General Principles for Sensory Coding 🔓

Tatyana O. Sharpee, Computational Neurobiology Laboratory, Salk Institute for Biological Studies

Summary

Sensory systems exist to provide an organism with information about the state of the environment that can be used to guide future actions and decisions. Remarkably, two conceptually simple yet general theorems from information theory can be used to evaluate the performance of any sensory system. One theorem states that there is a minimal amount of energy that an organism has to spend in order to capture a given amount of information about the environment. The second theorem states that the maximum rate with which the organism can acquire resources from the environment, relative to its competitors, is limited by the information this organism collects about the environment, also relative to its competitors.

These two theorems provide a scaffold for formulating and testing general principles of sensory coding but leave unanswered many important practical questions of implementation in neural circuits. These implementation questions have guided thinking in entire subfields of sensory neuroscience, and include: What features in the sensory environment should be measured? Given that we make decisions on a variety of time scales, how should one solve trade-offs between making simpler measurements to guide minimal decisions vs. more elaborate sensory systems that have to overcome multiple delays between sensation and action. Once we agree on the types of features that are important to represent, how should they be represented? How should resources be allocated between different stages of processing, and where is the impact of noise most damaging? Finally, one should consider trade-offs between implementing a fixed strategy vs. an adaptive scheme that readjusts resources based on current needs. Where adaptation is considered, under what conditions does it become optimal to switch strategies? Research over the past 60 years has provided answers to almost all of these questions but primarily in early sensory systems. Joining these answers into a comprehensive framework is a challenge that will help us understand who we are and how we can make better use of limited natural resources.

Keywords: efficient coding, information theory, adaptation, predictive coding, predictive information, receptive fields

Subjects: Computational Neuroscience, Sensory Systems

How Sensory Systems Increase Fitness

Why are brains and, in particular, sensory systems needed? The answer to this question is sometimes motivated exclusively in the context of movement (Wolpert, 2011): Without the possibility of movement, there is no need for the brain. And in fact, there are creatures, such as sea squirts, that digest their own brain and spinal cord once they have chosen a stone to settle permanently on. Expanding the definition of movement to include chemical communication, one can begin to view intracellular signal processing as fulfilling the roles of both the sensory and motor systems in the brain (Bray, 1995). It is certainly possible to move in a random fashion without relying on sensing, as some species do when foraging (Viswanathan, da Luz, Raposo, & Stanley, 2011). Under what circumstances then are sensory systems explicitly needed?

Information Rate Determines Maximal Population Growth Rate

Using information theory, the answer to this question can be given in surprisingly general terms (Bialek, 2013; Hilbert, 2017; Kelly, 1956) that apply across organisms, as well as human population activities, such as economics. For example, many aspects of animal behavior such as pursuit-evasion games can be described using game theory (Isaacs, 1999). Suppose that each day a herd of animals can access water at one of two locations. These locations are also familiar to a hunter or predators, who should also decide which location to explore. If escape is easier at one location, then provided this decision is carried out just once, all prey animals should go to that location. However, this strategy would not be optimal in the long term because the predator would know to go straight to that location (Gal, Alpern, & Casas, 2015; Gal & Casas, 2014). So, to find the optimal long-term solution, let us denote the fraction of prey animals getting water at locations 1 and 2 as f_1 and f_2, the probability that the hunter visits each of these two locations as p_1 and p_2, and the relative population growth at these two locations as g_1 and g_2. Starting with initial number of animals N in the herd, after one day the number of animals will be $N g_2 f_2$ if the hunter visited location 1 killing all animals there and $N g_1 f_1$ if the hunter visited location 2. Using a binary variable, n_1 takes the value of 1 if location one comes under attack or value 0 if location 2 is attached, both of the possible outcomes can be summarized using a factor $F_1 = n_1 g_2 f_2 + (1 - n_1) g_1 f_1$, which describes a relative change in the population. For the analysis of long-term gain or loss, it is convenient to write the factor F_1 as an exponential taking advantage of the binary character of variable n_1:

$$F_1 = \exp\left(n_1 \ln\left(g_2 f_2\right) + \left(1 - n_1\right) \ln\left(g_1 f_1\right)\right).$$

After many days, the change in the number of animals will be given by a product of factors F_i, where index i denotes a day. In the long term, the average change per day will be given by:

$$\exp\left(\Lambda\right) = \exp\left(p_1 \ln\left(g_2 f_2\right) + p_2 \ln\left(g_1 f_1\right)\right)$$

(1)

To find the optimal strategy that minimizes the expected loss, we differentiate taking into account that f1+f2=1. This yields:

$$\frac{p_1}{1 - f_1} = \frac{p_2}{f_1}.$$

Or equivalently, $\frac{p_1}{p_2} = \frac{1 - f_1}{f_1}$, which allows for a simple solution:

$$p_1 = 1 - f_1 = f_2, p_2 = 1 - p_1 = f_1$$

(2)

We have thus recovered a version of the matching law, which animals have been observed to follow in a number of situations (Poling, Edwards, Weeden, & Foster, 2011). Interestingly, in this situation the optimal allocation of animals between locations does not depend on the gains/resources available at each location, only by the probability of loss/win. Further, the result generalizes to the case of many possibilities as $p_i = 1 - f_i$ as long as one can define a unique gain associated with each outcome (or if gains are location-independent). Substituting the optimal time allocation to the expression (1) for the resource rate, we find that is given by:

$$\exp\left(\Lambda\right) = \exp\left(p_1 \ln g_2 + \left(1 - p_1\right) \ln g_1 + p_1 \ln p_1 + \left(1 - p_1\right) \ln\left(1 - p_1\right)\right).$$

The first two terms in this expression were set by the gains and are not under prey animals' control. The second two terms together correspond to the negative of the entropy of a binary process (Cover & Thomas, 1991):

$$H_{binary}\left(p\right) = -plnp - \left(1 - p\right) \ln\left(1 - p\right).$$

The entropy is a positive quantity, and so this is the limiting factor that sets the maximal rate of resource accumulation:

$$\Lambda = p_1 lng_1 + \left(1 - p_1\right) lng_2 - H_{binary}\left(p\right).$$

(3)

Before we discuss the implication of this equation for sensory systems and the brain, let me note in passing that from the hunter's point of view the maximal gain is obtained when their behavior is as random as possible in order to maximize $H_{binary}(p)$. In this toy problem this means visiting the two locations with equal probabilities of 0.5. It turns out that humans are especially bad in generating random sequences compared to other primates (Henrich, 2015; Martin, Bhui, Bossaerts, Matsuzawa, & Camerer, 2014). To compensate for this, many hunting rituals involve procedures that serve to randomize the behavior. For example, as described in (Henrich, 2015) the hunting location is selected by Naskapi foragers in Labrador, Canada based on the pattern of cracks in bones heated over fire or by the type of bird augury. Without such rituals, the hunter would tend to choose one location more often than the other, and this could then be exploited by the prey animals to reduce their loss.

This brings us to the problem of estimating probabilities p_1 and p_2. In reality these probabilities are not given and need to be estimated from experience. At least initially, if we are working with a stationary process, the estimated values q_1 and q_2 for these probabilities will deviate from the true probabilities p_1 and p_2. How will this impact the growth rate? Instead of $H_{binary}(p)$ we will obtain:

$$\Lambda_{\text{est}} = p_1 ln q_2 + (1 - p_1) ln q_1 + p_1 ln q_1 + (1 - p_1) \ln(1 - q_1)$$

(4)

The difference between the truly optimal growth rate Λ and the one we can achieve based on estimated probabilities is:

$$\Lambda - \Lambda_{\text{est}} = p_1 ln \frac{p_1}{q_1} + (1 - p_1) \ln \frac{1 - p_1}{1 - q_1} = D_{KL}(p \parallel q)$$

(5)

This quantity is known as the Kullback–Leibler (KL) distance and it specifies the cost of misestimating one probability *p* by the probability *q*. This function is always positive, reaching zero only when probabilities coincide. Thus, there is a tangible cost in terms of growth for misestimating probability values.

The analysis above was done for the stationary distribution where the probabilities were constant in time. So far we have not answered the question of why the brain and sensory systems would be useful from the perspectives of the information and game theories. The sensory systems can increase overall fitness by delivering more accurate probabilities that are specific to a given sensory situation. Let us analyze this mathematically. Suppose that probabilities p_1 and p_2 vary in time in a manner that correlates with some environmental

variable x. Now instead of using average probabilities p_1 and p_2 across different environmental conditions, we can rely on more accurate estimates by measuring environmental variable x. The increased growth rate from observing x is:

$$\Delta_x - \Delta = D_{KL}\left(p\&x \parallel p\right) > 0$$

(6)

The KL distance in Eq. (7) is the Shannon mutual information between the attach probability p and the environmental variable x. It is worth writing down this expression explicitly:

$$\Delta_x - \Delta = \int dx P\left(x\right)\left[p_1 \ln\left[\frac{p_1\left(x\right)}{p_1}\right] + \left(1 - p_1\right)\ln\left[\frac{1 - p_1\left(x\right)}{1 - p_1}\right]\right],$$

(7)

where $p_1\left(x\right)$ specifies how the attack probability at location 1 depends on the variable x, with the average probability $p_1 = \int dx\, p_1\left(x\right) P\left(x\right)$ being an integral over the input probability distribution $P\left(x\right)$. Eq. (8) demonstrates that the effectiveness of any sensory system or its part can be quantified as the increase in the maximum rate of resource accumulation.

Finally, let use briefly discuss the impact that gains parameters g_1 and g_2 have on the growth rate. Imagine that they are set by or dependent on an adversarial opponent. For example, there can be another prey species consuming the same resources, and the population growth rate $g1$ and $g2$ at each location can be inversely related to the total number of animals of both kinds: $g_1 \propto \frac{1}{f_1 + \tilde{f}_1}$ If both species have access to correct probabilities, on average their numbers will increase similarly. However, if one species can estimate the probabilities better, by for example, investing in better sensory systems, then it will gain a systematic advantage over the other. This advantage will lead to an exponential gain in resources over the long term. This is then what provides a conceptual impetus for designing the sensory systems.

The Metabolic Cost of Information Transmission

The discussion so far focused on the minimum information that the organism needs to obtain on the current state of the environment in order to attain a specified growth rate. However, there is a complimentary requirement of minimum energy that the animal needs to spend in order to transmit a given number of bits. This bound is specified by the so-called channel coding theorem (Cover & Thomas, 1991). The metabolic requirements associated with spiking have also been quantified (Laughlin, de Ruyter van Steveninck, & Anderson, 1998).
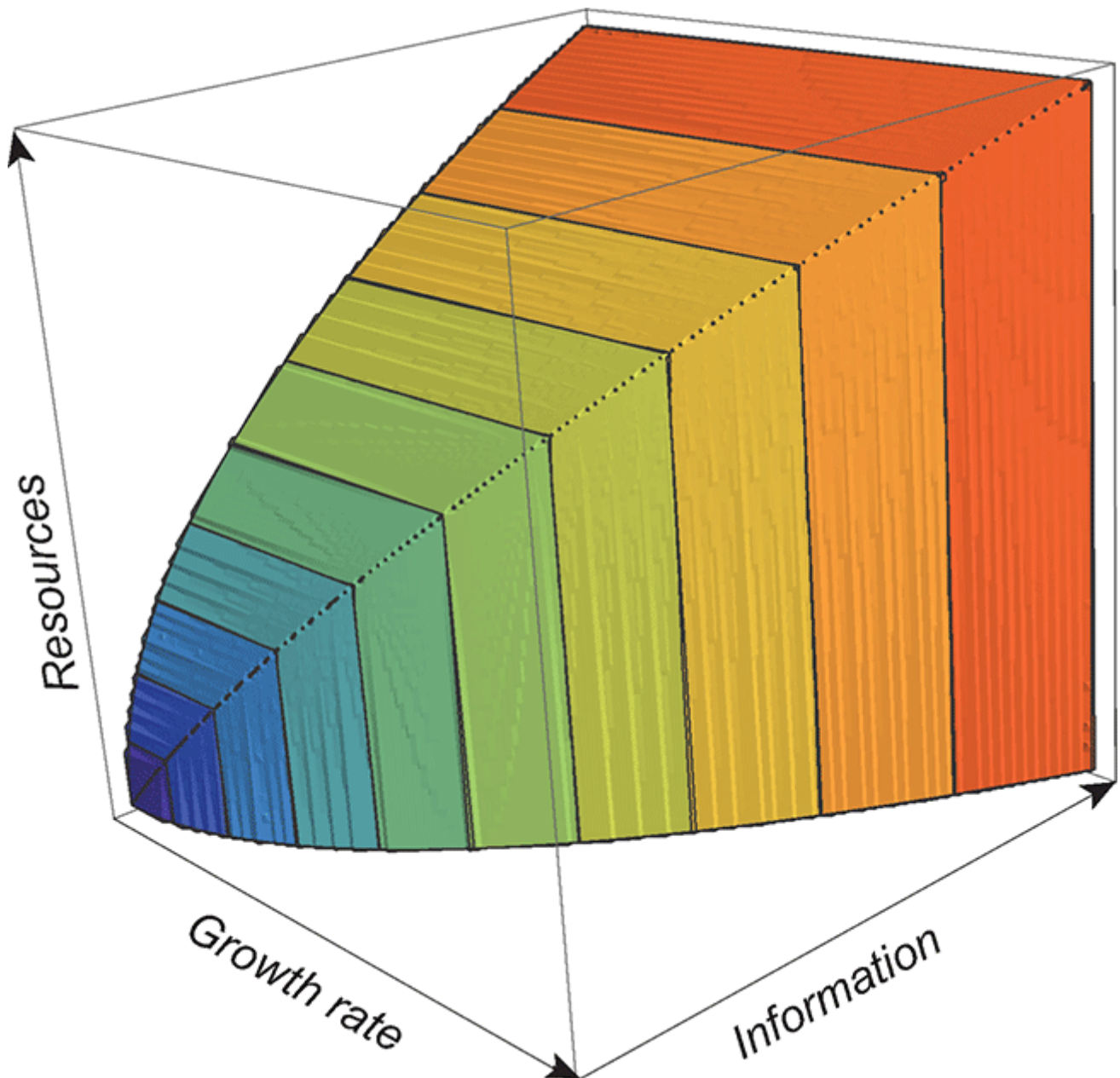
Figure 1. To sustain a given growth rate, the organism needs to transmit a minimum amount of information in bits. This in turn determines a minimum amount of resource or metabolic rate.

Overall, information theory provides a quantitative framework for allocating resources in the organism: the minimum amount of energy that needs to be invested in a given sensory neural circuit can be related to the amount of gain in metabolic resources that such sensory measures could potentially produce (Figure 1). One must also factor in the cost of making computations based on the sensory measurements that are necessary to implement optimal behavioral strategy (Tishby & Polani, 2011), but we will not discuss this aspect further as the

discussion will take us away from sensory systems. Nevertheless, we arrive at a general description of the optimal allocation of resources between different sensory systems based on the amount of gain in future growth that they could produce.

To summarize, information theory makes it possible to compartmentalize the analysis of sensory circuits by asking (1) how much information a given sensory measurement provides in general and about specific possible action, (2) how costly that action is metabolically, and separately (3) how effective motor and behavioral systems are in making optimal decisions based on the available sensory evidence. This compartmentalization is especially useful for the analyses of large neural systems where the space of both possible sensory measurements and behavioral choices is large. In what follows, we will discuss insights gained into the organization of the sensory circuits from the application of information theoretic principles.

Predicting Sensory Features That Should Be Represented in Neural Circuits

Natural signals have high dimensionality but also exhibit structure on a variety of time scales. Thus, various aspects of sensory signals differ in the amount of relevant information they convey for a particular action. In simpler neural circuits where sensory and motor actions are more directly related to other, one can expect that sensory circuits will encode those input features that are maximally informative about particular motor actions. Indeed, classic studies in the frog retina have identified a number of neurons that serve to detect predator approach (Lettvin, Maturana, McCulloch, & Pitts, 1957). These include "looming" detectors that trigger on shadows as predators approach from above (Gollisch & Meister, 2010). These types of detectors have also been found in the mammalian retina (Gollisch & Meister, 2010). The representation of looming stimuli in sensory areas beyond the retina has become an active research area (Temizer, Donovan, Baier, & Semmelhack, 2015; Zhao, Liu, & Cang, 2014), perhaps due to the importance of the signals that these neurons encode. The relationship between sensory stimuli and other actions is less direct. For more complex and less hardwired events, one can expect early sensory systems to provide a general representation of sensory events. Because such representations are not tied to a specific action, one could ask which features of the natural signals carry the maximal information about the (uncompressed) input pattern. The answers derived using theoretical analyses based on the statistics of natural signals have accounted for properties of neural responses in the retina (Atick & Redlich, 1990, 1992; Bialek & Owen, 1990), thalamus (Dong & Atick, 1995; Dan, Atick, & Reid, 1996), the primary visual cortex (Bell & Sejnowski, 1997; Olshausen & Field, 1996, and higher visual areas (Cadieu & Olshausen, 2012). Similarly in the auditory system, the observed neural filters provide maximal information about naturalistic auditory signals (Rieke, Bodnar, & Bialek, 1995) and had spectral characteristics that match maximally informative solutions (Smith & Lewicki, 2006). The statistics of natural stimuli in other sensory modalities are less studied, but one can only assume that neural properties in these other systems are also tuned to maximize information transmission. Experimental verification of this hypothesis will serve to solidify the relevance of information theory to sensory coding.

Optimal Nonlinearities in Sensory Coding

In general, extracting signals relevant to behavior requires nonlinear computations. The nonlinear computations are necessary for at least two reasons. First, nonlinear operations are necessary to make constructs that are not directly provided by the receptor neurons but nevertheless are highly informative about particular events in the environment. One example is the computation of motion, where optimal nonlinearities can be understood from information theoretic principles (Bialek & de Ruyter van Steveninck, 2005; Potters & Bialek, 1994). Another example is contrast invariant representation that can provide stable perception across different light intensity conditions (Ben-Yishai, Bar-Or, & Sompolinsky, 1995). The extraction of ethologically relevant features such as hands and faces (Desimone, Albright, Gross, & Bruce, 1984; Tsao & Livingstone, 2008) in the visual system or selectivity to bird's own song in the auditory system (Doupe & Kuhl, 1999) constitutes a third example.
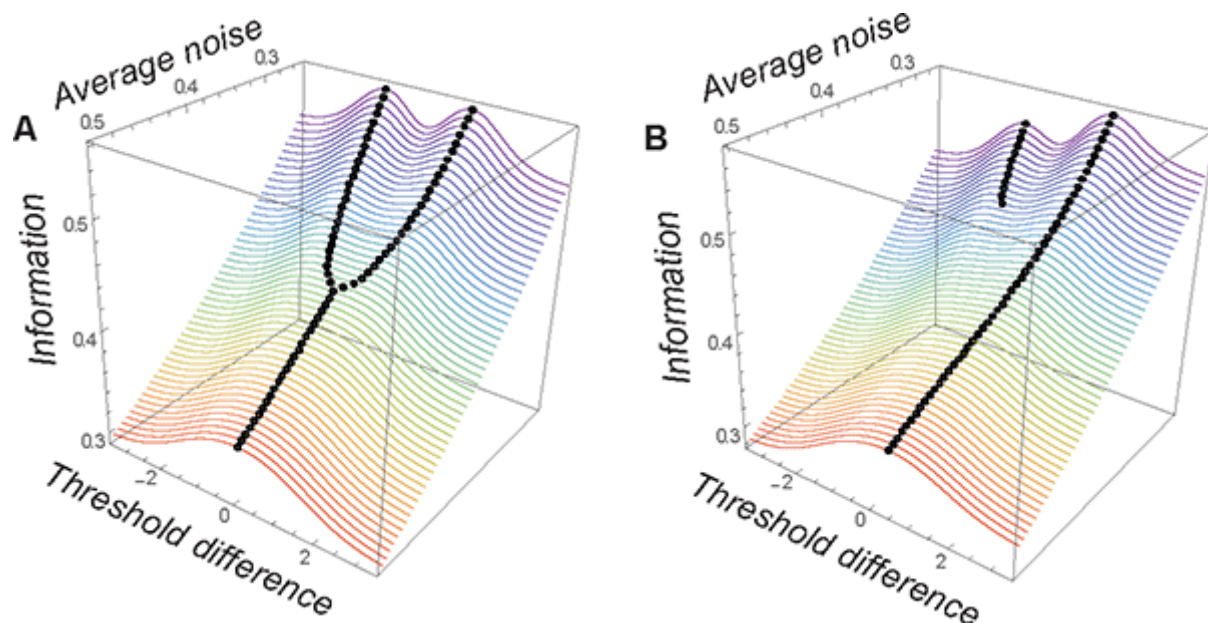
Figure 2. Optimal nonlinear codes depend on the amount of input noise. Information transmitted by a pair of neurons is plotted as a function of average noise in the two pathways and the difference in thresholds between neurons. Both neurons have the same amount of input noise in (A) and finite difference in noise in (B). All curves are plotted for a fixed total spike rate of the two neurons across a range of inputs, described by a Gaussian function. Threshold values and noise are given in units of the input standard deviation. In (A) there is a critical amount of noise below which neuronal specialization into two classes occurs. The two maxima are equivalent because both pathways have the same amount of noise. In (B) finite noise differences adds a slope to the picture in (A). Breaking the symmetry between neurons, it becomes optimal to associate positive thresholds differences with positive noise differences. In other words, neuronal pathway with lower input noise should optimally have a threshold closer to the mean of the input distribution.

Second, nonlinear processing can greatly improve the efficiency of the nervous system, where efficiency is defined as the number of bits transmitted for a given metabolic cost. For example, one can derive an optimal ratio between the number of analogue and digital operations that can minimize the overall metabolic cost (Sarpeshkar, 1998). Further, even the specific shape of the nonlinear function can significantly increase the amount of information transmitted if it is adjusted to properly match the input distribution. Under the assumption of small additive noise, the optimal nonlinearity is given by a cumulative distribution of input signals (Bialek, 2013; Laughlin, 1983). In a classic study, Laughlin demonstrated that the nonlinearity of neurons in the retina did indeed match the cumulative distribution of contrast in the natural environment (Laughlin, 1983). This result can be extended to analyze the optimal parameters of nonlinearities for different noise models and numbers of neurons (Gjorgjieva, Sompolinsky, & Meister, 2014; Harper & McAlpine, 2004; Kastner, Baccus, & Sharpee, 2015; McDonnell, Stocks, Pearce, & Abbott, 2005; McDonnell, Stocks, Pearce, & Abbott, 2006; Nikitin, Stocks, Morse, & McDonnell, 2009). For example, when two neurons jointly encode the same sensory feature in the presence of small input noise, there is an optimal difference between their thresholds that depends on the amount of noise (McDonnell et al., 2005, 2006; Nikitin et al., 2009; Gjorgjieva et al., 2014; Kastner et al., 2015). One neuron responds only when the signal magnitude is very strong, whereas the other neurons respond more often, for less strong inputs. The larger the noise, the smaller the optimal separation between thresholds becomes (Figure 2). Curiously, there is a maximum amount of noise above which the optimal difference between thresholds turns to zero and remain zero regardless of any further noise increase. When thresholds are the same, each neuron's encoding properties become indistinguishable and they form the same neuron type, from a functional perspective. Thus, these arguments indicate that the number of different cell types necessary for maximally informative encoding depends on the amount of sensory noise. These predictions can be tested in the retina, where the amount of noise differs between the On pathway, responsible for encoding light increments, and the Off pathway, responsible for encoding light decrements (Balasubramanian & Sterling, 2009; Chichilnisky & Kalmar, 2002; Ratliff, Borghuis, Kao, Sterling, & Balasubramanian, 2010). The sensory noise in the Off pathway is smaller and, importantly, below the critical value for the separation of thresholds. In contrast, noise in the On pathway is higher than the critical value. Thus, maximally informative solutions could explain why there are two Off neuron types reporting the same type of light intensity fluctuations in the negative directions, while only one type of On neurons reports the analogous types of light intensity fluctuations in the positive direction (Kastner et al., 2015). A closer look at the data indicates that there are small but significant differences between the amount of noise in the high and low threshold Off neurons. To account for these differences, the maximally informative solutions were analyzed with nonzero differences in the amount of noise between the two neural pathways (Kastner et al., 2015), cf. Figure 2B. In this case, the maximally informative solutions assign neurons in the pathway with lower thresholds to also have smaller thresholds relative to the mean of the distribution. This prediction agrees with the measurements of different retinal cell types encoding light decrements (Kastner et al., 2015).

Extending these results to encompass more neuron types and testing in larger neural circuits seems to offer a path towards a systematic theory of neuronal types that could explain how different cell types from different species are related to each other and the number of neuronal types that one may expect to find in the brain (Balasubramanian & Sterling, 2009; Sharpee, 2014). Such a comprehensive theory has not been developed to date, but some theories are available for certain limiting cases. For example, for small sensory noise, the thresholds of neurons in a large population would be optimally distributed according to the cumulative distribution of the input signals (Brunel & Nadal, 1998). This results echoes the initial derivation for one neuron in the presence of small additive output noise (Laughlin, 1983). It also demonstrates how calculations carried out in different regimes, such as for one neuron with small output noise and many neurons operating in the presence of small input noise, can be related and amalgamated.

Dynamical Aspects of Neural Processing

The discussion so far has largely ignored the dynamical aspects of neural processing. We discussed that the presence of certain features in the environment, such as a looming stimulus, can serve as indicators of certain events and predicate the corresponding action. However, we have ignored the inevitable delays in signal processing that separate stimulus detection from action. This added perspective suggests that one should look for features in the natural sensory environment that carry the most information about future actions. In simplified terms, we are looking for those stimulus features that are correlated with stimulus values at future times. In representing such features one needs to strike a balance between, on one hand, representing features that have predictive power and are therefore correlated with future stimulus values, and, on the other hand, not sending essentially the same stimulus value when correlations in time become too strong. The arguments that emphasize the first idea are known in sensory neuroscience as extracting *predictive information about the future*, whereas the second constraints is known as *predictive coding*. We discuss these two lines of reasoning in turn.

Predicting the Future

In the case of early sensory systems, one can simplify this argument to look for features of sensory processing that carry the most information about the future state of the environment (Bialek, Nemenman, & Tishby, 2001). Recent work in the retina shows that, at least in controlled experiments where the amount of information between past and future events is known, small groups of retinal neurons provide the maximum amount of information possible about future stimulus values (Palmer, Marre, Berry, & Bialek, 2015).

Making predictions for the future can also lead to more efficient representations within the nervous system itself. This is because neural codes can be structured to report just the deviation between their inputs and what is expected based on both their past own activity and the activity of other neurons in the circuit (Rao & Ballard, 1999). When applied to neurons at

the same level of neural processing, this concept predicts a tendency for nearby neural responses to be more "decorrelated" than their inputs (Barlow, 1959). This concept is also known as redundancy reduction. It is also part of maximally informative code because smaller redundancy between neural responses increases the overall amount of information that is conveyed by the neural population as a whole (Atick, 1993).

Recent studies in the retina provide direct support for this prediction when they find that the responses of retinal ganglion cells are less correlated than their inputs (Simmons et al., 2013). However, early work on the responses of individual neurons recorded one at a time also provided support for a redundancy reduction based on the structure of the neural receptive fields. At bright intensities, neural receptive fields in the retina have a center-surround organization (Srinivasan, Laughlin, & Dubs, 1982), and thus report the differences between the light intensities from the surrounding areas. This structure of the neural codes contributes to the decorrelation of neural responses (Balasubramanian & Sterling, 2009; Barlow, 1959; Borghuis, Ratliff, Smith, Sterling, & Balasubramanian, 2008; Garrigan et al., 2010). At low light intensities, the center surround organization becomes less pronounced (Srinivasan et al., 1982). This is also consistent with predictions of redundancy reduction because the increased noise under such conditions makes neural responses less correlated in the first place, reducing the need for additional correlation. Importantly, one can obtain a quantitative match between the structure of the neural receptive fields and predictions of maximally informative solutions at a variety of noise levels (Atick, 1992; Doi et al., 2012; Doi, Inui, Lee, Wachtler, & Sejnowski, 2003; Doi & Lewicki, 2005, 2007; Karklin & Simoncelli, 2011; Redlich & Atick, 1990).

Predictive Coding

Redundancy reduction can be applied temporally as well as spatially. In particular, thalamic neurons have been shown to have filtering properties that would lead to a decorrelation of stimuli with temporal correlations typical of natural scenes (Dong & Atick, 1995; Dan et al., 1996). More generally, one can apply ideas of the redundancy reduction principle to incorporate predictions made between stages of sensory processing, as was proposed by Rao and Ballard (1999). Predictions that are made by higher stages of neural processing are conveyed via feedback connections to lower stages, where they are subtracted from incoming signals. This approach has many desirable features and can account for a number of properties of cortical circuits. First, it is by construction hierarchical, and therefore can be replicated in an iterative fashion at different cortical stages. Second, the exchange between signals required to implement predictive coding matches almost completely the structure of connections within the cortical column (Bastos et al., 2012). Third, the model also explains why neurons in the superficial layers of cortex synchronize in the gamma range of frequencies, whereas neurons in the deep layer show synchronization in the lower alpha/beta frequencies. The reason is that in the predictive coding model neurons in the deep layers signal predictions to lower cortical areas, whereas neurons in the superficial signal prediction errors to subsequent cortical areas. In the model, predictions are obtained by a linear summation of prediction errors. Linear integration-type filtering serves as a low-pass filter

operation, explaining the tendency for lower frequencies in the deeper layers. Conversely, prediction errors are nonlinear functions of predictions. Because nonlinear operation can create and enhance signals are higher frequencies, the predictive coding model can explain synchronization in the higher frequency range compared to prediction signals carried by neurons in the deep layers.

Fourth, the predictive coding model reproduces many of the intriguing properties of extraclassical receptive fields, such as length tuning for hyper complex cells and the dynamics of neural responses to stimuli defined by differences in texture (Rao & Ballard, 1999). When applied to spatiotemporal natural stimuli, the hierarchical predictive coding model produced receptive fields that corresponded to optic flow patterns that match the properties of neurons in the dorsal visual pathway (Jehee, Rothkopf, Beck, & Ballard, 2006). It would be intriguing to see the predictions that this approach would generate for the responses of high-level sensory neurons in other modalities. Finally, it is noteworthy that the concept of redundancy reduction in time can be applied to the spike-generation process itself (Deneve, 2008; Jones, Johnson, & Ratnam, 2015). In this model, spikes signal the occurrence of new inputs that cannot be predicted based on the neuron's prior activity. The model can also be expanded to incorporate past activity from other neurons. Such signals would be transmitted via recurrent connections and can implement elements of the computations necessary for pattern completion (Roelfsema, 2006).

Neural Adaptation and Intermittent Structure of Natural Stimuli

Natural stimuli not only exhibit correlations on many time scales, but they are also *intermittent* (Figure 3). The term intermittency refers here to the observation that natural stimuli change between bouts of signals with different variances. A patch of signals with low variance will be followed by a patch of signals with large variance. While locally a bout of natural signals may be approximated as a Gaussian distribution across time, the sum of Gaussian distributions produces a non-Gaussian distribution with heavy tails (Ruderman & Bialek, 1994; Simoncelli & Olshausen, 2001). Such heavy-tailed distributions have been observed for the visual (Ruderman & Bialek, 1994; Simoncelli & Olshausen, 2001), auditory (Singh & Theunissen, 2003; Theunissen & Elie, 2014), and olfactory systems (Vickers, Christensen et al., 2001). The presence of correlations in time and space is also important here because without correlations, based on the central limit theorem, neurons will be receiving effectively Gaussian inputs, even if the initial distributions were non-Gaussian.
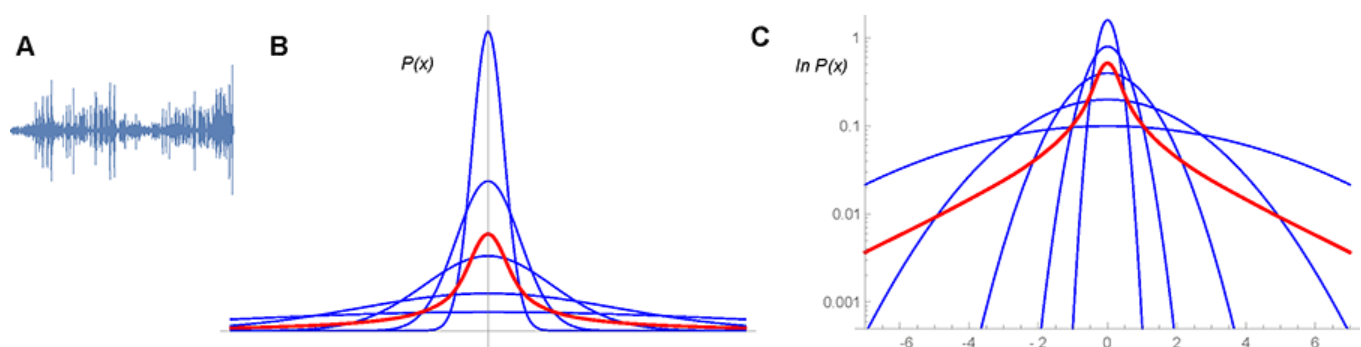
Figure 3. The intermittent structure of natural stimuli (A) leads to non-Gaussian statistics (B,C). Different patches of natural stimuli have different variances. As a result, even if the distribution at one moment in time or space can be described by a Gaussian function, the average distribution will be strongly non-Gaussian. This can be seen in the log-Plot (C) where the resultant distribution decreases linearly with the deviation from the mean, whereas Gaussian distributions correspond to parabolas in these coordinates.

Why is the intermittent structure of natural scenes important for neural coding? This is because the neural code can encode signals more efficiently by adjusting its properties to encode signals of a given variance (Brenner, Bialek, & de Ruyter van Steveninck, 2000; Fairhall, Lewen, Bialek, & de Ruyter van Steveninck, 2001; Kim & Rieke, 2001; Rieke, 2001). In this way the neural code can achieve a higher accuracy of representing sensory inputs for a given metabolic cost. For example, in vision, the mean light intensity varies broadly as a function of day and whether we are outside or inside a building. Correspondingly, retinal neurons adapt to a variety of mean light intensities (Enroth-Cugell & Shapley, 1973). The fact that retinal neurons can set their gain independently helps us appreciate the beauty of a sunset or sunrise. Such scenes are difficult to photograph because in a camera the gain is set equally across all pixels, whereas people can appreciate their beauty by adjusting to higher intensities around the sun while maintaining the ability to perceive lower light intensities in the periphery. Further, under typical conditions, the contrast varies significantly between different parts of the visual scene. Typically, sky regions have lower contrast (variance) between nearby positions then signals arising closer from the ground. Thus, retinal neurons that encode different parts of the visual field can set their gain independently to better encode their respective input signals. The adaptation is not limited to the mean and variance of input signals, but also takes place with respect to changes in spectrotemporal correlations (Kompaniez, Sawides, Marcos, & Webster, 2013; Sharpee et al., 2006; Simmons et al., 2013), color (Webster, Mizokami, Svec, & Elliott, 2006; Webster & Mollon, 1997), and facial features (Webster & MacLeod, 2011). Further, an adaptation to changes in variance has been observed in many sensory circuits studied, including the somatosensory system (Maravall, Petersen, Fairhall, Arabzadeh, & Diamond, 2007) and the olfactory system (Burgstaller & Tichy, 2012).

Timescales of Adaptation

The adaptive neural code makes it necessary to estimate parameters of the input distribution and to decide when it is worthwhile to switch to a new operating regime. It also raises further the question of how perception can remain stable when the neural responses are context-dependent (Fairhall et al., 2001). The question of stable perception has not been fully resolved at present, and there are certainly situations where perception is affected, as can be witnessed by the presence of certain illusions, such as the waterfall illusion. On the other hand, the decision as to when neural circuits should to switch to a different operating regime has been analyzed in considerable detail experimentally and theoretically. When considering this question, one may note that such questions are statistical in nature, and thus compare the timing of adaptive changes in the neural code with predictions of the optimal statistical

estimator. This situation that has been most thoroughly studied is what happens when input variance changes in step-like manner, such as during transitions from low to high variance and from high to low variance. One maybe surprised to learn that the transition from low to high and from high to low variance differ in the minimal time it takes to detect them: The transition to high variance is easier to detect because a single observation of a large input value is sufficient to provide strong evidence that the input variance is no longer small. In contrast, the evidence that the input variance has decreased is only available once a large number of small input values are received in sequence. After all, the observation of small values is also highly likely for distributions of large variance. The minimal time that it takes to detect changes in the variances of the input distribution were quantitatively predicted by (DeWeese & Zador, 1998) under a number of conditions, such as when there are sudden jumps in variance or while the variance was smoothly changing. Their theoretical results indicated that previously observed longer adaptation times to sudden decreases, compared to increases, in the mean light intensity (Barlow & Mollon, 1982) and contrast (Smirnakis, Berry, Warland, Bialek, & Meister, 1997) were not due to limitations of the biological machinery, as was previously assumed. Rather, they correspond to the requirements of the optimal statistical estimation strategy.

Studies of adaptive neuron dynamics have produced another startling phenomenon. Suppose that the input variance is switched between high and low values every five seconds or every 10 seconds. Provided that the neural dynamics have reached the steady state under both conditions, one might expect that from the fully adapted state to a new value would only depend on the new variance value and should not depend on how long the previous variance value was presented. Yet, experiments have indicated that this is not so (Fairhall et al., 2001). Furthermore, the time constant determining the initial dynamics after the switch scaled linearly with how long each value of the input variance was maintained in the experiment (Fairhall et al., 2001). What could be the cause of such dependence? It turns out that the differences in temporal dynamics could be explained as implementing an optimal statistical estimation strategy (Wark et al., 2009). The intuition behind this explanation is that the longer the circuit is exposed to a given fixed value of the input variance, the tighter the estimation accuracy of the variance value around this value is, and the smaller the probability that the value will switch to a new value. That the optimal timescale increases approximately linearly with the period of switching between different variance values follows from detailed computations and depends on the difference between variance values probed (Wark et al., 2009). While this discussion could create an impression that neural circuits could adapt to any kind of statistics, experimental studies did find apparent exceptions. One exception is the kurtosis or the fourth order statistics of the input distribution: Changes in kurtosis do not appear to trigger adaptation by themselves (Bonin, Mante, & Carandini, 2006; Tkačik, Ghosh, Schneidman, & Segev, 2014), although they do affect the dynamics of adaptation (Wark et al., 2009). For example, an adaptation to changes in the width of non-Gaussian (binary) distributions was faster than to changes in the variances of a Gaussian distribution.

Conclusion

In discussing possible principles of sensory coding, we have seen that information maximization can provide both a conceptual framework for understanding the mutual dependence between information and energy resources, and a number of specific predictions for the organization and dynamics of sensory circuits. So far, the bulk of supporting evidence comes from the early sensory processing where most of available information in the sensory signals has relevant/predictive value for future actions. The conceptual frameworks for understanding the optimality of sensory systems that are close to motor circuits and directly participate in the sensorimotor transformation have been formulated (Tishby & Polani, 2011) but only sparsely tested so far (Calhoun, Chalasani, & Sharpee, 2014; Sharpee, Calhoun, & Chalasani, 2014). Expanding these tests to a wider array of sensorimotor transformations is an interesting direction for future research, with potential broad impacts on engineered robotic systems. Another aspect of understanding behaviors from the information theory perspective is the optimality of reward systems and signals. In principle, reward signals are internal to the organism and therefore can be associated with random behaviors (as in addiction). Nevertheless, analyses of maximally efficient behavioral strategies suggest that those actions that lead to states with larger predictive values of the information should be rewarded more (Tishby & Polani, 2011). In other words, the internal reward function should mirror the predictive value of the information available in different sensory states.
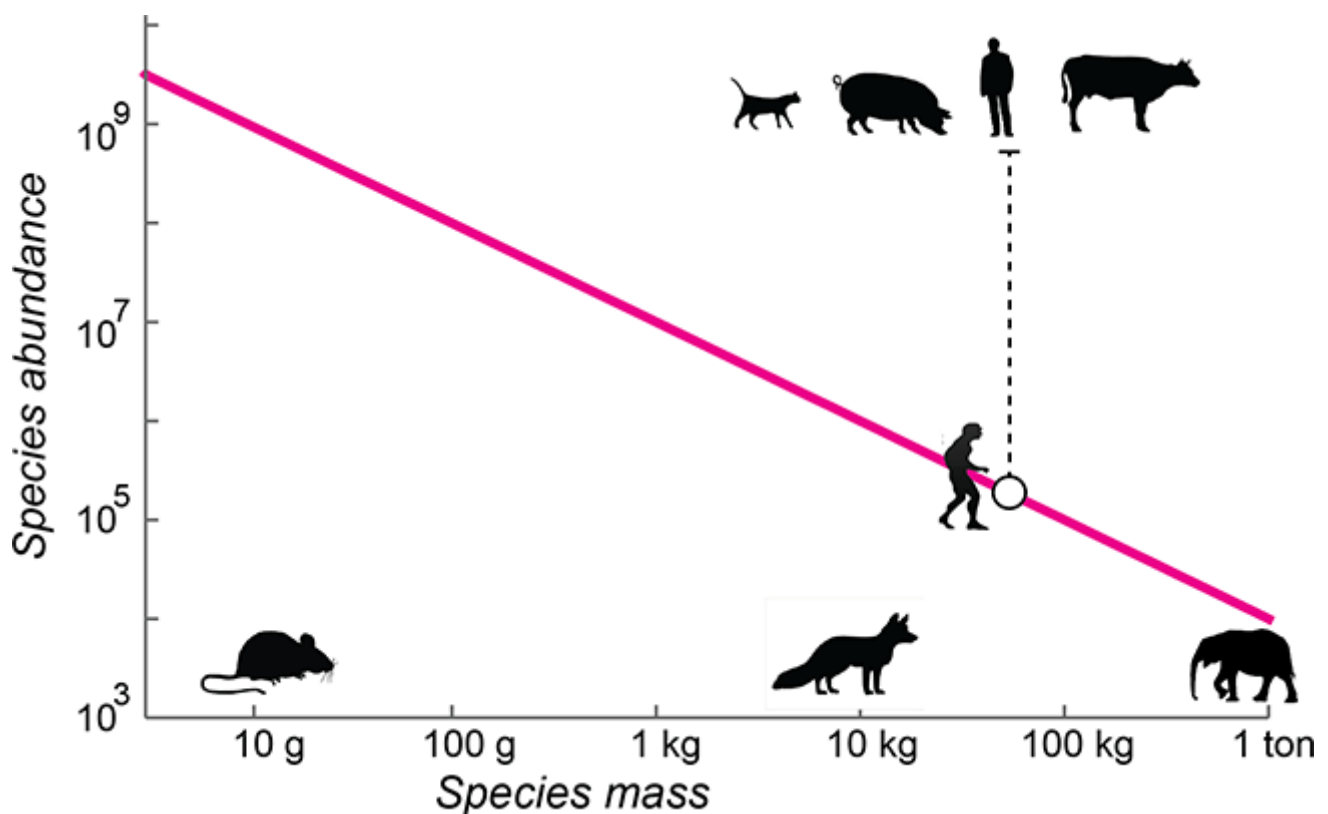
Figure 4. The scaling relationship between body size and species abundance. The power law shown here for mammals also extends to various ectotherms and bacteria (White et al., 2007). While the number of ancient humans agrees with this relationship, modern humans and domesticated animals deviate by five orders of magnitude from the predicted abundance for our mass (Kapitza, 1999). The increase and limitations are argued to be set by information exchange between humans rather than by environmental resources (Kapitza, 1999). The figure is created based on data from (Kapitza, 1999).

Finally, it is perhaps useful to contemplate the mutual constrains of the rates of information and resource acquisition at the level of human populations. Across many different species, from bacteria to large mammals, the species abundance follows an inverse scaling relationship with the body size (White et al., 2007). That is, there are fewer elephants than mice in this world. This scaling relationship correctly predicts the size of the prehistoric human population (10^5 at 1.6 million years ago). Modern humans exceed the predicted abundance for our body size by five orders of magnitude, together with various domesticated animals (Figure 4) (Kapitza, 1999). While some theories posit that population growth is limited by resources, other theories argue that the growth is determined by information exchange between individuals rather than by resources. The latter theory can account for smaller rates of growth among isolated human populations that have not been limited by resources per se (Henrich, 2015; Kapitza, 1999, 1999). For example, the population of Polar Inuit of northwest Greenland suffered a steady decline in population starting from 1820, when they became isolated by losing the technological know-how for making bows and kayaks, the latter being crucial for maintaining contacts with other groups. In 1862, following a chance visit from another Inuit group, the population started to increase again once the lost technologies were regained. (Boyd, Richerson, & Henrich, 2011). The information-exchange theory also can account for the growth of the global human population over the past three million years, specifically explaining the observed hyperbolic growth. So far the information-exchange theory in ecology primarily relies on the colloquial notion of information. However, it would be exciting to see if the initial attempts (Dolgonosov & Naidenov, 2006) to connect these ideas with Shannon's theory of information can be formalized to yield further insights into how we can make better use of the available natural resources.

Further Reading

Bialek W. (2013). *Biophysics: Searching for principles*. Princeton, NJ: Princeton University Press.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley-Interscience.

Henrich, J. P. (2015). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, U.K.: Cambridge University Press.

Kapitza, S. P. (1996). The phenomenological theory of world population growth. *Physics— Uspekhi*, *39*(1), 57–71.

Sterling, P., & Laughlin, S. (2015). Principles of neural design.

Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

References

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network*, *3*, 213–251.

Atick, J. J. (1993). Princeton Lectures on Biophysics. W. Bialek, World Scientific Publishing Company.

Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, *2*, 308–320.

Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, *4*, 196–210.

Balasubramanian, V., & Sterling P. (2009). Receptive fields and functional architecture in the retina. *Journal of Physiology*, *587*(12), 2753–2767.

Barlow, H., & Mollon, J. D. (1982). *The senses*.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *Proceedings of the symposium on the mechanization of thought processes*. London: HM Stationery Office.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fris, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, *37*(23), 3327–3338.

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Science of the U.S.A.*, *92*(9), 3844–3848.

Bialek, W. (2013). *Biophysics: Searching for principles*. Princeton, NJ: Princeton University Press.

Bialek, W., & de Ruyter van Steveninck, R. R. (2005). Features and dimensions: Motion estimation in fly vision. arxiv.org/abs/q-bio.NC/0505003 *<http://arxiv.org/abs/q-bio.NC/ 0505003>*: 1–18.

Bialek, W., I. Nemenman, & Tishby, N. (2001). "Predictability, complexity, and learning." *Neural Computation*, *13*(11), 2409–2463.

Bialek, W. & Owen, W. G. (1990). Temporal filtering in retinal bipolar cells: Elements of an optimal computation? *Biophys. J.*, *58*(5), 1227–1233.

Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *Journal of Neuroscience*, *26*(23), 6346–6353.

Borghuis, B. G., Ratliff, C. P., Smith, R. G., Sterling, P., & Balasubramanian, V. (2008). Design of a neuronal array. *Journal of Neuroscience*, *28*(12), 3178–3189.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Science of the U.S.A.*, *108*, Suppl. 2, 10918–10925.

Bray, D. (1995). Protein molecules as computational elements in living cells. *Nature*, *376*(6538), 307–312.

Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, *26*(3), 695–702.

Brunel, N., & Nadal, J. P. (1998). Mutual information, fisher information, and population coding. *Neural Computation*, *10*(7), 1731–1757.

Burgstaller, M., & Tichy, H. (2012). Adaptation as a mechanism for gain control in cockroach ON and OFF olfactory receptor neurons. *European Journal of Neuroscience*, *35*(4), 519–526.

Cadieu, C. F., & Olshausen, B. A. (2012). Learning intermediate-level representations of form and motion from natural movies. *Neural Computation*, *24*(4), 827–866.

Calhoun, A. J., Chalasani, S. H., & Sharpee, T. O. (2014). Maximally informative foraging by Caenorhabditis elegans. *ELife*, *3*.

Chichilnisky, E. J., & Kalmar, R. S. (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *Journal of Neuroscience*, *22*(7), 2737–2747.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley-Interscience.

Dan, Y., Atick, J. J., & Reid R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*, *16*(10), 3351–3362.

Deneve, S. (2008). Bayesian spiking neurons I: Inference. *Neural Computation*, *20*(1), 91–117.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*(8), 2051–2062.

DeWeese, M., & Zador, A. (1998). Asymmetric dynamics in optimal variance adaptation. *Neural Computation*, *10*, 1179–1202.

Doi, E., Gauthier, J. L., Field, G. D., Shlens, J., Sher, A., Greschner, M., . . . Simoncelli, E. P. (2012). Efficient coding of spatial information in the primate retina. *Journal of Neuroscience*, *32*(46), 16256–16264.

Doi, E., Inui, T., Lee, T. W., & Sejnowski, T. J. (2003). Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Computation*, *15*(2), 397–417.

Doi, E., & Lewicki, M. S. (2005). Sparce coding of natural images using an overcomplete set of limited capacity units. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems*, *17* (pp. 377–384). Cambridge, MA: MIT Press.

Doi, E., & Lewicki, M. S. (2007). A theory of retinal population coding. In B. Scholkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems*, *19* (pp. 353–360). Cambridge, MA: MIT Press.

Dolgonosov, B. M., & Naidenov, V. I. (2006). An informational framework for human population dynamics. *Ecological Modelling*, *198*, 375–386.

Dong, D. W., & Atick, J. J. (1995). Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network*, *6*, 159–178.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631.

Enroth-Cugell, C., & Shapley, R. M. (1973). Adaptation and dynamics of cat retinal ganglion cells. *Journal of Physiology*, *233*(2), 271–309.

Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature*, *412*(6849), 787–792.

Gal, S., Alpern, S., & Casas, J. (2015). Prey should hide more randomly when a predator attacks more persistently. *Journal of the Royal Society Interface*, *12*(113), 20150861.

Gal, S., & Casas, J. (2014). Succession of hide-seek and pursuit-evasion at heterogeneous locations. *Journal of the Royal Society Interface*, *11*(94), 20140062.

Garrigan, P., Ratliff, C. P., Klein, J. M., Sterling, P., Brainard, D. H., & Balasubramanian, V. (2010). Design of a trichromatic cone array. *PLOS Computational Biology*, *6*(2), e1000677.

Gjorgjieva, J., Sompolinsky, H., & Meister, M. (2014). Benefits of pathway splitting in sensory coding." *Journal of Neuroscience*, *34*(36), 12127–12144.

Gollisch, T., & Meister, M. (2010). Eye smarter than scientists believed: Neural computations in circuits of the retina. *Neuron*, *65*(2), 150–164.

Harper, N. S., & McAlpine, D. (2004). Optimal neural population coding of an auditory spatial cue. *Nature*, *430*(7000), 682–686.

Henrich, J. P. (2015). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.

Hilbert, M. (2017). The more you know, the more you can grow: An information theoretic approach to growth in the Information Age. *Entropy*, *19*(82), e19020082.

Isaacs, R. (1999). *Differential games: A mathematical theory with applications to warfare and pursuit, control and optimization*. Mineola, NY: Dover Publications.

Jehee, J. F., Rothkopf, C., Beck, J. M., & Ballard, D. H. (2006). Learning receptive fields using predictive feedback. *Journal of Physiology Paris*, *100*(1–3), 125–132.

Jones, D. L., Johnson, E. C., & Ratnam, R. (2015). A stimulus-dependent spike threshold is an optimal neural coder. *Frontiers in Computational Neuroscience*, *9*, 61.

Kapitza, S. P. (1996). The phenomenological theory of world population growth. *Physics—Uspekhi*, *39*(1), 57–71.

Kapitza, S. P. (1999). How many people have ever lived, are living, and will live on the Earth. Moscow: Nauka.

Karklin, Y., & Simoncelli, E. P. (2011). Efficient coding of natural images with a population of noisy Linear-Nonlinear neurons. *Advances in Neural Information Processing Systems*, *24*, Denver 1989, Curran Associates, Inc.

Kastner, D. B., Baccus, S. A., & Sharpee, T. O. (2015). Critical and maximally informative encoding between neural populations in the retina. *Proceedings of the National Academy of Science of the U.S.A.*, *112*(8), 2533–2538.

Kelly, J. L. (1956). A new interpretation of the information rate. *The Bell System Technical Journal*, *35*(4), 917–926.

Kim, K. J., & Rieke, F. (2001). Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *Journal of Neuroscience*, *21*(1), 287–299.

Kompaniez, E., Sawides, L., Marcos, S., & Webster, M. A. (2013). Adaptation to interocular differences in blur. *Journal of Vision*, *13*(6), 19.

Laughlin, S. (1983). Matching coding to scenes to enhance efficiency. In Braddick & Sleigh (Eds.), *Physical and biological processing of images* (pp. 42–51). Berlin: Springer Verlag.

Laughlin, S. B., de Ruyter van Steveninck, R. R., & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, *1*(1), 36–41.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1957). What the frog's eye tells the frog brain. *Proceedings of the IRE*, *47*(11), 1940–1951.

Maravall, M., Petersen, R. S., Fairhall, A. L., Arabzadeh, E., & Diamond, M. E. (2007). Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLOS Biology*, *5*(2), e19.

Martin, C. F., Bhui, R., Bossaerts, P., Matsuzawa, T., & Camerer, C. (2014). Chimpanzee choice rates in competitive games match equilibrium game theory predictions. *Scientific Reports*, *4*, 5182.

McDonnell, M. D., Stocks, N. G., Pearce, C. E. M., & Abbott, D. (2006). Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance. *Physics Letters*, *352*, 183–189.

McDonnell, M. D., Stocks, N. G., Pearce, C. E. M., & Abbott, D. (2005). Point singularities and suprathreshold stochastic resonance in optimal coding. Retrieved from https://arxiv.org/abs/cond-mat/0409528 *<https://arxiv.org/abs/cond-mat/0409528>*

Nikitin, A. P., Stocks, N. G., Morse, R. P., & McDonnell, M. D. (2009). Neural population coding is optimized by discrete tuning curves. *Physical Review Letters*, *103*(13), 138101.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Palmer, S. E., Marre, O., Berry, M. J., & Bialek, W. (2015). Predictive information in a sensory population. *Proceedings of the National Academy of Science of the U.S.A.*, *112*(22), 6908–6913.

Poling, A., Edwards, T. L., Weeden, M., & Foster, T. M. (2011). The matching law. *Psychological Record*, *61*(2), 313–322.

Potters, M., & Bialek, W. (1994). Statistical mechanics and visual signal processing. *Journal de Physique I*, *4*(11), 1755.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Ratliff, C. P., Borghuis, B. G., Kao, Y. H., Sterling, P., & Balasubramanian, V. (2010). Retina is structured to process an excess of darkness in natural scenes. *Proceedings of the National Academy of Science of the U.S.A.*, *107*(40), 17368–17373.

Redlich A. N., & Atick, J. J. (1990). Mathematical model of the simple cells in the visual cortex. *Biological Cybernetics*, *63*, 99–109.

Rieke, F. (2001). Temporal contrast adaptation in salamander bipolar cells. *Journal of Neuroscience*, *21*(23), 9445–9454.

Rieke, F., Bodnar, D. A., & Bialek, W. (1995). Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society B: Biological Sciences*, *262*(1365), 259–265.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*, 203–227.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters*, *73*, 814–817.

Sarpeshkar, R. (1998). Analog versus digital: extrapolating from electronics to neurobiology. *Neural Computation*, *10*(7), 1601–1638.

Sharpee, T. O. (2014). Toward functional classification of neuronal types. *Neuron*, *83*(6), 1329–1334.

Sharpee, T. O., Calhoun, A. J., Chalasani, S. H. (2014). Information theory of adaptation in neurons, behavior, and mood. *Current Opinion in Neurobiology*, *25*, 47–53.

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*(7079), 936–942.

Simmons, K. D., Prentice, J. S., Tkačik, G., Homann, J., Yee, H. K., Palmer, S. E., . . . Balasubramanian, V. (2013). Transformation of stimulus correlations by the retina. *PLOS Computational Biology*, *69*(12), e1003344.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *Journal of the Acoustical Society of America*, *114*(6 Pt 1), 3394–3411.

Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W, & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature*, *386*(6620), 69–73.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 978–982.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B*, *216*, 427–459.

Temizer, I., Donovan, J. C., Baier, H., & Semmelhack, J. L. (2015). A visual pathway for looming-evoked escape in larval zebrafish. *Current Biology*, *25*(14), 1823–1834.

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, *15*(6), 355–366.

Tishby, N., & Polani, D. (Eds.). (2011). *Information theory of decisions and actions: Perception-action cycle: Models, architectures, and hardware*. New York: Springer.

Tkačik, G., Ghosh, A., Schneidman, E., & Segev, R. (2014). Adaptation to changes in higher-order stimulus statistics in the salamander retina. *PLOS One*, *9*(1), e85841.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437.

Vickers, N. J., Christensen, T. A., Baker, T. C., & Hildebrand, J. G. (2001). Odour-plume dynamics influence the brain's olfactory code. *Nature*, *410*(6827), 466–470.

Viswanathan, G. M., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (2011). *The physics of foraging: An introduction to random searches and biological encounters*. Cambridge, U.K.: Cambridge University Press.

Wark, B., Fairhall, A., & Rieke, F. (2009). Timescales of inference in visual adaptation. *Neuron*, *61*(5), 750–761.

Webster, M. A., & MacLeod, D. I. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1702–1725.

Webster, M. A., Mizokami, Y., Svec, L. A., & Elliott, S. L. (2006). Neural adjustments to chromatic blur. *Spatial Vision*, *19*(2–4), 111–132.

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research*, *37*(23), 3283–3298.

White, E. P., Ernest, S. K., Kerkhoff, A. J., & Enquist, B. J. (2007). The relationship between size and abundance in ecology. *Trends in Ecology and Evolution*, *22*, 323–330.

Wolpert, D. M. (2011). The real reason for brains. TedGlobal talk. Retrieved from https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains?language=en *<https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains?language=en>*.

Zhao, X., Liu, M., & Cang, J. (2014). Visual cortex modulates the magnitude but not the selectivity of looming-evoked responses in the superior colliculus of awake mice. *Neuron*, *84*(1), 202–213.

Related Articles

Active Electroreception in Weakly Electric Fish

Auditory Hair Cells and Sensory Transduction

Deep Neural Networks in Computational Neuroscience

Evolution of Neocortex for Sensory Processing

Normalization Principles in Computational Neuroscience

Gastropod Feeding Systems: Evolution of Neural Circuits that Generate Diverse Behaviors

What Is a Neuronal Ensemble?