

Natural Images: Coding Efficiency

D J Graham and D J Field, Cornell University,
Ithaca, NY, USA

© 2009 Elsevier Ltd. All rights reserved.

Introduction

Images and movies of the natural world are known to share a variety of statistical regularities. Such stimuli show consistent spatial statistics, spatiotemporal statistics, contrast and intensity distributions, and chromatic structure. Consider the images shown in [Figure 1](#). The top row shows white noise and the bottom row shows natural scenes. A process that generates random white noise images will generate all possible images with equal probability. However, because of the statistical regularities of natural scenes, the possibility of seeing a natural scene generated from a white noise process is extremely low. For even a small 8×8 pixel patch, the entropy of the white noise patches with 8 bits of gray level is $8 \times 8 \times 8$, or 512, bits. This results in a total of 2^{512} , or about 10^{154} equally probable images. Natural scenes are considerably redundant (i.e., they have much lower entropy), with estimates that they contain approximately 40% of the entropy of white noise for small (8×8) patches, and still lower relative entropy for larger patches. This difference would in turn imply that a single white noise pattern out of 10^{90} would have the basic statistics of natural scenes (assuming a flat distribution). Although the precise ratio requires an estimate of the true distribution, the estimate does demonstrate just how redundant natural scenes are. Because of this redundancy, it is possible to build a visual system that is efficient if it is specifically dedicated to representing such an environment. A visual system can reduce the size of its problem space and focus its efforts on what it is likely to encounter in the world. Because of this redundancy, visual systems can be shaped by evolution and development to take advantage of the environment.

Using the supposition that the visual system employs codes that approach optimal efficiency with respect to predictable (i.e., redundant) structure, one can go some way in explaining why visual neurons show the coding properties they do. A number of studies have shown success at predicting many of the basic linear properties of visual neurons. In the following sections we will review some of this work. But we note that this work and any extension of it depend on the appropriate definition of efficiency. In the classical definition of efficiency used in engineering, a system is most efficient if all of its work (within limits imposed by

thermodynamics) is done in service of its task. In this view, efficiency presumes that the system has some well-defined tasks, and that the system's design is a direct reflection of the need to do as little work as possible to achieve those tasks. But how do we define these tasks for biological systems?

Efficient for What Task?

Many sensory systems, the task or 'goal' of which is often readily definable and measurable, have sensitivity that is near their physical limits. With respect to the brain as a whole, however, the only clearly defined task is the most general one in biology: differential reproductive success. But this goal tells us little about why a visual system would have the structure it does. Clearly, the visual system is involved in vision, but to treat it as an independent subsystem with some basic input/output relationship is too much of a simplification. Certainly, the output of the visual system cannot be reduced to a behavior. And although it may seem computationally reasonable, there is little reason to presume that the output of the visual system is simply some 'object-detector.' There are three reasons for this: (1) object recognition is not a well-localized processing task in cortex, (2) no known congenital disorder obliterates this ability exclusively, and (3) patients with visual agnosia caused by lesions typically show severe deficits in other visual faculties. Even though object recognition is distributed, one might still propose that object recognition could be the collective goal of the visual system's many parts. But object recognition surely competes for neural real estate with areas whose goals include accurate spatial mapping, motion predictions, and a host of other tasks.

This difficulty in defining the visual system's general goal is likely to prove a significant hurdle for studies attempting to apply efficient coding theories to later stages of processing. In addition, our current ability to predict a neuron's behavior is significantly limited at these later stages. Much of the success of efficient coding techniques has been in the application to neurons relatively early in the visual pathway (retina, lateral geniculate nucleus (LGN), and V1). For neurons in these areas, the family of response properties is relatively well defined and it is possible to talk about how the information in the visual image is represented with the array of neurons in each of these areas. This in turn allows us to address the question of whether the array of neurons is processing the information efficiently. However, the approach still requires a precise definition of efficiency.

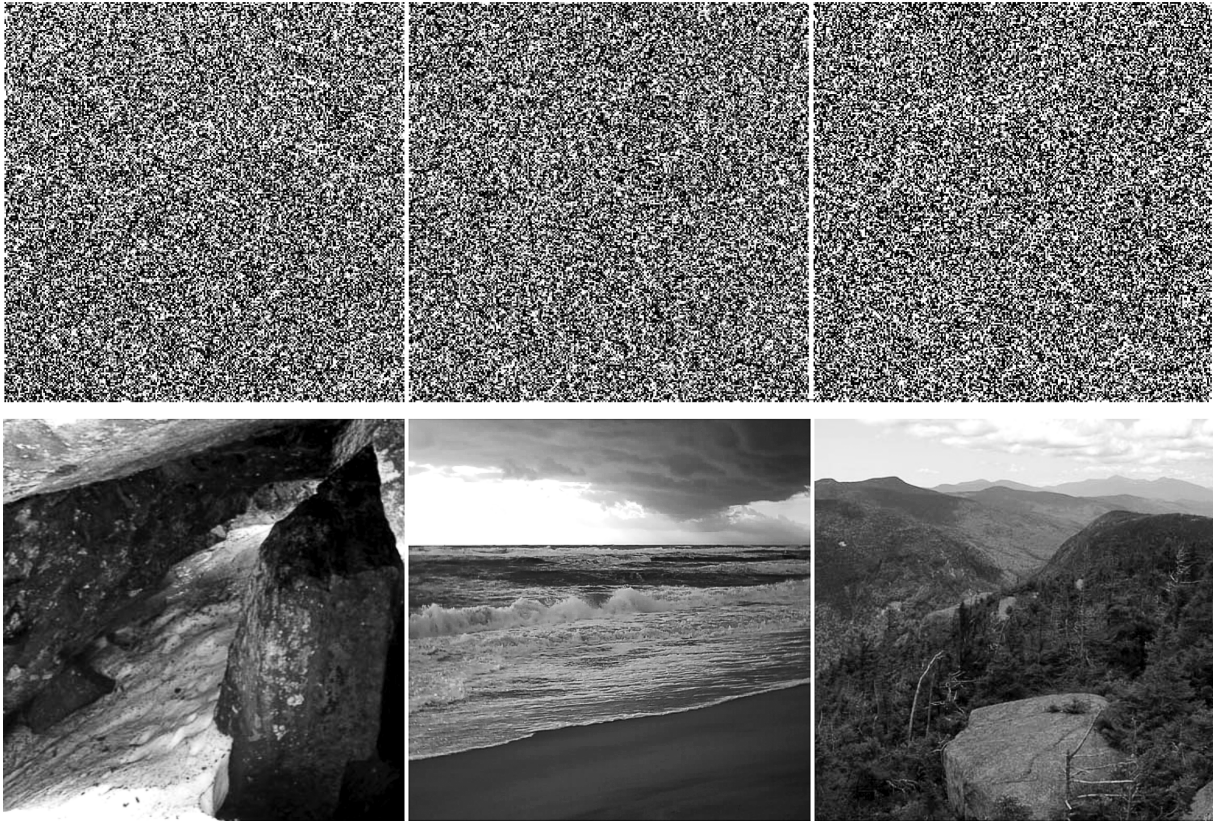


Figure 1 A process that chooses pixel intensity randomly will produce white noise images (top). Images resembling natural scenes (bottom), would almost never occur in such a process.

Defining Efficiency

For those who have proposed theories of optimality in early visual coding, much of the discussion has centered on the metric of efficiency. The majority of the work in the field focuses on what we will call representational efficiency. Such studies employ the tools of information theory, and they have explored the properties of neurons that are involved in representing the image. These studies have focused on issues relating to correlations and statistical independence in the firing rates of neurons. Much of this work has involved neural networks and computational models of visual areas. A second line of research has focused on what we call metabolic efficiency. Several influential reports have described investigations of the metabolic costs of generating spikes, and others have argued that constraints that minimize wiring are important for explaining known neuronal properties. In this article, we propose a third form of efficiency that we call learning efficiency. We argue that an important consideration for any sensory system is the challenge of learning about the relative probability of events in the world based on a handful of samples. To the extent that the visual system is

optimally efficient at carrying out its multitude of tasks, these efficiency rules are likely to each contribute significantly to a description of why the visual system is designed as it is. Visual systems may approach optimality (in the engineering sense of the word) across each of these three dimensions, but as we will argue, one of these dimensions – learning efficiency – is one that engineers rarely consider. These dimensions are not necessarily orthogonal to one another, nor do they currently all have well-defined units of measurement. These dimensions are mere sketches of the terrain over which the human visual system appears to have been optimized through evolution. Furthermore, we believe these dimensions can find application in other modalities as well, though here we focus on the visual system. Together, these efficiency dimensions may also suggest ways to design efficient artificial visual systems.

Representational Efficiency

Marcelja, in 1980, was the first to propose that neurons found in the primary visual cortex (V1) show a number of similarities to the mathematical functions

described by Gabor's 1946 theory of communication. Later studies (by Field and Tolhurst, in 1986; and Jones and Palmer, in 1987) confirmed that Marcelja's Gaussian-modulated sinusoid model (i.e., the Gabor function model) provided a good first-order approximation to the receptive field properties of these neurons. Some of the first computational models of the early visual system (e.g., by Watson, in 1983; and Daugman, in 1985) demonstrated how an array of neurons with these properties can represent a visual image. However, this work left unresolved the important question of 'why' such a solution might evolve.

One approach suggests that an understanding of the 'why' question requires consideration of the environment in which the system functions. This approach suggests that theories should be guided by an understanding of the statistics of natural scenes. Early television researchers, and later, in 1987, Field and Burton and Moorhead, found that the Fourier spatial frequency spectra of natural scenes typically fall off as $1/f^k$, where f is spatial frequency and k is approximately 1.2. This regularity expresses the same redundancy as the autocorrelation function of scenes, which measures how similar neighboring points are in terms of luminance. The $1/f$ structure follows from two properties: (1) neighboring points are correlated and (2) the images are roughly scale invariant. A number of studies have shown that there are a wide variety of forms of redundancy found in natural scenes that go beyond the pairwise correlations and spatial frequency spectra (see later). However, pairwise correlations have been the source of a number of theories regarding efficient coding and many researchers imply that these are the most relevant statistics, so we begin with a discussion of such correlations.

Correlation and Decorrelation

When the notion of stimulus redundancy is considered, most analyses typically refer to the pairwise correlations in the data. Many of the earliest theories of efficient sensory coding developed from the notion that if the neural responses to two stimuli are correlated, then an efficient system should strive to represent the data with reduced correlation. This is certainly one important form of redundancy. In recent work, Chandler and Field argue that the pairwise correlations (as represented by the power spectra) account for approximately 40% of the total redundancy in natural scenes (for 8×8 natural scene patches).

In theory, any representation with significant correlations implies that most of the signal lies in a subspace within the larger space of possible representations. By choosing a representation that codes for

only that subspace, it is possible to represent the data with significantly reduced dimensionality (e.g., a reduced number of neurons, or a smaller dynamic range of responses). Srinivasan, Laughlin, and Dubs showed that the center-surround structure of fly large monopolar cells is well matched to the correlations in a collection of scenes they considered relevant to the fly. Given a certain level of noise, they found that a weighted, linear sum over space transmits the greatest amount of information about their collection of natural inputs. Atick and Redlich continued this line of inquiry, arguing that the amplitude spectra of natural scenes are effectively flattened once they reach retinal outputs. The argument is that the roughly linear rise in sensitivity with increasing spatial frequency is inversely related to the $1/f$ -distributed fall-off in natural scene spectra, thus allowing neighboring neurons to be uncorrelated (i.e., have a flat spectrum). Atick and Redlich argued that this decorrelation is reflected in a flattening of the spatial frequency response of retinal ganglion cells. However, this argument is dependent on appropriate spacing of retinal neurons. Moreover, the current evidence suggests that neighboring neurons are significantly correlated in the presence of natural scenes.

We have recently argued that an independent goal of retinal coding is to achieve compression while maintaining equal response magnitude across the array of neurons of different sizes. This approach achieves a form of response spectrum flattening but it is quite independent of whether the neurons are tuned and spaced ideally to achieve decorrelation. This work suggests that, in order to explain the details of neural response properties, we must acknowledge that decorrelation in the retina and LGN is far from complete and that efficient coding schemes must consider additional constraints.

Optimal Information Transfer

Another strategy attempts to utilize each level of response magnitude with equal frequency. To this end, Laughlin compared the distribution of local contrasts in the blowfly environment with contrast responses in large monopolar cells in the creature's eye. The intensity-response function measured for these cells is well matched to a distribution of natural contrasts, such that the range of possible responses approximates the information-theoretic ideal. In other words, contrast responses in the fly are distributed across the range of environmental contrasts in such a way as to minimize the number of 'code words' necessary for transmitting information about the entire range of contrasts at a given fidelity. Later studies have shown that adaptation of visual neurons with respect to local luminance,

orientation, and contrast can also be addressed using this maximum information transfer (infomax) approach. A more elaborate extension of infomax models combines feedback of stored predictors from higher cortical areas with typical natural inputs. These schemes optimize over families of representations that minimize error between the input and a top-down representation. Feedback representation optimization models of this sort effectively produce a system whereby primary visual areas reduce their response as higher areas provide better descriptions of input 'content,' in line with recent imaging findings. It should be noted that in many infomax models, optimal redundancy reduction (decorrelation) and maximum information transfer strategies are equivalent.

Beyond Correlations: Sparseness and Independence

The pairwise correlations found in natural scenes represent only one form of redundancy. **Figure 2** shows two images – a natural scene and noise – with similar $1/f$ structure and therefore similar pairwise correlations. There are a number of ways to describe the differences between these two images. For example, they differ in their phase spectra. But the two images can also be described in terms of differences in their sparse structure. For the $1/f$ noise image, all linear representations produce response distributions that are Gaussian. However, for the natural scene, some projections of the data are non-Gaussian. That is, when the appropriate array of linear filters is used to represent a natural scene, the histogram of activity will be a non-Gaussian histogram.

As noted by Field, a non-Gaussian histogram implies low entropy in the first-order responses and relatively high entropy in the higher order relationships between the filters for a linear transform (i.e., more independent). In other words, a system that produces maximally non-Gaussian histograms produces a representation wherein the neurons are

maximally independent. This is the basic idea behind sparse coding algorithms and independent components analysis (ICA).

In a code with maximal independence, the firing of each neuron provides maximal unique information (i.e., the sharing of information with other neurons has been minimized). If the data consist of an array of relatively rare, sparse events, then matching the neurons to those events will produce activity that is sparse. The definitions of sparseness have varied in the literature. In general, 'sparse' implies a relatively high probability of no activity across the population, and some proportion of relatively active neurons. In the computational literature, in which neurons are often modeled as linear operators, the kurtosis (the fourth statistical moment) of the response histogram can be used to describe relative sparseness. Other metrics, such as the sparseness index, have proved more useful for spike trains.

Field demonstrated that arrays of linear neurons with properties like those found in primary visual cortex appear to maximize the sparse response to natural scenes. Olshausen and Field further demonstrated that a neural network that attempts to represent natural scenes and maximize sparseness will produce an array of neurons with spatial properties like those found in cortical simple cells. That is, a system that is forced to produce a faithful representation of the input using only a handful of neurons (each firing near its maximum response when it is active) gives simple-cell-like receptive fields. This suggests that at least at the level of primary visual cortex, visual system representations show evidence of being efficiently matched to the sparseness of natural scenes. Similar results have been found for spatiochromatic stimuli and spatiotemporal patterns.

It must be emphasized that sparse outputs of these networks result from the sparse structure of the data. It would be relatively simple to produce a



Figure 2 Noise with a spatial frequency amplitude spectrum like that of natural scenes (left) has the same pairwise correlations as does the natural scene (right), but lacks other statistical regularities of scenes.

nonlinearity that forces a sparse output independent of the input, one that would not be an efficient coding strategy. The networks described earlier expressly search for the sparse structure that exists in the data. Moreover, recordings of primate visual neurons in response to natural scenes have also been shown to produce a sparse output. Visual neurons show sparse responses both in early stages of cortical processing, such as V1, and in higher levels, such as inferotemporal cortex. Highly sparse firing is a widespread phenomenon in a variety of brain areas and species, including monkey association areas, rabbit motor areas, rat somatosensory areas, rat auditory areas, rat hippocampus, and centers believed to be involved in bird song generation.

However, the sparse response of these neurons is not direct evidence that these neurons are efficiently representing the environment. All of these neurons are nonlinear, and higher level neurons are very nonlinear. A proof that these codes are efficient would require a clear understanding of how the array of neurons represents the input image. We are not at this level of understanding – and we may never be. Most visual neurons beyond the retina can be modulated by higher levels of processing. Although a number of studies have implied that a decent theory of primary visual cortex is close at hand, Olshausen and Field have argued that we are still a long way from such a theory. Part of the argument comes from the work of the Gallant lab. This work has been an effort to measure the responses of primate visual neurons to a set of natural stimuli and to predict from these responses how each neuron would respond to an arbitrary natural scene. Gallant and colleagues have found that even in V1, typically less than 50% of the response variance to new stimuli can be predicted. In higher levels of the visual system, that prediction accuracy is further reduced. In addition, as noted by Olshausen and Field in 2005, recording studies select only a small portion of the neurons in any given area, and researchers typically record only from large pyramidal neurons that produce large spikes. Other neuron types (e.g., granular neurons that produce smaller spikes) are often not surveyed in these studies, which has led to the interesting suggestion by Shoham and colleagues that there exists a neural ‘dark matter’ problem.

Optimality with Nonlinear Systems

Despite the difficulties in providing a clear metric for defining efficiency in nonlinear systems, a number of efforts have been made to generalize efficiency arguments to account for the known nonlinear properties of visual neurons. One particularly fruitful approach

involves a technique called slow feature analysis. According to this model, a stimulus moving across an image results in quick changes to the neurons responding to the input (for a linear array of neurons). If we consider the histogram of neural activity integrated over some time period, we would find that the activity is less sparse than in a single time frame. However, the changes in images are not random and are often caused by objects or backgrounds showing consistent movement. Slow feature analysis attempts to take into account this redundancy in the movements of features. The technique attempts to find nonlinear solutions that are capable of describing the moving images with a relatively consistent set of neurons, despite the changes that are occurring. Slow feature analysis has been demonstrated to produce nonlinear behavior like that shown with complex cells. It has been argued that current implementations of slow feature analysis are mathematically equivalent to a form of spatiotemporal ICA. Given that object identities are relatively invariant over time, there is hope that these lines of work will eventually be capable of producing object-level representations like those found in inferotemporal cortex. However, we have not yet reached that point.

There exist a variety of techniques that come under the general heading of nonlinear ICA. Although this area of research is too vast to be reviewed here, it should be noted that many of these studies have generated some of the nonlinear properties of visual neurons. However, since we do not yet have a complete model of the nonlinearities in these neurons, we cannot yet argue that such techniques are capable of accounting for the full array of neural properties in any region of visual cortex.

Without a clear account of how information is represented in any given area, we are left without any kind of proof that the information is represented efficiently. Linear models that optimize sparseness and independence do produce simulated neurons with many properties like those found in V1. This certainly supports the notion that V1 is directed toward an efficient code for natural scenes. But as we will argue in the next sections, there are both metabolic and learning costs for high sparseness and therefore a more complete understanding of efficient coding requires examining factors beyond those considered in proposals of representational efficiency. Both the design of artificial visual systems and the development of algorithms for representing natural scenes will likely benefit from our understanding of a larger range of evolutionary constraints, like those described in the following sections. These additional biological constraints require a broadening of current notions of efficient design beyond representational

efficiency. We may someday find that the best artificial systems look remarkably like systems we find in biological systems, and it is likely that a full account of biological systems will require an understanding of these biological requirements.

Metabolic Efficiency

Consideration of representational efficiency as discussed in the preceding sections leads to important insights into the design of visual systems. It provides a number of metrics of efficiency that are generally independent of the limitations of energy or of neural hardware. In this section, we consider attempts to explain the properties of neurons from a consideration of metabolic constraints. We discuss two broad hypotheses of efficient neural design: spike efficiency and neural wiring optimization.

Spike Efficiency

Any information-processing strategy in the nervous system will incur metabolic costs. If we accept that information is primarily transmitted through the use of spikes, then a relevant question is whether one of the constraints on information processing is the number of spikes. Two studies have made detailed efforts to estimate the metabolic cost of spikes, and both have come to the conclusion that the high cost of spikes indeed results in an important constraint on neural processing. Both Attwell and Laughlin, in 2001, and Lennie, in 2003, found that the total available metabolic resources and the cost of a spike limit the firing rate to less than 1 Hz – and probably less than 0.2 Hz. As subsequently noted by Olshausen and Field, most studies have found firing rates significantly higher, suggesting that the cells that are typically recorded have unusually high firing rates. Levy and Baxter earlier argued that when the cost of spiking is considered, maximum information transfer is attained when only 2–16% of the neurons are firing. Lennie estimated that the limited resources imply that at any given moment, only 1/50th of the population of cortical neurons will show high firing rates. Thus, in order to save energy individually and across populations, visual neurons must adopt highly sparse patterns of firing. Therefore, we can argue that from both a representational point of view and a metabolic point of view, sparse firing is efficient.

A related debate centers on the optimal histogram of activity for a real neuron that can only produce a positive firing rate. If the goal of a neuron is to maximize the total information rate while minimizing the mean activity, the most efficient distribution is the exponential distribution. There is certainly a similarity

between the histograms of activity of visual neurons (in response to natural signals) and the exponential distribution. However, on close inspection, the exponential model does not appear to fit. For example, Treves et al. found that for neurons in inferotemporal cortex, the exponential model could be rejected in 84% of cases. The lack of any particular successful model led these authors to conclude that there was “no special optimization principle or purpose to the firing distributions found.”

Moreover, we question whether the goal should be to optimize the information rate. Consider a case in which there are n possible causes or features in an image and there exist n neurons available to represent those features. One could argue from a representational viewpoint that one should match the neurons to the features. If each feature had a particular response probability, then matching that feature would produce a histogram of activity that matched the histogram of the feature, not one that necessarily matched the histogram corresponding to the optimal compression algorithm. Certainly we do not expect that there is a match between the number of features and the number of neurons. But this example is meant to show that if we expect neurons to provide an explicit representation of their environment, part of the impetus for that explicit representation may be to match the probability distributions of the environment – not simply to maximize information rates.

Minimum Wiring

In addition to metabolic constraints, any neural architecture will be dependent on the anatomical ‘wiring’ available. This so-called wiring optimization principle dates back to Ramón y Cajal. This type of metabolic efficiency is typically measured by comparing the wiring volume and distribution of real neurons with calculations of the minimum volume needed to connect model cells and cortical areas to one another. Mitchison has argued that stripe patterns in cortex (e.g., ocular dominance columns) are efficient at minimizing the volume of dendritic wiring needed by areas receiving input from two separate sources, compared to alternative arrangements. Durbin and Mitchison found that a highly reduced model of cortical wiring arrangement matched real wiring in certain ways and minimized a model of wiring cost. Along the same lines, Koulakov and Chklovskii found that pinwheel patterns of orientation preference in cortex allow efficient wiring. In a related study by Chklovskii et al., neural wiring density and local signal delay have similarly reached an optimal compromise. Together, these studies suggest that intracortical wiring optimization can help explain why

cortical maps have the topography they do. Cherniak et al. and others have shown that minimal wiring may also help explain why brain areas are placed and interconnected as they are. In an early study using a reduced anatomical model of the *Caenorhabditis elegans* ganglia, Cherniak employed a connection-optimizing algorithm to search for connectivities that minimized total wiring. The actual wiring of the ganglia matched the optimal arrangement from a family of possible orderings.

Some have argued that brain evolution and development impose invariant connectivity ratios, which may tend to enforce optimal wiring. For example, Changizi and Shimojo suggest that across phyla the average number of synapses per neuron scales with the number of neurons per physiologically defined area, and that the number of connections among areas (per area) scales with the number of areas. However, this approach has several limitations. It should be noted that the number of ways that an efficient wiring system could strive to optimize the cost of adding cell volume, increasing cell metabolism, delaying and attenuating the signal, and making projections during development is essentially unlimited. But when considered in concert with representational constraints, minimum wiring arguments can provide insights about efficient receptive field design.

Learning Efficiency

In addition to the fact that neural systems, unlike human-engineered ones, are difficult to assess in terms of their goals and costs, they also differ from traditionally engineered systems because they are produced through development and learning rather than through carefully planned assembly. Indeed, part of the reason why the level of efficiency of a brain is hard to measure is that it becomes specialized to a variety of tasks – often concurrently – during development, the combination of which helps the adult creature reproduce. Therefore, in order to explain why the visual system of the adult has the structure it does, given a proposal of efficient processing of natural scenes, we must consider how visual system representations may depend on efficient learning and development. This efficiency dimension in particular is an example of somewhere our intuition about the brain seeking desperately to conserve energy can lead us astray. Indeed, learning efficiency could be at odds with metabolic cost optimizations, since the proportion of metabolic energy consumption in infant human brains (which are most subject to this constraint) is roughly 3 times that for adults.

In discussions of efficiency in sensory systems, many studies have used various types of neural networks and

other learning algorithms to generate early sensory processes. There are certain basic problems that any learning algorithm must overcome. We examine one of these issues – invariance. Then we describe evidence of contributions to learning efficiency from innate, purely learned, and hybrid strategies, which span the well-known nature/nurture debate.

Sparseness and Invariance

Certain constraints are imposed on a system that must learn about its environment from a relatively small number of examples. In essence, efficient learning is dependent on finding a balance between selectivity of neurons for specific features and invariance across examples having features that vary in irrelevant ways (e.g., lighting). This task is further complicated because each example is likely to be seen just once in a lifetime. A second instance is unlikely to be presented with the same lighting, the same position, the same size, and the same orientation as the first. To allow any calculation of the probability of what is likely, it is critical that the system be capable of generalizing to multiple instances of that object.

Sparse coding, in the strict version, does not help with this problem. The typical argument against hierarchical sparse codes is that a system that develops a neuron for every object would require too many object detectors in order to function. This ‘grandmother cell’ hypothesis has been critiqued elsewhere, but it is worth noting that plausible models of object recognition suggest the need for grandmother cell-like coding for certain tasks (e.g., distinguishing between different faces) and sparse population codes for other tasks (e.g., categorization). That is, object recognition may require a variety of strategies that vary in their degree of sparseness.

It has also been argued that optimizing for sparseness and independence can assist a system in identifying novel inputs and in detecting new relationships. Consider the case in which objects are uniquely represented by just one neuron versus the case in which the object is identified by the relative activity of 100 neurons. In the single-neuron case, to learn that two objects often co-occur requires a relatively simple algorithm (e.g., Hebbian learning). But when identification requires a particular activity profile among many neurons, the learning algorithm would require a far more complex association among the 100 units involved in the representation.

Overcompleteness

An important property of visual representations beyond the optic nerve is that they are highly overcomplete. Visual cortex in humans contains on the order of 1000 times more neurons than do the two optic

nerves. In macaque V1 alone, there are approximately 50 times more output fibers than input fibers. These overcomplete codes must involve significant redundancy. Overcompleteness has been suggested by Barlow as an optimally efficient way “to model the redundancy in images, not necessarily to reduce it.” And as Riesenhuber and Poggio pointed out, overcompleteness allows any particular signal to be represented with a higher degree of sparseness than is possible with complete codes, a property which is useful for generalization during learning, as described earlier. In order to achieve efficient learning, one must consider solutions that may not be optimally efficient from a strict representational efficiency point of view. That is, solutions that are representationally efficient (e.g., maximum information transfer arguments) may not help explain how the system achieves efficient learning.

‘Hard-Coded’ Efficiency

Though every part of the visual system undergoes developmental change, some efficient properties of the adult cells are thought to be innate. Maloney found that the distribution of natural chromaticities could be efficiently and accurately coded using just three wavelength vectors, and that the three cone spectral sensitivities in particular may optimally achieve color constancy based on typical surface spectral reflectances in the natural world. Others have proposed that the distribution of spectral sensitivities in primates is optimally weighted so as to separate nutritious fruit from surrounding foliage, a genetic adaptation that could have aided propagation of trichromatic primates.

Moreover, at the level of the retina, there is a deep conservation of ‘virtually all functional and structural features of importance’ and of the developmental program in the vertebrate retina according to Finlay et al. This may indicate that visual system development is optimized to allow successful elaboration, robustness, and specialization (what is referred to as ‘evolvability’) for a range of species that live in a variety of habitats using a single developmental and organizational scheme. It can therefore be seen as an efficiency which is innate, but which is connected to the types of variations present in a changing visual environment.

The development of classical receptive field organization in cortex does not appear to require learning, and relatively little spatial refinement is needed to achieve adult level acuity. One area of debate centers on the question of the necessity of visual experience in order to achieve adult receptive field structures, which presumably contribute to efficient representations in the brain. Recent studies have suggested the prevailing view – which proposes that animals dark-reared

during the critical period of development are left with unrefined receptive fields because visual experience had no chance to ‘prune’ dendritic arbors – could be in need of reevaluation.

Efficient Learning from the Environment

In higher levels of cortex, learning leads to a marked decrease in mean response in prefrontal and inferotemporal cortex, which has been linked to the detection of novel stimuli. In particular, learning based on natural stimulus matching in monkeys led to systematically lower mean responses to learned natural stimuli in prefrontal cortex, and these responses did not vary with the addition of noise. At lower levels of cortex, learning has a minor effect on classical response properties such as orientation tuning and receptive field size in V1 cells). Learning also has a relatively small effect on responses in V4 to noise-degraded natural stimuli. Therefore, efficient learning strategies may be an important principle, especially in higher levels of visual cortex.

Hybrid Strategies: Efficient Innate Learning

Finally, there is evidence that spontaneous activity in the visual system, in a time period before a creature’s eyes open, may help refine the organization of the visual system. There is good evidence that suppressing spontaneous activity in the retina (so-called retinal waves) can affect the refinement of retinal projections to the LGN. As some have noted, spontaneous, patterned activity is observed in cortex, hippocampus, thalamus, retina, and spinal cord in a variety of creatures. Simple programs of the sort that could theoretically produce spontaneous, patterned activity could well be coded into genes, and the ‘running’ of these programs could produce the necessary statistical properties that a developing visual system needs to operate at a basic level once the eyes open. We refer to this as notion as ‘innate learning,’ since it employs elements of both learning and innateness. These innate learning strategies are efficient in the sense that in the absence of external stimuli, they require far fewer genetic instructions in order to develop proper response properties, compared to full genetic specificity.

However, innate learning in the form of spontaneous retinal activity is not required for the initial formation of ocular dominance columns or of orientation selectivity. There is strong evidence that retinotopic maps and ocular dominance patterns fail to refine properly in animals reared in altered visual environments (e.g., strobe environments, binocular deprivation, environments with altered spatial statistics). But it should be noted, especially considering the brief time during which spontaneous activity has been so far studied,

as stated by Carrasco and colleagues, that “the extent to which spontaneous and visually driven activity contribute to the development and maintenance of stimulus specificity is unclear.”

In short, learning efficiency may require solutions that would not be predicted by representational efficiency arguments or metabolic efficiency arguments alone. The difficulty of learning invariant properties of objects based on a handful of presentations is one reason why this is so. The degree to which the visual system is efficiently engineered to learn is dependent on innate properties, on learned associations, and on patterns of spontaneous activity, the combination of which may approach ‘optimal’ efficiency in concert with metabolic and representational constraints.

Conclusion

Visual systems have been argued to be operating near theoretical limits of optimality – not simply parsimoniously – for certain tasks. But given the profusion of ways that brain structure, cell development and function, and neural representation are believed to be optimal in some way, we argue that visual system design represents a compromise among these many demands for efficiency. We have attempted to elucidate a variety of ways in which visual systems can strive for efficiency. No single quantity (e.g., energy consumption) applies strict limits to brain design (except in the absolute sense), nor is the system optimized to perform a single, ‘optimizable’ task. There is much to be learned from investigating the contribution from a variety of constraints on efficiency.

See also: Activity in Visual Development; Animal Models of Inherited Retinal Degenerations; Fovea: Primate; Information Coding; Retina: An Overview; Retinal Pharmacology; Inner Retinal Layers; Retinal Color Mechanisms; Retinal Development: An Overview; Retinal Development: Cell Type Specification; Retinal Models; Visual Cortical Models of Orientation Tuning.

Further Reading

- Atick JJ and Redlich AN (1992) What does the retina know about natural scenes? *Neural Computation* 4: 196–210.
- Attwell D and Laughlin SB (2001) An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism* 21: 1133–1145.
- Barlow HB (2001) Redundancy reduction revisited. *Network: Computer Neural Systems* 12: 241–253.
- Carrasco MM, Razak KA, and Pallas SL (2005) Visual experience is necessary for maintenance but not development of refined retinotopic maps in superior colliculus. *Journal of Neurophysiology* 94: 1962–1970.

- Chandler DM and Field DJ (2007) Estimates of the information content and dimensionality of natural scenes from proximity distributions. *Journal of the Optical Society of America* 24: 922–941.
- Chklovskii DB and Koulakov AA (2004) Maps in the brain: What can we learn from them? *Annual Review of Neuroscience* 27: 369–392.
- Daugman JG (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America* 2: 1160–1169.
- David SV, Vinje WE, and Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of V1 neurons. *Journal of Neuroscience* 24: 6991–7006.
- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A* 4: 2379–2394.
- Field DJ (1993) Scale-invariance and self-similar ‘wavelet’ transforms: An analysis of natural scenes and mammalian visual systems. In: Farge M, Hunt JCR, and Vassilicos JC (eds.) *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*, pp. 151–193. Oxford: Oxford University Press.
- Field DJ (1994) What is the goal of sensory coding? *Neural Computation* 6: 559–601.
- Finlay BL, Silveira LCL, and Reichenbach A (2005) Comparative aspects of visual system development. In: Kremers J (ed.) *The Structure, Function and Evolution of the Primate Visual System*, pp. 37–72. New York: Wiley.
- Graham DJ, Chandler DM, and Field DJ (2006) Can the theory of ‘whitening’ explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research* 46: 2901–2913.
- Jones JP and Palmer LA (1987) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* 58: 1233–1258.
- Laughlin SB, de Ruyter van Steveninck RR, and Anderson JC (1998) The metabolic cost of neural information. *Nature Neuroscience* 1: 36–41.
- Lennie P (2003) The cost of cortical computation. *Currents in Biology* 13: 493–497.
- Maloney LT and Wandell BA (1986) Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A* 3: 29–33.
- Mitchison G (1991) Neuronal branching patterns and the economy of cortical wiring. *Proceedings of the Royal Society of London – Series B* 245: 151–158.
- Olshausen BA and Field DJ (1996) Emergence of simple cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607–609.
- Olshausen BA and Field DJ (2005) How close are we to understanding V1? *Neural computation* 17: 1665–1699.
- Osorio D and Vorobyev M (1996) Colour vision as an adaptation to frugivory in primates. *Proceedings of the Royal Society of London – Series B* 263: 593–599.
- Shoham S, O’Connor DH, and Segev R (2006) How silent is the brain: Is there a ‘dark matter’ problem in neuroscience? *Journal of Comparative Physiology A* 192: 777–784.
- Srinivasan MV, Laughlin SB, and Dubs A (1982) Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London – Series B: Biological Sciences* 216: 427–459.
- Treves A, Panzeri S, Rolls ET, et al. (1999) Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Computation* 11: 601–631.