Principles of Neural Information Theory

A Tutorial Introduction

James V Stone

Title: Principles of Neural Information Theory
A Tutorial Introduction
Author: James V Stone

First Edition, 2017.
Typeset in LaTeX 2_ε.
Cover design: Stefan Brazzo.
First printing.
File: book_NeuralInfoTheory_v24.tex, .

ISBN 978-0-9933679-2-2

For Teleri.

Contents

Preface

> *To understand life, one has to understand not just the flow of energy, but also the flow of information.*
> W Bialek, 2012.

Here in the 21st century, where we rely on computers for almost every aspect of our daily lives, it seems obvious that information is important. However, it would have been impossible for us to know just how important it is before Claude Shannon almost single-handedly created *information theory* in the 1940s. Since that time, it has become increasingly apparent that information, and the energy cost of each bit of information, imposes fundamental, unbreachable limits on the form and function of all organisms. In this book, we concentrate on one particular function, information processing in the brain. In our explorations, we will discover that information theory dictates exactly how much information can be processed by each neuron, and how the staggeringly high cost of that information forces the brain to treat information like biological gold dust. Almost all of the facts presented in this book reflect the harsh realities implied by the application of information theory to neuronal computation, and the predictions of one particular idea, known as the *efficient coding hypothesis.*

The methods we use to explore the efficient coding hypothesis lie in the realms of mathematical modelling. Mathematical models demand a precision unattainable with purely verbal accounts of brain function. With this precision, comes an equally precise quantitative predictive power. In contrast, the predictions of purely verbal models can be vague, and this vagueness also makes them virtually indestructible, because predictive failures can often be explained away. No such luxury exists for mathematical models. In this respect, mathematical models are easy to test, and if they are weak models then they are easy to disprove. So, in the Darwinian world of mathematical modelling, survivors tend to be few, but those few tend to be supremely fit.

Of course, this is not to suggest that purely verbal models are always inferior. Such models are a necessary first step in understanding. But continually refining a verbal model into ever more rarefied forms cannot be said to represent scientific progress. Eventually, a purely verbal model should evolve to the point where its predictions can be tested against measurable physical quantities. Happily, most branches of neuroscience reached this state of scientific maturity some time ago. Accordingly, this book is intended as a tutorial account of how one particular mathematical framework (information theory) is being used to test the quantitative predictions of a candidate general principle of brain function: the efficient coding hypothesis.

Feynman's Legacy. Every writer of scientific texts aspires to acquire the deceptively easy style of the great physicist Richard Feynman. In his famous lecture series (http://feynmanlectures.caltech.edu/), he defined what it means to write simply, and without jargon, whilst providing the reader with a rigorous and intuitive understanding of physics. However, Feynman's style was borne of deep insights, based on many years of study. This, in turn, engendered a confidence which allowed him to un-grasp the mathematical hand-holds, which re-assure, but also constrain, other scientists. Inspired by such eloquent writing, the style adopted here in *Principles of Neural Information Theory* is an attempt to describe the raw science of neural information theory, un-fettered by the conventions of standard textbooks, which can confuse rather than enlighten the novice. Accordingly, key concepts are introduced informally, before being described mathematically; and each equation is accompanied by explanatory text.

So, unlike most textbooks, and like the best lectures, this book is intended to be both informal and rigorous, with prominent sign-posts as to where the main insights are to be found, and many warnings about where they are not. Using this approach, it is hoped that the diligent reader may gain an intuitive understanding of key facts, which are sometimes well presented, but often well camouflaged, in more formal accounts of neural computation and information theory.

What Is Not Included. An introductory text cannot cover all aspects of a subject in detail, and choosing what to leave out is as important as choosing what to include. In order to compensate for this necessity, pointers to material not included, or not covered in detail, can be found in the annotated Further Reading section.

PowerPoint Slides of Figures. Most of the figures used in this book can be downloaded from
http://jim-stone.staff.shef.ac.uk/BookNeuralInfo/NeuralInfoFigs.html

Corrections. Please email corrections to j.v.stone@sheffield.ac.uk.
A list of corrections can be found at
http://jim-stone.staff.shef.ac.uk/BookNeuralInfo/Corrections.html

Acknowledgments. Thanks to all those involved in developing the freely available LaTeX 2ε software, which was used to typeset this book. Shashank Vatedka deserves a special mention for checking the mathematics in a final draft of this book. Thanks to Caroline Orr for meticulous copy-editing and proofreading. For reading draft versions of this book, I am very grateful to Karl Friston, Nikki Hunkin, Stuart Wilson and ...

Jim Stone, Sheffield, England, 2017.

Chapter 1

All That We See

When we see, we are not interpreting the pattern of light intensity that falls on our retina; we are interpreting the pattern of spikes that the million cells of our optic nerve send to the brain.
Rieke, Warland, De Ruyter van Steveninck, and Bialek, 1997.

1.1. Introduction

All that we see begins with an image focussed on the *retina* at the back of the eye (Figure 1.1). Initially, this image is recorded by 126 million photoreceptors within the retina. The outputs of these photoreceptors are then *encoded*, via a series of intermediate connections, into a sequence of digital pulses or *spikes*, that travel through the one million nerve fibres of the *optic nerve* which connect the eye to the brain.

The fact that we see so well implies that the brain must be extraordinarily good at encoding the retinal image into spikes, and equally good at *decoding* those spikes into all that we see (Figure 1.2). But the brain is not only good at translating the world into spikes, and spikes into perception, it is also good at transmitting information from the eye to the brain whilst expending as little energy as possible. Precisely how good, is the subject of this book.

1.2. Efficient Coding

Neurons communicate information, and that is pretty much all that they do. But neurons are expensive to make, maintain, and run[55]. For example, half of the total energy used by a child at rest is required just to keep the brain ticking over. Of this, about 13% is used to transmit spikes along neurons, and the rest is for maintenance. The cost of using neurons is so high that only 2-4% of them can be active at any one time[57].

Given that neurons and spikes are so expensive, we should be unsurprised to find that when the visual data from the eye is encoded as a series of spikes, each neuron and each spike conveys as much information as possible. These considerations have given rise to the *efficient coding hypothesis*[5;10;12;28;75;97;98], an idea developed over many years by Horace Barlow (1959)[9].

The efficient coding hypothesis is conventionally interpreted to mean that neurons re-package sensory data in order to transmit as much information as possible. Even though it is not usually made explicit, if data are encoded efficiently as described above then this often implies that the amount of energy paid for information is as small as possible. In order to avoid any confusion, we adopt a more specific interpretation of the efficient coding hypothesis here: namely, that neurons re-package sensory data in order to transmit as much information as possible *per Joule of energy expended*[47;49;63;67;68;85;96].

There are a number of different methods which collectively fall under the umbrella term 'efficient coding'. However, to a first approximation, the results of applying these various methods tend

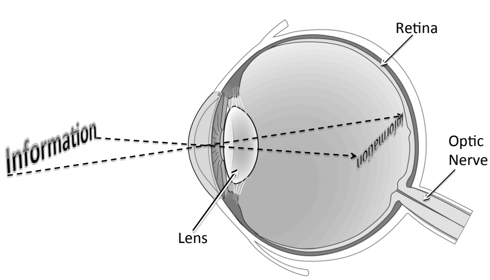

Figure 1.1. Cross section of eye.

to be quite similar[75], even though the methods themselves appear quite different. These methods include sparse coding[35], principal component analysis, independent component analysis[11;86], information maximisation (infomax)[58], predictive coding[72;84] and redundancy reduction[3]. We will encounter most of these broadly similar methods throughout this book, but we place special emphasis on predictive coding because it is based on a single principle, and it has a wide range of applicability.

1.3. General Principles

The test of a theory is not just whether or not it accounts for a body of data, but also how complex the theory is in relation to the complexity of the data being explained. Clearly, if a theory is, in some sense, more convoluted than the phenomenon it explains then it is not much of a theory. As an extreme example, if each of the 86 billion neurons in the brain required its own unique theory then the resultant collective theory of brain function would be almost as complex as the brain itself. This is why we favour theories that explain a vast range of phenomena with the minimum of words or equations. A prime example of such a parsimonious theory is Newton's theory of gravitation, which explains (amongst other things) how a ball falls to Earth, how atmospheric pressure varies with height above the Earth, and how the Earth orbits

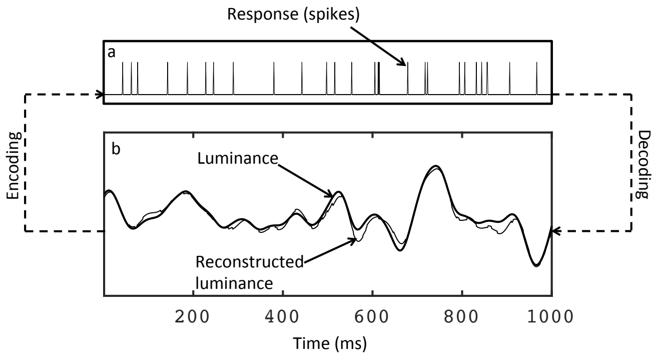

Figure 1.2. Encoding and decoding. Rapidly changing luminance (bold curve in b) is *encoded* as a neuronal *spike train* (a), which can be *decoded* to reconstruct an estimate of the luminance (thin curve in b).

the Sun. In essence, we favour theories which rely on a *general principle* to explain a range of physical phenomena.

With this in mind, there are a finite number of general principles which may explain the design of the brain. Briefly, and within the context of physical theories, some prime candidates for a general principle are: 1) the supply of energy is the single most important factor in the design of the brain, 2) information throughput is the single most important factor in the design of the brain, and, 3) information per Joule of energy expended is the single most important factor in the design of the brain (i.e. the efficient coding hypothesis). However, even though complex systems are affected by many factors, usually only one of them dominates its behaviour[13] (see Section 6.1).

Whichever theory is correct, if we want to understand how the brain works then we need more than a theory which is expressed in mere words. For example, if the theory of gravitation were stated only in words then we could say that each planet has an approximately circular orbit, but we would have to use many words to prove precisely why each orbit must be elliptical, and to state exactly how elliptical each orbit is. In contrast, a few equations would express these facts exactly, and without ambiguity. Thus, whereas words are required to provide theoretical context, mathematics imposes a degree of precision which is extremely difficult, if not impossible, to achieve with words alone. To quote one of the first great scientists,

> *The universe is written in this grand book, which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language in which it is written. It is written in the language of mathematics, without which it is humanly impossible to understand a single word of it.*
> Galileo Galilei, 1623.

In the spirit of Galileo's recommendation, a rigorous theory of information processing in the brain should begin with a quantitative definition of information.

1.4. Information Theory

Information theory was developed almost exclusively by Claude Shannon during the 1940s. His classic paper published in 1948, and the subsequent book by Shannon and Weaver (1949)[81], heralded a transformation in our understanding of information. Before the publication of Shannon's work, information had been regarded as a kind of poorly defined miasmic fluid. But afterwards, it became apparent that information is a well-defined and, above all, *measurable* quantity.

Shannon considered information to be as fundamental as physical quantities like energy and mass (see the quotation which opens Chapter 2). Even though we cannot sense information in the same way that we can sense the effects of energy (e.g. as heat) or mass (e.g. as weight), information is just as important for life, for us, and for our brains.

Shannon's theory of information provides a mathematical definition of information, and describes precisely how much information can be communicated between different elements of a system. This may not sound like much, but Shannon's theory underpins our understanding of how signals and noise are related, and why there are definite limits to the rate at which information can be communicated within *any* system, whether man-made or biological.

1.5. Neurons, Signals and Noise

When a question is typed into a computer search engine, the results provide useful information, but this is buried in a sea of mostly useless data. In this internet age, it is easy for us to appreciate the difference between information and mere data, and we have learned to treat the information as useful *signal* and the rest as useless *noise*. This experience is now so commonplace that phrases like *signal to noise ratio* are becoming part of everyday language. Even though most people are unaware of the precise meaning of this phrase, they know intuitively that data comprise a combination of signal and noise.

The ability to separate signal from noise, to extract information from data, is crucial for modern telecommunications. For example, it allows a television picture to be compressed or encoded to its bare information bones and transmitted to a satellite, and then to a TV, before being decoded to reveal the original picture on the TV screen.

More importantly, this type of scenario is ubiquitous in the natural world. The ability of eyes and ears to extract useful signals from noisy sensory data, and to package those signals efficiently, is the key to survival[88]. Indeed, the efficient coding hypothesis suggests that the evolution of sense organs, and of the brains that process data from those organs, is primarily driven by the need to minimise the energy expended for each bit of information acquired from the environment. Moreover, because information theory tells us how to measure information precisely, it provides an objective benchmark against which the performance of neurons can be compared.

The maximum rate at which information can be transmitted through a neuron can be increased in a number of different ways. However, whichever way we (or evolution) chooses to do this, doubling the maximum information rate costs more than a doubling in neuronal hardware, and more than twice the amount of power (energy per second)[85]. This is a universal phenomenon, which implies a diminishing information return on every additional micrometre of neuron diameter, and on every additional Joule of energy invested in transmitting spikes along a neuron. This, in turn, imposes fundamental and unbreachable limits on information processing in neuronal systems.

The extraordinarily high cost of information means that the brain cannot depend on physiological mechanisms which require extravagant amounts of information. Whereas an astronomer can quadruple the amount of light in an image by quadrupling the area of his telescope's objective lens, any nocturnal animal which attempted the same trick would pay in myriad ways, and would therefore almost certainly reduce its Darwinian fitness. Far better, far more *efficient*, to extract as much information as possible from a relatively dim retinal image, and to re-package it to its informational essence before sending it to the brain.

Information theory does not place any conditions on what type of mechanism implements this re-packaging; in other words, on exactly *how* it is to be achieved. However, unless there are unlimited amounts of power available, relatively little information will reach the brain without some form of re-packaging. In other words, information theory does not specify how any task, such as vision, is implemented, but it does set fundamental limits on what is achievable by any implementation, biological or otherwise.

Because these limits are unbreachable, and because they effectively extort such a high price, there seems to be little alternative but to evolve brains which are exquisitely sensitive to the many trade-offs between time, neuronal hardware, energy and information. As we shall see, whenever such a trade-off is encountered, the brain seems to maximise the amount of information gained for each Joule of energy expended.

1.6. An Overview of Chapters

This section contains technical terms which are explained fully in the appropriate chapter, and in the Glossary.

In order to fully appreciate the evidence referred to above, some familiarity with the basic elements of information theory is required; these elements are presented in Chapter 2. We then consider (in Chapter 3) how to apply information theory to the problem of measuring the amount of information in the output of a spiking neuron, and how much of this information (i.e. mutual information) is related to changes in the neuron's input. We also consider how often a neuron should produce a spike in order to maximise its information content, and we find that this coincides with an important property, linear decodability. In Chapter 4, we discover that one of the consequences of information theory (specifically, Shannon's noisy coding theorem) is that the cost of information rises inexorably and disproportionately with information rate. This steep rise suggests that neurons should set particular physical parameters like axon diameter, the distribution of axon diameters, and synaptic conductance to minimise the cost of information; evidence is presented which supports this suggestion.

In Chapter 5, we consider how the correlations between the inputs to neurons sensitive to different colours always reduce information rates, and how this can be ameliorated by pre-processing in the retina to decorrelate outputs. This pre-processing involves principal component analysis, which can be used to maximise neuronal information throughput. The lessons learned so far are then applied (in Chapter 6) to the problem of encoding time-varying, correlated visual inputs. We explore how a standard neuron model can be used for efficient coding of the temporal structure of retinal images, and how a predictive coding model yields similar results to the standard model. In Chapter 7, we explore how the spatial structure of the retinal image can be encoded, and how information theory predicts different encoding strategies under high and low luminance conditions. Evidence is presented that these strategies are consistent with those used in the retina, and which are also implemented by predictive coding.

Once colour, spatial or temporal structure has been encoded by a neuron, the result must pass through the neuron's non-linear input/output (transfer) function. Accordingly, in Chapter 8, we consider what form this transfer function should adopt in theory, in order to maximise information throughput. Crucially, we find that this theoretically optimal transfer function matches those found in visual neurons. Finally, the problem of how to decode neuronal outputs is addressed in Chapter 9, where the importance of prior knowledge or experience is explored in the context of Bayes' theorem. In each chapter, we will explore particular neuronal mechanisms, how they work, and (most importantly) *why* they work in the way they do.

Chapter 2

Information Theory

A basic idea in information theory is that information can be treated very much like a physical quantity, such as mass or energy.
C Shannon, 1985.

2.1. Introduction

Every physical quantity, like a sound or a light, consists of data which has the potential to provide information about some aspect of the world. For an owl, the sound of a mouse rustling a leaf may indicate a meal is below; for the mouse, a flickering shadow overhead may indicate it is about to become a meal.

Precisely how much information is gained by a receiver from data depends on three things. First, and self-evidently, the amount of information in the data. Second, the relative amounts of relevant information or *signal*, and irrelevant information or *noise*, in the data. Third, the ability of the receiver to separate the signal from the noise.

Once the data reach the sensory apparatus of an animal, it is up to that animal to ensure that the information in the data is preserved so that it reaches the animal's brain. The limits on an animal's ability to capture data from the environment, to package them efficiently, and to extract the information they contain, is dictated by a few fundamental *theorems*, which represent the foundations on which information theory is built (a theorem is a mathematical statement which has been proved to be true). The theorems of information theory are so important that they deserve to be regarded as the *laws* information.

Just as a bird cannot fly without obeying the laws of physics, so, a brain cannot function without obeying the laws of information. And, just as the shape of a bird's wing is ultimately determined by the laws of physics, so the structure of a neuron is ultimately determined by the laws information. In order to understand how these laws are related to neural computation, it is necessary to have a sound grasp of the essential facts of Shannon's theory of information.

Being both a mathematician and an engineer, Shannon stripped the problem of communication to its bare essentials, depicted in Figure 2.1. He then provided the fundamental theorems of information theory, which can be summarised as follows. For any communication channel: 1) there is a definite upper limit, the *channel capacity*, to the amount of information that can be communicated through that channel, 2) this limit shrinks as the amount of noise in the channel increases, 3) this limit can very nearly be reached by judicious packaging, or encoding, of data before it is sent through the channel. For our purposes, an important corollary of these theorems is that the cost of information rises very rapidly as the information rate increases.

Note that this chapter can be skipped on a first reading of the book, and returned to as necessary.

2.2. Finding a Route, Bit by Bit

Information is usually measured in *bits*, and one bit of information allows you to choose between two equally probable alternatives. In order to understand why this is so, imagine you are standing at the

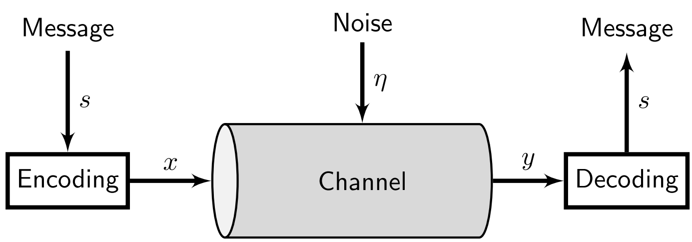

Figure 2.1. The communication channel. A message (data) is encoded before being used as input to a communication channel, which adds noise. The channel output is decoded by a receiver to recover the message.

References

[1] EH Adelson and JR Bergen. Spatiotemporal energy models for the perception of motion. *JOSA A*, 2(2):284–299, 1985.

[2] JJ Atick, Z Li, and AN Redlich. Understanding retinal color coding from first principles. *Neural Computation*, 4(4):559–572, 1992.

[3] JJ Atick and AN Redlich. Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320, 1990.

[4] JJ Atick and AN Redlich. What does the retina know about natural scenes? *Neural Computation*, 4(2):196–210, 1992.

[5] F Attneave. Some informational aspects of visual perception. *Psychological Review*, pages 183–193, 1954.

[6] D Attwell and SB Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, 21(10):1133–1145, 2001.

[7] V Balasubramanian, D Kimber, and MJ Berry. Metabolically efficient information processing. *Neur Comp*, 13(4):799–815, 2001.

[8] V Balasubramanian and P Sterling. Receptive fields and functional architecture in the retina. *J Physiology*, 587(12):2753–2767, 2009.

[9] HB Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. In DV Blake and AM Utlley, editors, *Proc Symp Mechanization Thought Processes (Vol2)*. 1959.

[10] HB Barlow. Possible principles underlying the transformations of sensory messages. *Sensory communication*, pages 217–234, 1961.

[11] AJ Bell and TJ Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[12] W Bialek. *Biophysics: Searching for Principles.* Princeton University Press, 2012.

[13] W Bialek. Perspectives on theory at the interface of physics and biology. *arXiv:1512.08954*, 2015.

[14] W Bialek, M DeWeese, F Rieke, and D Warland. Bits and brains: Information flow in the nervous system. *Physica A: Statistical Mechanics and its Applications*, 200(1-4):581–593, 1993.

[15] R Bogacz. A tutorial on the free-energy framework for modelling perception and learning. *J. of Mathematical Psychology*, 2015.

[16] BG Borghuis, CP Ratliff, RG Smith, P Sterling, and V Balasubramanian. Design of a neuronal array. *J Neuroscience*, 28(12):3178–3189, 2008.

[17] DH Brainard, P Longere, PB Delahunt, WT Freeman, JM Kraft, and B Xiao. Bayesian model of human color constancy. *Journal of Vision*, 6(11), 2006.

[18] G Buchsbaum and Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc Roy Soc London, B*, 220(1218):89–113, 1983.

[19] J Burge and WS Geisler. Optimal disparity estimation in natural stereo images. *Journal of Vision*, 14(2):1, 2014.

[20] J Bussgang. Cross-correlation function of amplitude-distorted Gaussian signals. *Lab. Elect., Mass. Inst. Technol., Cambridge, MA, USA, Tech. Rep*, 216, 1952.

[21] P Dagum and M Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial intelligence*, 60(1):141–153, 1993.

[22] Y Dan, JJ Atick, and RC Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J of Neuroscience*, 16(10):3351–3362, 1996.

[23] P Dayan and DC Abbott. *Theoretical Neuroscience.* MIT Press, New York, NY, USA, 2001.

[24] MH DeGroot. *Probability and Statistics, 2nd edition.* UK, Addison-Wesley, 1986.

[25] MR Deweese. *Optimization principles for the neural code.* PhD thesis, Princeton University, 1995.

[26] MR DeWeese and W Bialek. Information flow in sensory neurons. *Il Nuovo Cimento D*, 17(7-8):733–741, 1995.

[27] E Doi, JL Gauthier, GD Field, J Shlens, A Sher, M Greschner, TA Machado, LH Jepson, K Mathieson, DE Gunning, et al. Efficient coding of spatial information in the primate retina. *The Journal of Neuroscience*, 32(46):16256–16264, 2012.

[28] E Doi and MS Lewicki. A simple model of optimal population coding for sensory systems. *PLoS Comp Biol*, 10(8).

[29] Dawei W Dong and Joseph J Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358, 1995.

[30] K Doya, S Ishii, A Pouget, and R Rao. *The Bayesian Brain.* MIT, MA, 2007.

[31] TE Duncan. Evaluation of likelihood functions. *Information and Control*, 13(1):62–74, 1968.

[32] P Dyan and LF Abbott. *Theoretical Neuroscience.* MIT Press, 2001.

[33] LA Finelli, S Haney, M Bazhenov, M Stopfer, and TJ Sejnowski. Synaptic learning rules and sparse coding in a model sensory system. *PLoS Comp Biol*, 2008.

[34] BJ Fischer and JL Pena. Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience*, 14:1061–1066, 2011.

[35] P Foldiak and M Young. Sparse coding in the primate cortex. *in The Handbook of Brain Theory and Neural Networks, ed. Arbib, MA*, pages 895–898, 1995.

[36] JP Frisby and JV Stone. *Seeing: The computational approach to biological vision.* MIT Press, 2010.

[37] P Garrigan, CP Ratliff, JM Klein, P Sterling, DH Brainard, and V Balasubramanian. Design of a trichromatic cone array. *PLoS Comput Biol*, 6(2):e1000677, 2010.

[38] JL Gauthier, GD Field, A Sher, M Greschner, J Shlens, AM Litke, and EJ Chichilnisky. Receptive fields in primate retina are coordinated to sample visual space more uniformly. *Public Library of Science (PLoS) Biology*, 7(4), 2009.

[39] AR Girshick, MS Landy, and EP Simoncelli. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, 14(7):926–932, 2011.

[40] D Guo, S Shamai, and S Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005.

[41] JJ Harris, R Jolivet, E Engl, and D Attwell. Energy-efficient information transfer by visual pathway synapses. *Current Biology*, 25(24):3151–3160, 2015.

[42] HK Hartline. Visual receptors and retinal interaction. *Nobel lecture*, pages 242–259, 1967.

[43] A Jarvstad, SK Rushton, PA Warren, and U Hahn. Knowing when to move on cognitive and perceptual decisions in time. *Psychological science*, 23(6):589–597, 2012.

[44] DC Knill. Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, 7:1–24, 2007.

[45] DC Knill and R Richards. *Perception as Bayesian inference.* Cambridge University Press, New York, NY, USA, 1996.

[46] DC Knill and JA Saunders. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43:2539–2558, 2003.

[47] K Koch, J McLean, M Berry, P Sterling, V Balasubramanian, and MA Freed. Efficiency of information transmission by retinal ganglion cells. *Current Biology*, 14(17):1523 – 1530, 2004.

[48] K Koch, J McLean, R Segev, MA Freed, MJ Berry, V Balasubramanian, and P Sterling. How much the eye tells the brain. *Current biology : CB*, 16(14):1428–1434, 07 2006.

[49] L Kostal, P Lansky, and MD McDonnell. Metabolic cost of neuronal information in an empirical stimulus-response model. *Biological cybernetics*, 107(3):355–365, 2013.

[50] L Kostal, P Lansky, and J-P Rospars. Efficient olfactory coding in the pheromone receptor neuron of a moth. *PLoS Computational Biology*, 4(4), 2008.

[51] SW Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16:3768, 1953.

[52] MF Land and DE Nilsson. *Animal eyes*. OUP, 2002.

[53] SB Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch*, 36c:910–912, 1981.

[54] SB Laughlin. Matching coding to scenes to enhance efficiency. In O.J. Braddick and A.C. Sleigh, editors, *Physical and biological processing of images*, pages 42–52. Springer, Berlin, 1983.

[55] SB Laughlin, RR de Ruyter van Steveninck, and JC Anderson. The metabolic cost of neural information. *Nature neuroscience*, 1(1):36–41, 1998.

[56] DC Lay. *Linear Algebra and its Applications*. New York, Addison-Wesley, 1997.

[57] P Lennie. The cost of cortical computation. *Current Biology*, 13:493–497, 2003.

[58] R Linsker. Self-organization in perceptual network. *Computer*, pages 105–117, 1988.

[59] M Meister and MJ Berry. The neural code of the retina. *Neuron*, 22:435–450, 1999.

[60] I Nemenman, GD Lewen, W Bialek, and RR de Ruyter van Steveninck. Neural coding of natural stimuli: Information at sub-millisecond resolution. *PLoS Comp Biology*, 4(3), 03 2008.

[61] S Nirenberg, SM Carcieri, AL Jacobs, and PE Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411(6838):698–701, June 2001.

[62] JE Niven, JC Anderson, and SB Laughlin. Fly photoreceptors demonstrate energy-information trade-offs in neural coding. *PLoS Biology*, 5(4), 03 2007.

[63] JE Niven and SB Laughlin. Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol*, 211:1792–804, 2008.

[64] H. Nyquist. Certain topics in telegraph transmission theory. *Proceedings of the IEEE*, 90(2):280–305, 1928.

[65] BA Olshausen. 20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked. In *20 Years of Computational Neuroscience*, pages 243–270. 2013.

[66] S Pajevic and PJ Basser. An optimum principle predicts the distribution of axon diameters in *PLoS ONE*, 2013.

[67] JA Perge, K Koch, R Miller, P Sterling, and V Balasubramanian. How the optic nerve allocates space, energy capacity, and information. *J Neuroscience*, 29(24):7917–7928, 2009.

[68] JA Perge, JE Niven, E Mugnaini, V Balasubramanian, and P Sterling. Why do axons differ in caliber? *J Neuroscience*, 32(2):626–638, 2012.

[69] E Persi, D Hansel, L Nowak, P Barone, and C van Vreeswijk. Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex. *PLoS Comput Biol*, 7(2):e1001078, 02 2011.

[70] JR Pierce. *An introduction to information theory: symbols, signals and noise.* Dover (2nd Edition), 1961 reprinted by Dover 1980.

[71] MD Plumbley. Efficient information transfer and anti-hebbian neural networks. *Neural Networks*, 6:823–833, 1993.

[72] RPN Rao and DH Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2:79 – 87, 1999.

[73] FM Reza. *Information Theory.* New York, McGraw-Hill, 1961.

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

[75] F Rieke, D Warland, RR van Steveninck, and W Bialek. *Spikes: Exploring the Neural Code.* MIT Press, Cambridge, MA, 1997.

[76] KF Riley, MP Hobson, and SJ Bence. *Mathematical methods for physics and engineering.* Cambridge University Press, 2006.

[77] TW Ruderman DL, Cronin and C Chiao. Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America*, 15:2036–2045, 1998.

[78] AB Saul and AL Humphrey. Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *Journal of neurophysiology*, 64(1):206–224, 1990.

[79] B Sengupta, MB Stemmler, and KJ Friston. Information and efficiency in the nervous system - a synthesis. *PLoS Computuational Biology*, July 2013.

[80] CE Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[81] CE Shannon and W Weaver. *The Mathematical Theory of Communication*. University of Illinois Press.

[82] EC Smith and MS Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 2006.

[83] WW Sprague, EA Cooper, I Tošić, and MS Banks. Stereopsis is adaptive for the natural environment. *Science Advances*, 1(4):e1400254, 05 2015.

[84] MV Srinivasan, SB Laughlin, and A Dubs. Predictive coding: A fresh view of inhibition in the retina. *Proc Roy Soc London, B.*, 216(1205):427–459, 1982.

[85] P Sterling and S Laughlin. *Principles of Neural Design*. MIT Press, 2015.

[86] JV Stone. *Independent Component Analysis: A Tutorial Introduction*. MIT Press, Boston, 2004.

[87] JV Stone. Footprints sticking out of the sand (Part II): Children's Bayesian priors for lighting direction and convexity. *Perception*, 40(2):175–190, 2011.

[88] JV Stone. *Vision and Brain: How we perceive the world*. MIT Press, 2012.

[89] JV Stone. *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*. Sebtel Press, Sheffield, England, 2013.

[90] JV Stone. *Information Theory: A Tutorial Introduction*. Sebtel Press, Sheffield, England, 2015.

[91] JV Stone, IS Kerrigan, and J Porrill. Where is the light? Bayesian perceptual priors for lighting direction. *Proceedings Royal Society London (B)*, 276:1797–1804, 2009.

[92] SP Strong, RR De Ruyter Van Steveninck, W Bialek, and R Koberle. On the application of information theory to neural spike trains. In *Pac Symp Biocomput*, volume 621, page 32, 1998.

[93] GS Tomassy, DR Berger, H Chen, N Kasthuri, KJ Hayworth, A Vercelli, HS Seung, JW Lichtman, and P Arlotta. Distinct profiles of myelin distribution along single axons of pyramidal neurons in the neocortex. *Science*, 344(6181):319–324, 2014.

[94] JH van Hateren. A theory of maximizing sensory information. *Biological Cybernetics*, 68:23–29, 1992.

[95] MJ Wainwright. Visual adaptation as optimal information transmission. *Vision research*, 39(23):3960–3974, 1999.

[96] Z Wang, X Wei, A Stocker, and D Lee. Efficient neural codes under metabolic constraints. In *NIPS2016*, 2016.

[97] X Wei and AA Stocker. Mutual information, Fisher information, and efficient coding. *Neural computation*, 28(2):305–326, 2016.

[98] L Zhaoping. *Understanding Vision: Theory, Models, and Data*. OUP Oxford, 2014.

Index