different exponents depending on their specific content [15]. This paper will review recent attempts to connect natural motion statistics to efficient prediction in the visual system, focusing on the retina. Tying temporal statistics of natural scenes to neural prediction will reveal what types of motion the brain can efficiently represent and therefore constrain the types of predictions the brain can perform.

The concept of efficient coding for prediction in the brain has been developed in two main ways: via theories of predictive coding [8, 32, 35, 44, 53, 54] in which temporal redundancy is minimized, and through analytical work to characterize the optimal trade-offs between representing the past and future sensory input [13, 14] (via information bottleneck calculations [19, 21, 56]). In this paper, we will review and relate these two approaches to neural coding in the retina, and propose methods for extending this work to the context of natural motion statistics.

It has been shown that retinal ganglion cells (RGCs), the output cells of the retina whose axons form the optic nerve, display a whole host of nonlinear processing characteristics that may be connected to prediction. RGCs respond differentially to object versus background motion [41]. Ganglion cells have also been shown to code for a variety of motion features in ways that cannot be accounted for by a simple receptive field model of encoding. This includes: motion anticipation, the coding in the retina for the anticipated position of an object moving at constant velocity [11]; the omitted-stimulus response, in which ganglion cells fire after the cessation of a sequence of visual flashes at the appropriate delay where the next flash in the sequence would be expected [48]; and reversal responses, where neurons in the retina fire a synchronous burst of activity after the reversal of a moving bar, irrespective of their relative receptive field positions [49]. All of these adaptive motion-processing features speak to the retina's complexity as an encoding device, and relate to the predictability of the future state of visual stimuli. Recent work has also shown that the retina solves a general prediction problem in a near-optimal way [42].

We will review this background material and discuss how extensions of this work could reveal how an organism's ecological niche shapes its predictive processing. In particular, it may be that the retinas of different species possess the capacity to solve different suites of motion prediction problems tailored to their natural environment. Evolution may shape which problems are hardwired in the early visual system, and exploring that can uncover just how far the retina is able to tune its predictive power to the statistics of its inputs.

## 2 Theories of Optimal Prediction

Sitting at the front end of the visual system and with a limited number of fibers to transmit all the visual information the brain receives, the retina has long been hypothesized to be an efficient and perhaps even optimal encoder of the visual world [3, 6, 7, 26, 32]. This notion of efficiency dictates that the retina's representation of the visual inputs to the photoreceptor

layer should be as lossless as possible, given the number of cables the retina has along which to transmit information to the brain, intrinsic noise in neural responses, and the fact that metabolic constraints limit the firing rate of neurons. To make the best use of each fiber, these signals should be independent in space and in time. Recent work from a variety of researchers expands on this simple notion of efficiency. Natural inputs to the retina are non-Gaussian [27,46], the noise spectrum in neural data is not white, retinal firing is certainly highly redundant [43], and not all information about the input is equally relevant to the organism. The concept of optimizing the predictive capacity of the retina assigns value to particular bits of information: it says that compression is only successful when the transmitted bits convey information about the future input [14]. The information bottleneck method [56] is a way of defining relevant information, in this case information about the future, as the distortion measure.

## 2.1 Efficient Coding in the Time Domain: Predictive Coding

The efficient coding hypothesis states that information about the input should be maximized, while minimizing the entropy of the response [51]. If not all bits of information about the input are retained, the problem can be formulated using rate distortion theory [10],

$$\min_{p(r_t|s_t)} I(R_t; S_t) + D(R_t, S_t), \quad (1)$$

where  $D$  is the average distortion,  $R$  is the neural response to the stimulus  $S$ , and the minimization finds the lowest transmitted bit rate, given  $D$ .

The core concept in predictive coding, in the time domain, is that temporal correlations in the output stream should be eliminated, so that only deviations from expected input, or those that are ‘surprising,’ are encoded [53]. If the input statistics are stationary, predictive coding aims to minimize the response of the system. A rate distortion theory for predictive coding might be formulated in this case using an information theoretic framework,

$$\min_{p(r_t|s_t)} I(R_t; R_{t+\Delta t}^*) - \beta I(R_t; S_{t_{\text{past}}}), \quad (2)$$

where the information about the future response is minimized, subject to a constraint, weighted by  $\beta$ , on the information retained about the past. Here, the information about the past, the converse of distortion, is fixed and we minimize predictive information subject to that constraint. Predictive coding is not specifically formulated for this type of input scenario, but for one in which the input statistics change in time. The role of neurons in a predictive coding paradigm is to code for changes in stimulus statistics, not the ongoing predictable events in a stationary world.

Predictive coding has been postulated to be achieved through feedback connections from higher areas onto sensory input streams [8,35,44,58], and early [53] as well as recent work in the retina hypothesizes feedforward adaptive mechanisms at the sensory periphery may result in predictive coding [34]. As such, predictive coding is highly efficient, because redundancy in time is eliminated. Mechanisms have been proposed by which the retina could implement predictive coding, via inhibitory interactions at bipolar terminals [32]. Also, the work of Denève shows how predictive coding may be a general self-organized property of neural networks [16].

## 2.2 Predictive Information

Information theoretic treatments of prediction in the brain focus on defining not just the code that retains the most stimulus information for a given output bit rate, but the one that retains the most information about the future stimulus. This addition of the notion of relevant information has sharpened discussions of early sensory processing in the context of prediction [14]. The theory for retaining the optimal amount of predictive information has been well-developed by Tishby et al. [19, 21, 56], and leads not only to elegant theoretical results but also testable experimental predictions. Recent experimental and theoretical work draws on these results and has shown that the salamander retina may be optimized for prediction of a simple motion stimulus [42].

The efficient representation of predictive information that is seen in [42] adds the notion of relevant information to the classical idea of efficient coding. The simplest version of the efficient coding hypothesis is that the retina processes visual inputs to remove redundancy, allowing the array of retinal ganglion cells to make fuller use of their limited capacity to transmit information [3, 4, 7]. The results in [42] suggest that the retina is not designed to represent all of the input light patterns impinging on its photoreceptors, but instead to represent those parts of the input that are most predictive of the future. The retina clearly throws away some aspects of the input light patterns, but perhaps only those parts that are irrelevant for the task of prediction.

*Information bottleneck approaches* The maximal amount of predictive information a system can possibly encode can be found by solving the following information bottleneck problem:

$$\min_{p(r_t|s_t)} I(R_t; S_{t,t-\Delta t,t-2\Delta t,\dots}) - \beta I(R_t; S_{t+\Delta t}), \quad (3)$$

where we have now written  $S_{(t)}_{\text{past}}$  explicitly in terms of a series of time points leading up to the response. This can also be understood as a rate distortion problem where the distortion metric is the predictive information. Here we can see how predictive information and predictive coding are really solving complementary optimization problems for stationary stimuli.

Providing an efficient representation of predictive information is nearly opposite of what one would expect from neurons doing predictive coding. In that type of code, signals are decorrelated in time so that predictable components are eliminated and neurons encode only the deviations from expectation, or surprise. After adaptation (i.e. higher level selection of the current best input model) to a particular predictable input stimulus, surprise should be eliminated. Without surprise driving response, neurons only spike at their basal noise level. The responses of neurons implementing a truly optimal predictive code in a stationary and predictable input environment would, thus, carry no predictive information about their own responses, beyond the cells' own intrinsic correlation times. In contrast, recent results [42] suggest that the retina has a large amount of response-driven predictive information, and that these responses efficiently separate predictive from non-predictive bits, and transmit the predictive bits preferentially. During the time when a new stimulus ensemble is being learned or adapted to, these two coding schemes might share many common response features. Designing experiments to directly test transient responses at the switch between two input statistics with different predictable structure would help to differentiate these two theories.

Predictive coding does not explicitly preserve information about the future stimulus in the peripheral input channel. It postulates top-down comparison of inputs to the current predicted state to negate or cancel out predictable features and reduce response when inputs are predictable. As such, it is hard to compare predictive coding to optimal predictive information schemes. The prescription for predictive coding is if  $S$  is the input, and  $R^*$  is the

expected response and  $R$  is the observed response: when  $p(r|s)$  is much different from  $p(r^*|s)$ , respond. Somewhere in the brain must live the model(s) that generates  $r^*$ . Higher cortical areas that inhibit early sensory areas, such as the lateral geniculate nucleus, might provide precisely this type of feedback, and have been implicated in experimental evidence for predictive coding, see e.g. [44]. In the retina, with little to no feedback from any downstream area, these models must be wired into the retinal circuit. An explicit representation of the components of the current input that best predict the future might be readily used to drive fast behaviors, such as an escape response.

*Coding for surprising stimuli* Predictive coding and optimal predictive information are seemingly opposed processing theories, but yield similar predictions about neural response in certain stimulus conditions, while they are diametrically opposed in others. We illustrate this by way of a few (somewhat pathological) toy examples: If we imagine a world that is wholly static, there is nothing to predict, no predictive information, and a predictive coder would have no response. In this null framework, when an organism is perhaps staring at a static image with perfectly stable eyes, the two theories agree: no response should be generated in either theory. If the world is instead completely stochastic, and in a form not properly anticipated by the coder, however, predictive coding would dictate that all noise signals that deviate strongly from the prior on the input generate a response, since each one is unexpected and therefore surprising. These surprising inputs are, however, uninformative about the future, because they are a pure noise signal. The predictive information present is zero and no response modulation should be encoded in the ‘maximize predictive information’ framework.

Predictive coding is, however, designed explicitly for non-stationary stimuli. While inputs are changing, the two theories are more closely aligned. Predictive information optimization would code for a surprising change in the inputs, because that surprising feature will have maximal information about the future state of the input. In that sense the two methods are aligned and preserving predictive information preferentially over other bits of past information is ‘efficient’, in the predictive coding sense.

The two theories differ in the basic layout of where models of the world are stored and how they are used to sculpt efficient predictive computations in the brain. Predictive information posits representation of the current content of the inputs that can best predict the future, and this information is explicitly represented in the neural response. This may be useful in early sensory processing stages, to quickly drive behavioral responses that rely on fast processing. When the world is in a stationary predictable state, predictive coding postulates that that current state should be offloaded from the sensory periphery to some higher-level brain area that stores all the possible configurations of the predictable input space. In this way, much brain real-estate and learning over time (perhaps evolutionary time) is invested in reducing the number of spikes at early sensory stages. The observation that cortical spiking is sparse lends support to such a scheme. In reality, the brain is most likely operating using a combination of explicit representation of the predictable events in the world (maximizing predictive information in the response) and suppressing continued response to ongoing predictable events (predictive coding).

*Where models of the input statistics are stored* Predictive coding, in its clear formulation by Rao and Ballard, postulates that a set of models of the input statistics are present in higher order areas that feedback onto sensory areas to suppress responses to predictable events [44]. Recent work has elegantly shown how this can be implemented in a hierarchical (Bayesian) framework [24]. With limited feedback impinging on the retina, however, the retina itself

must store the models of the input statistics it will receive. It has been demonstrated how adaptive gain mechanisms in the retina, that have presumably been encoded over evolutionary time, can instantiate predictive coding in the retina [32]. To further test these ideas in the context of natural scene statistics, one needs to define the set of motion types an organism encounters in its natural environment.

### 3 Basic Models of Early Visual Processing

Many models of early visual processing predict spiking in retinal ganglion cells, the output cells of the retina, based on filtering and subsequent thresholding of the visual input. In linear-nonlinear-Poisson (LN or LNP) models, the probability of spiking is an instantaneous, nonlinear function of a linearly filtered version of the sensory input. In the case of the retina that we study here, the input is the pattern of light impinging on the photoreceptors, a function of space and time,  $s(\mathbf{x}, t)$ . Thus, if we write the probability per unit time of a spike (the firing rate), we have

$$r_{\text{LN}}(t) = r_0 g(z), \quad (4)$$

where  $r_0$  sets the scale of firing rates and  $g(z)$  is a dimensionless nonlinear function whose argument,  $z$ , is defined by

$$z(t) = \int_0^t d\tau \int d\mathbf{x} f(\mathbf{x}, \tau) s(\mathbf{x}, t - \tau); \quad (5)$$

where the function  $f(\mathbf{x}, \tau)$  is the receptive field.

If we deliver stimuli that are drawn from a Gaussian white noise ensemble, then

$$f(\mathbf{x}, \tau) \propto \langle s(\mathbf{x}, t - \tau) \delta(t - t_{\text{spike}}) \rangle, \quad (6)$$

where  $t_{\text{spike}}$  is the time of a spike and  $\langle \cdots \rangle$  denotes an average over the stimulus ensemble [20]. This filter can be applied to predict responses to any novel input stimulus, once the nonlinearity,  $g(z)$ , is fit to training data. The resulting model firing rate  $r_{\text{LN}}(t)$  can be used to generate spikes drawn from a Poisson process with that rate. Spikes are correlated in time through input stimulus correlations and the length of the temporal component of the model's receptive field.

In many contexts these, our simplest models of retinal ganglion cell firing, fail to recapitulate the actual response properties of the retina. This is most clearly demonstrated for motion stimuli. LN models fail for a variety of moving stimuli and cannot account for responses that include: motion anticipation [11], the reversal response [49], object motion detection [41], and the omitted-stimulus response [48]. The myriad ways in which a simple LN model fails to reproduce known retinal response properties are described in detail in recent reviews from Gollisch and Meister [29] and Berry and Schwartz [12]. When simple bars of light move across the retina, reverse their path, blink on and off, or move in more complex ways, many kinds of nonlinear processes in the retina are activated. None of these effects are captured by this basic version of the LN model for retinal firing. Complex adaptive gain control mechanisms or cascades of nonlinear processing steps need to be added to these models to explain the observed spiking response of the retina [11, 48, 49]. Perhaps the best way to dissect which variant on the LN theme is relevant for the retina is to find a universal model that captures all of these behaviors.

Additionally, results in [42] reveal that the LN model fails to recapitulate the near-optimal behavior of the real neurons, as illustrated here in Fig. 1. In this work, a bar stimulus was

presented to the retina whose trajectory contained both predictable and stochastic motion components. The center of mass of the moving bar evolved in time,  $t$ , according to an Ornstein–Uhlenbeck process with an added damping term,

$$x_{t+\Delta\tau} = x_t + v_t \Delta\tau \quad (7)$$

$$v_{t+\Delta\tau} = [1 - \Gamma \Delta\tau]v_t - \omega^2 x_t \Delta\tau + \xi_t \sqrt{D \Delta\tau}, \quad (8)$$

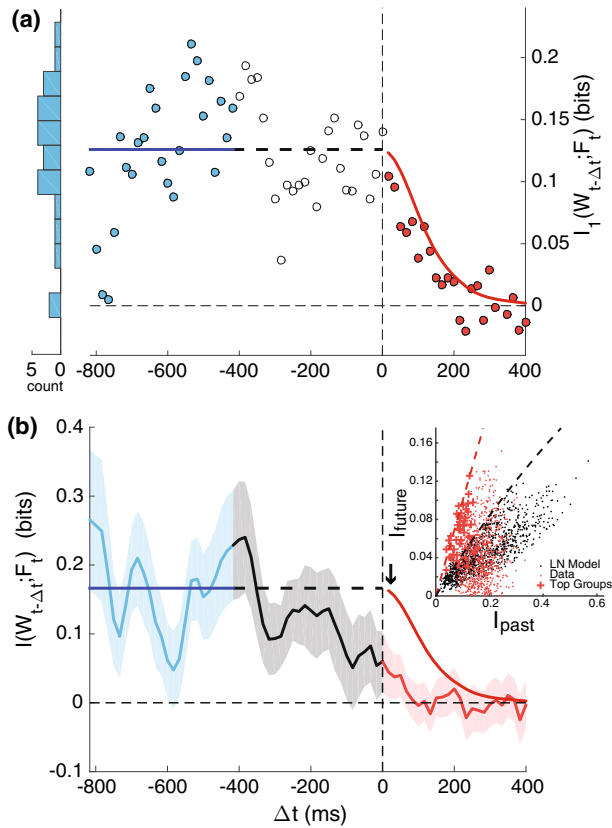
where  $\xi_t$  is a Gaussian random variable with zero mean and unit variance. With a natural frequency  $\omega = 2\pi \times (1.5 \text{ rad/s})$ , and the damping  $\Gamma = 20 \text{ s}^{-1}$ ,  $\zeta = \Gamma/2\omega = 1.06$ , the dynamics are slightly overdamped. The time step  $\Delta\tau = 1/60 \text{ s}$  is the update time of the noise process, and  $D = 2.7 \times 10^6 \text{ pixel}^2/\text{s}^3$ . These parameters were chosen to drive robust response in the larval salamander retina and to contain a somewhat complex mixture of predictable and non-predictable motion. Many such trajectories were sampled, and displayed to the retina so that 100 independent paths converged onto a single common future trajectory, for 30 such ‘futures’. In this way, the future could be repeated while varying the past, allowing for the computation of information about the future directly [14, 42]. Information about the identity of the future contained in the retinal response was computed to test for saturation of the maximal predictive information given by the information bottleneck calculation for the same stimulus. Results of this calculation are reproduced in Fig. 1. While the average group of 5 cells does not reach the average bound on predictive information, Fig. 1b, particular groups do, Fig. 1a. We do not expect every group in the retina to be maximally predictive, but notably, every cell in the recorded population participates in a 5-cell group that does saturate the bound on the maximal predictive information possible, given the information retained about the past, Fig. 1b (inset, red crosses).

If we take the LN model derived from responses to a random checkerboard stimulus and use it to produce neural responses to the temporally correlated and stochastically noisy moving bar stimulus described in [42], the predictive information carried by the neurons differs dramatically from that obtained from the real neurons (Fig. 1, inset). All model groups fall away from the bound determined by  $\Delta t = 17 \text{ ms}$ , the delay between the current response and the onset of the common future. When we compute information about the future, we assume that the future starts now, and do not make any allowances for processing delays. We could, instead, compare the performance of the LN model with bounds calculated assuming that there is a delay between past and future, so that  $\Delta t^* = \Delta t + t_{\text{delay}}$ . The bound for  $\Delta t^*$  is chosen to be  $t_{\text{delay}} = 117 \text{ ms}$ , comparable to the delay one might estimate from the peak of the information about position, or from the structure of the receptive fields themselves. Interestingly, the model neurons do come close to this delayed bound.

Of course, these model cells might not be optimal at any delay; instead they could fail to represent all of the predictable components of the stimulus, such as the velocity. Real salamander RGCs have a delay of at least 50 ms, as measured by the time to the peak firing rate induced by a flash. The data reveal that the retina has a mechanism that allows it to saturate the bound on the predictive information with almost zero effective processing delay when responding to a predictably moving stimulus.

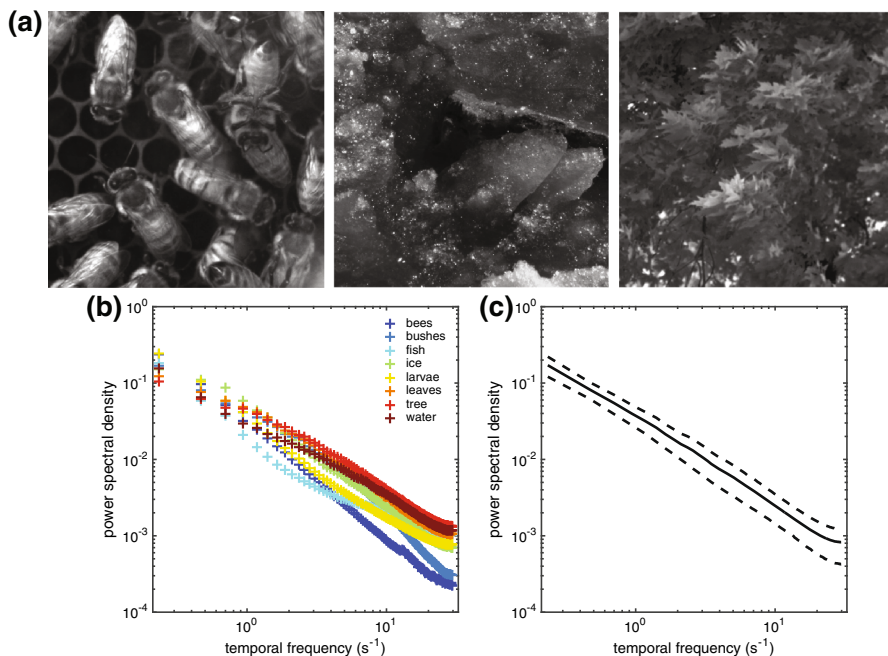
This work only scratches the surface of optimal coding for the future stimulus in the retina. The motion statistics were chosen to include both two time scales of predictable motion as well as a purely stochastic motion component, while still being soluble via Gaussian information bottleneck techniques [19]. Important extensions of this work will test other parametric motion models with different statistics, but one of the most important directions of future research in this area is to explore motion models that mimic the properties of natural scenes.





**Fig. 1** Predictive information in the salamander retina reaches the bound on the predictive information for small groups of cells. **a** Moving bar stimuli are presented to the retina and contain temporal correlations as well as stochastic noise. These stimuli converge from a random set of 3000 initial states to 30 common future trajectories, which begin at  $\Delta t = 0$  along the  $x$ -axis. Responses of the retina from a particular 5-cell group are recorded at time  $\Delta t$ , relative to the onset of the common future, such that positive  $\Delta t$ 's (red circles) correspond to responses leading up to the common future, and negative  $\Delta t$ 's correspond to responses recorded after the onset of the 30 common trajectories (blue and white circles). Information about the identity of the future can be computed as a function of  $\Delta t$  by comparing response entropy over all futures to response entropy for a given future in a particular time bin as in [42]. Blue circles and their information histogram plotted on the left correspond to responses during which the responses have become stationary with respect to the repeated 30 common trajectories. White circles denote the transition in response between past (blue circles) and future (red circles). The information about the identity of the common trajectory, or the past information, can be estimated from the blue portion of the information curve. The bound on the maximal amount of information about the future a group response with a given amount of information about the past stimulus is denoted with the red line. **b** Population data for 100 random groups for each of the 49 recorded cells (for a total of 4900 5-cell groups). Shaded areas indicate mean  $\pm$  1SD. The mean information about the past across all groups is indicated by the dark blue line, and the corresponding maximal information about the future is denoted with a red line. (inset) For a  $\Delta t$  of 16ms, indicated by the arrow, the predictive information present in 1000 groups of 5 retinal ganglion cells (RGCs, red dots), as well as for model neurons fit with linear-nonlinear (LN) models (black dots), and the groups containing each cell that came closest to the bound (red crosses). Some LN model groups have more information about the past because the LN model does not capture the low-level noise in the response of the RGCs. Other model groups have less info than observed in real data because they fail to capture the stimulus driven response of these cells. No LN model groups capture as much information about the future of the stimulus as the most predictive groups of the cells in the retina. Data here are reproduced from [42] (Color figure online)



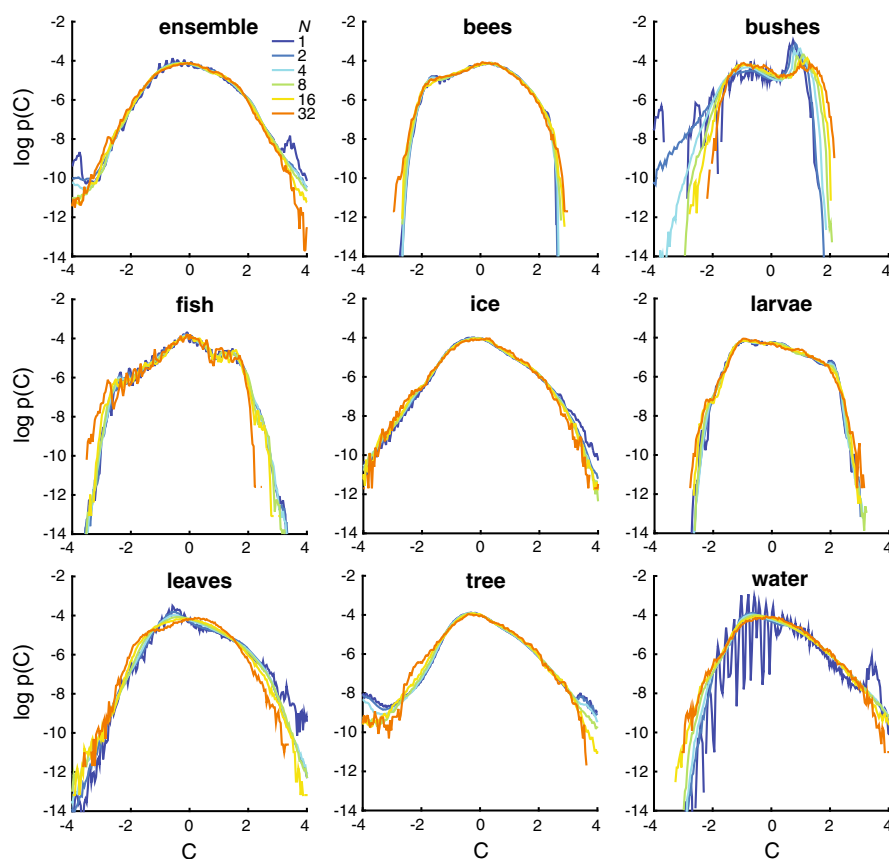


**Fig. 2** Natural motion has a heavy-tailed temporal power spectrum. **a** Representative frames from 3 natural movies: bees on a honeycomb (a plate of glass exposes the hive from the side); pack ice flowing in Lake Michigan; tree blowing in the wind. **b** Temporal power spectra for 8 natural movie clips. Spectra were computed for each pixel using a sliding Hamming window of 256 frames with 50 % overlap, then averaged across pixels. Spectra were normalized so that the integral under the curve is 1, yielding a power spectral density plot. **c** Average temporal power spectral density (solid line) across the 8 clips, with dashed lines indicating  $\pm 1$  SD (Color figure online)

## 4 Towards Naturalistic Motion Stimuli

Natural scenes have heavy-tailed distributions of many quantities of interest, including intensity, contrast, and temporal modulation frequency [15, 27, 40, 46]. This means that there exists no single length scale in space or time one can use to coarse-grain natural scenes without sacrificing large amounts of structure in the data, and that potentially salient fluctuations exist on all scales. We illustrate some basic statistics of natural scenes, taken from our own database of natural movies. In Fig. 2, we show some example frames from three of our clips, along with their temporal modulation spectra. In Fig. 2c, we plot the average power spectrum across 8 representative movie clips. Fitting these data to a power law,  $1/f^\alpha$ , yields an  $\alpha = 1.16$ . This agrees with data from Billock et al. [15], who computed this same function for the movie *Bladerunner* and other Hollywood movies, and with Dong & Atick [27], who also analyzed Hollywood movies in addition to natural movies they recorded.

The distribution of contrast in natural scenes is also known to be heavy-tailed [46]. We plot contrast distributions for 8 individual movie clips as well as for the ensemble of all clips in Fig. 3, for a variety of spatial scales. While individual scenes have varying slope exponential tails, the ensemble of all clips shows a fairly consistent contrast distribution across many length scales. The particular distribution of contrasts and its behavior at different



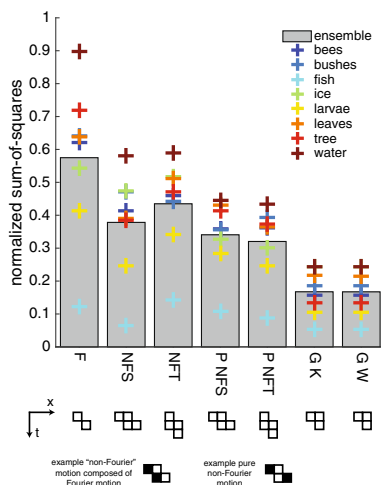
**Fig. 3** Contrast distributions in the motion database are heavy tailed and vary somewhat in shape across clips. (*upper left*) Contrast distribution for an ensemble of 8 natural movie clips of 20 s each, filmed at 60 Hz. The contrast of each pixel is defined as  $C = I/I_0$  where  $I$  is the 8-bit intensity value for that pixel and  $I_0$  is chosen for each frame such that the average contrast for that frame is 0, as in Ruderman [46]. Distributions were estimated at different spatial scales by averaging contrast over  $N \times N$  blocks. Contrast distributions are normalized to have unit variance. Individual contrast distributions are also plotted in the same fashion and labeled by clip (Color figure online)

scales may be important for organisms that only experience a restricted visual environment.

The heavy-tailed nature of natural scene statistics do not offer up a coarse-graining length scale over which we might smooth our inputs, and it is not possible to measure the information content of a neuron's response to the pixel-by-pixel representation of an image sequence of any useful size. Proxies for this calculation include computing information about time within a long and ergodic stimulus sequence [17], but a better approach is to find some reduced and parametrizable (and therefore sample-able) representation of the motion present in the natural environment that retains the features relevant to the organism.

We also computed several higher-order motion features of our selection of natural motion clips. Following the local template motion estimation procedure of [39], we reproduce their finding that natural clips have the same ratio of motion components, as shown here in Fig. 4.

To fully test whether the retina conveys information to the brain in a way that is optimized for prediction, we must consider what prediction problems the retina evolved to solve. In



**Fig. 4** Higher-order motion statistics for clips from the database have similar ratios of component motion. “Rule match opponent” (RMO) scores for 8 natural movie clips, and their combined (“ensemble”) scores. RMO scores are the average squared output of multi-point correlators of a given template and its rotations combined in an opponent fashion with binarized natural movie input, normalized by the average output with uncorrelated (white noise) input. The templates investigated were Fourier (F), non-Fourier spatial (NFS) and temporal (NFT), pure non-Fourier spatial (P NFS) and temporal (P NTF), and black (G K) and white (G W) gliders. The spatiotemporal arrangements of pixels in each template are diagrammed below. Non-Fourier templates will detect correlations caused by Fourier motion in addition to purely non-Fourier motion; thus, pure non-Fourier spatial (P NFS) and temporal (P NTF) scores were computed by excluding these Fourier motion patterns. Glider template scores depend on whether black (G K) or white (G W) is assigned positive polarity. See [39] for details (Color figure online)

particular, if the retina solves only a subset of the possible spatiotemporal prediction problems that could possibly be contained in its input, it should solve those that are present in the natural environment. It could even be the case that the retinas of different animals evolved to encode most efficiently prediction on the scales and correlation structures present in their respective ecological niches. To test this, we need a framework for enumerating the prediction problems present in natural visual stimuli.

#### 4.1 Quantifying Statistics of Natural Motion

Early theories of efficient and predictive coding focus on the pairwise correlation statistics of natural scenes, particularly their spatial and temporal power spectra, to explain the linear filter properties of sensory neurons in terms of whitening or decorrelation [3, 23, 28, 53]. From an information theoretic perspective, this is similar to solving the information bottleneck problem for jointly Gaussian stimuli, in which case the predictive features of the stimulus are completely determined by these pairwise statistics, and the solution boils down to an eigen-decomposition similar to canonical correlation analysis, i.e., a set of linear filters [19]. For natural scenes, with highly non-Gaussian statistics, this may be a good solution to first order, but their additional structure compared to correlated noise gives us hope of finding better solutions by leveraging higher order statistics.

The structure of natural scenes ultimately derives from the configuration of various objects and light sources in the environment relative to the observer. A useful abstraction for thinking about vision at this level is the concept of the plenoptic function, i.e., a complete holographic representation of the environment that can be observed at every point in space and time and

from every viewing position [1]. From this omniscient perspective (given a perfect model of the current state of the world), the dimensionality of the prediction problem is drastically reduced, from millions of individual pixels or photoreceptors down to the trajectories of moving objects and the observer. For example, an observer moving through a static environment with a smooth trajectory will experience highly predictable visual input, since the motion induces spatial transformations whose effects are easily predictable given the current input and an accurate estimate of these transformations [38].

*Self-motion* Spatial transformations due to self-motion involve all points in visual space at once, and are therefore fundamentally higher-order, but they also affect the pairwise statistics in characteristic ways. For example, the spatiotemporal rearrangement of static stimuli caused by fixational eye movements leads to a whitening of the power spectrum, which is itself a form of predictive coding through natural statistics matching [36,50]. Eye [45], head, and body movements will all affect the input to the retina in predictable ways, but will generally differ from system to system. These retinal trajectories through visual space have been measured well in the case of head-fixed macaques and humans [37,52], unrestrained humans [5], and freely moving rats [59]. In addition, there are non-visual sources for estimation of self-motion, such as efference copies of motor commands [22], which may play an important role in this kind of prediction.

*Object motion* The motion of objects in the world will cause local translations of pixels that contrast with global self-motion signals, and selectivity for such opponent motion signals is already observed in some retinal ganglion cells [41]. Objects in the world are subject to physical laws, such as gravity and inertia, making their trajectories predictable. One might think this kind of knowledge of physical laws requires complex cortical processing, but the fact that retinal ganglion cells can anticipate objects moving with constant velocities [11] is arguably an instantiation of knowledge about inertia. Object trajectories are potentially measurable from scenes using object tracking algorithms, allowing for the comparison of a neural system's prediction performance in response to stimuli that do and do not match the statistics of the animal's natural environment.

*Higher-order statistics of motion* Deciding how to quantify a natural scene can be challenging. The high dimensionality of natural inputs to the visual system means that direct approaches to quantifying the information transmission are wrought with sampling error pitfalls or completely impossible. Making educated guesses about what features of natural motion to quantify, or searching directly for a lower dimensional representation of the structure of natural scenes are two promising approaches to this problem. Recent work has defined a set of motion primitives, correlation structures in space and time that find their basis in early theories of motion processing. The Fourier, non-Fourier, and glider components of local motion can be readily computed from natural movies [33,39]. This approach reveals that certain ratios of these components may be prevalent in natural scenes [39]. The brain might make use of this fact to tailor its motion processing to just these types of input. If so, deviations from this natural ratio should lead to noisier, less efficient coding for the future stimulus. Downstream of the retina, this could lead to motion perception deficits.

Such local motion signatures only characterize part of the total motion signal in natural scenes. Machine learning efforts have been launched to find longer-range, collective components of natural image and motion statistics [18,30,31,47]. Work from these groups has shown that some physical models of long range fluctuations may be applicable to natural motion. This is exciting because it could lead to a generative model for such motion, opening

up the possibility of more stringent, parametrized tests of optimal coding for motion in the retina.

## 5 Discussion

Testing theories of optimal prediction in the visual stream requires an integration of existing theories of optimal coding in the retina and beyond, with a careful quantification of the motion statistics present in the natural environment. By examining how well information processing in the brain is tuned to natural inputs, we may discover new hard-wired and adaptive features of the predictive part of the neural code.

While we have focused here on processing in the retina, we hypothesize that a similar information bottleneck problem may be solved in a hierarchical fashion by subsequent visual areas, such as the visual cortex. Effective stacking of optimal prediction stages within a recurrent network in a single area is also possible. By iterating this kind of predictive architecture, the visual system may build up longer-time predictions, eventually bridging to the seconds- and minutes-long timescales underlying decision making and reward.

Differential encoding of predictable versus non-predictable stimuli have been noted in fMRI data both in the visual cortex and in the striatum [2, 25], supporting some combination of predictive coding (after a stimulus becomes predictable) and explicit representation of predictive information (while stimulus statistics are learned or adapted to).

It is possible that the same principles at work in the visual system, a uniquely well studied and tractable set of brain areas, extend to other sensory modalities, such as audition and somatosensation. The theory of optimal coding for predictive information in a changing environment generalizes to any input stream and can be formulated for any timescale of relevance in that channel. Complex behaviors rely on integrating input from many sensory streams and a common pre-processing architecture might facilitate that integration. Ultimately, directly tying efficient coding for prediction in early sensory areas to accurate behavioral outcomes will be the most satisfying test of these theories.

**Acknowledgments** We thank E. Nitzany for sharing his recent template motion analysis of movie clips from our database and D. J. Schwab for comments on the manuscript. Support for this work was provided by: The Alfred P. Sloan Foundation, a FACCTS Grant from the France Chicago Center, and a Chateaubriand Fellowship to JS.

## References

1. Adelson, E.H., Bergen, J.R.: The plenoptic function and the elements of early vision. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology (1991)
2. Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L.: Stimulus predictability reduces responses in primary visual cortex. *J. Neurosci.* **30**(8), 2960–2966 (2010)
3. Atick, J.J., Redlich, A.N.: What does the retina know about natural scenes? *Neural Comput.* **4**(2), 196–210 (1992)
4. Attneave, F.: Some informational aspects of visual perception. *Psychol. Rev.* **61**(3), 183–193 (1954)
5. Aytekin, M., Victor, J.D., Rucci, M.: The visual input to the retina during natural head-free fixation. *J. Neurosci.* **34**(38), 12701–12715 (2014)
6. Barlow, H.B.: Summation and inhibition in the frog's retina. *J. Physiol.* **119**(1), 69–88 (1953)
7. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. *Sensory Communication*, 217–234 (1961)
8. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J.: Canonical microcircuits for predictive coding. *Neuron* **76**(4), 695–711 (2012)

9. Bell, A.J., Sejnowski, T.J.: The independent components of natural scenes are edge filters. *Vis. Res.* **37**(23), 3327–3338 (1997)
10. Berger, T.: Rate-Distortion Theory. Encyclopedia of Telecommunications. Prentice-Hall, Englewood Cliffs (1971)
11. Berry, M.J., Brivanlou, I.H., Jordan, T.A., Meister, M.: Anticipation of moving stimuli by the retina. *Nature* **398**(6725), 334–8 (1999)
12. Berry, M.J., Schwartz, G.: Predictions in the Brain: Using Our Past to Generate a Future. The retina as embodying predictions about the visual world. Oxford University Press, Oxford (2011)
13. Bialek, W., Nemenman, I., Tishby, N.: Predictability, complexity, and learning. *Neural Comput.* **13**(11), 2409–2463 (2001)
14. Bialek, W., de Ruyter van Steveninck, R.R., Tishby, N.: Efficient representation as a design principles for neural coding and computation. In: Proceedings of the International Symposium on Information Theory 2006 ([arXiv:0712.4381](https://arxiv.org/abs/0712.4381) [q-bio.NC] (2007)) (2006)
15. Billock, V.A., de Guzman, G.C., Kelso, J.S.: Fractal time and 1/f spectra in dynamic images and human vision. *Phys. D* **148**(1), 136–146 (2001)
16. Boerlin, M., Machens, C.K., Denève, S.: Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput. Biol.* **9**(11), e1003258 (2013)
17. Brenner, N., Strong, S.P., Koberle, R., Bialek, W., de Ruyter van Steveninck, R.R.: Synergy in a neural code. *Neural Comput.* **12**(7), 1531–1552 (2000)
18. Cadieu, C.F., Olshausen, B.A.: Learning intermediate-level representations of form and motion from natural movies. *Neural Comput.* **24**(4), 827–866 (2012)
19. Chechik, G., Globerson, A., Tishby, N., Weiss, Y.: Information bottleneck for gaussian variable. *JMLR* **6**, 165–188 (2005)
20. Chichilnisky, E.: A simple white noise analysis of neuronal light responses. *Network* **12**(2), 199–213 (2001)
21. Creutzig, F., Globerson, A., Tishby, N.: Past-future information bottleneck in dynamical systems. *Phys. Rev. E* **79**(4), 041925 (2009)
22. Crowell, J.A., Banks, M.S., Shenoy, K.V., Andersen, R.A.: Visual self-motion perception during head turns. *Nature Neurosci.* **1**(8), 732–737 (1998)
23. Dan, Y., Atick, J.J., Reid, R.C.: Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* **16**(10), 3351–3362 (1996)
24. Denève, S.: Bayesian spiking neurons i: inference. *Neural Comput.* **20**(1), 91–117 (2008)
25. den Ouden, H.E., Daunizeau, J., Roiser, J., Friston, K.J., Stephan, K.E.: Striatal prediction error modulates cortical coupling. *J. Neurosci.* **30**(9), 3210–3219 (2010)
26. Doi, E., Gauthier, J.L., Field, G.D.: Efficient coding of spatial information in the primate retina. *J. Neurosci.* **32**(46), 16256–16264 (2012)
27. Dong, D.W., Atick, J.J.: Statistics of natural time-varying images. *Network* **6**(3), 345–358 (1995)
28. Dong, D.W., Atick, J.J.: Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network* **6**(2), 159–178 (1995)
29. Golisch, T., Meister, M.: Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* **65**(2), 150–164 (2010)
30. Häusler, C., Susemihl, A.: Temporal autoencoding restricted boltzmann machine. *arXiv preprint [arXiv:1210.8353](https://arxiv.org/abs/1210.8353)* (2012)
31. Häusler, C., Susemihl, A., Nawrot, M.P.: Natural image sequences constrain dynamic receptive fields and imply a sparse code. *Brain Res.* **1536**, 53–67 (2013)
32. Hosoya, T., Baccus, S.A., Meister, M.: Dynamic predictive coding by the retina. *Nature* **436**(7047), 71–77 (2005)
33. Hu, Q., Victor, J.D.: A set of high-order spatiotemporal stimuli that elicit motion and reverse-phi percepts. *J. Vis.* **10**(3), 9 (2010)
34. Kastner, D.B., Baccus, S.A.: Spatial segregation of adaptation and predictive sensitization in retinal ganglion cells. *Neuron* **79**(3), 541–554 (2013)
35. Kilner, J.M., Friston, K.J., Frith, C.D.: Predictive coding: an account of the mirror neuron system. *Cogn. Process.* **8**(3), 159–166 (2007)
36. Kuang, X., Poletti, M., Victor, J.D., Rucci, M.: Temporal encoding of spatial information during active visual fixation. *Curr. Biol.* **22**(6), 510–514 (2012)
37. Mukherjee, T., Battifarano, M., Simoncini, C., Osborne, L.C.: Shared sensory estimates for human motion perception and pursuit eye movements. *J. Neurosci.* **35**(22), 8515–8530 (2015)
38. Nakaya, Y., Harashima, H.: Motion compensation based on spatial transformations. *IEEE Trans. Circuits Syst. Video Technol.* **4**(3), 339–356 (1994)

39. Nitzany, E.I., Victor, J.D.: The statistics of local motion signals in naturalistic movies. *J. Vis.* **14**(4), 1–15 (2014)
40. Olshausen, B.A., Field, D.J.: Natural image statistics and efficient coding\*. *Network* **7**(2), 333–339 (1996)
41. Ölveczky, B.P., Baccus, S.A., Meister, M.: Segregation of object and background motion in the retina. *Nature* **423**(6938), 401–408 (2003)
42. Palmer, S.E., Marre, O., Berry, M.J., Bialek, W.: Predictive information in a sensory population. *Proc. Natl. Acad. Sci.* **112**(22), 6908–6913 (2015)
43. Puchalla, J.L., Schneidman, E., Harris, R.A., Berry, M.J.: Redundancy in the population code of the retina. *Neuron* **46**(3), 493–504 (2005)
44. Rao, R.P., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neurosci.* **2**(1), 79–87 (1999)
45. Rucci, M., Victor, J.D.: The unsteady eye: an information-processing stage, not a bug. *Trends Neurosci.* **38**(4), 195–206 (2015)
46. Ruderman, D.L.: Origins of scaling in natural images. *Vision Res.* **37**(23), 3385–3398 (1997)
47. Saremi, S., Sejnowski, T.J.: Hierarchical model of natural images and the origin of scale invariance. *Proc. Natl. Acad. Sci.* **110**(8), 3071–3076 (2013)
48. Schwartz, G., Harris, R., Shrom, D., Berry, M.J.: Detection and prediction of periodic patterns by the retina. *Nature Neurosci.* **10**(5), 552–554 (2007)
49. Schwartz, G., Taylor, S., Fisher, C., Harris, R., Berry, M.J.: Synchronized firing among retinal ganglion cells signals motion reversal. *Neuron* **55**(6), 958–969 (2007)
50. Segal, I.Y., Giladi, C., Gedalin, M., Rucci, M., Ben-Tov, M., Kushinsky, Y., Mokeichev, A., Segev, R.: Decorrelation of retinal response to natural scenes by fixational eye movements. *Proc. Natl. Acad. Sci.* **112**(10), 3110–3115 (2015)
51. Shannon, C.E.: A mathematical theory of communication. *Bell Sys. Tech. J* **27**(379–423), 623–656 (1948)
52. Spering, M., Gegenfurtner, K.R.: Contextual effects on motion perception and smooth pursuit eye movements. *Brain Res.* **1225**, 76–85 (2008)
53. Srinivasan, M.V., Laughlin, S.B., Dubs, A.: Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lon. B* **216**(1205), 427–459 (1982)
54. Srinivasan, R., Rao, K.: Predictive coding based on efficient motion estimation. *IEEE Trans. Commun.* **33**(8), 888–896 (1985)
55. Stephens, G.J., Mora, T., Tkačik, G., Bialek, W.: Statistical thermodynamics of natural images. *Phys. Rev. Lett.* **110**(1), 018701 (2013)
56. Tishby, N., Pereira, F.C., Bialek, W.: The information bottleneck method. In: *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, 37 ([arXiv:physics/0004057](https://arxiv.org/abs/physics/0004057) (2000)), 368–377 (1999)
57. van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lon. B* **265**(1394), 359–366 (1998)
58. Wacongne, C., Changeux, J.P., Dehaene, S.: A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* **32**(11), 3665–3678 (2012)
59. Wallace, D.J., Greenberg, D.S., Sawinski, J., Rulla, S., Notaro, G., Kerr, J.N.: Rats maintain an overhead binocular field at the expense of constant fusion. *Nature* **498**(7452), 65–69 (2013)