

Contents

Preface ix

Acknowledgments xi

Introduction xiii

- 1 What Engineers Know about Design 1
- 2 Why an Animal Needs a Brain 11
- 3 Why a Bigger Brain? 41
- 4 How Bigger Brains Are Organized 57
- 5 Information Processing: From Molecules to Molecular Circuits 105
- 6 Information Processing in Protein Circuits 125
- 7 Design of Neurons 155
- 8 How Photoreceptors Optimize the Capture of Visual Information 195
- 9 The Fly Lamina: An Efficient Interface for High-Speed Vision 235
- 10 Design of Neural Circuits: Recoding Analogue Signals to Pulsatile 265
- 11 Principles of Retinal Design 277
- 12 Beyond the Retina: Pathways to Perception and Action 323
- 13 Principles of Efficient Wiring 363
- 14 Learning as Design/Design of Learning 399
- 15 Summary and Conclusions 433

Principles of Neural Design 445

Notes 447

References 465

Index 519

Preface

Neuroscience abounds with stories of intellectual and technical daring. Every peak has its Norgay and Hillary, and we had imagined telling some favorite stories of heroic feats, possibly set off in little boxes. Yet, this has been well done by others in various compendia and reminiscences (Strausfeld, 2012; Glickstein, 2014; Kandel, 2006; Koch, 2012). Our main goal is to evince some principles of design and note some insights that follow. Stories deviating from this intention would have lengthened the book and distracted from our message, so we have resisted the natural temptation to memoir-ize.

Existing compendia tend to credit various discoveries to particular individuals. This belongs to the storytelling. What interest would there be to the Trojan Wars without Odysseus and Agamemnon? On the other hand, dropping a name here and there distorts the history of the discovery process—where one name may stand for a generation of thoughtful and imaginative investigators. Consequently, in addition to forgoing stories, we forgo dropping names—except for a very few who early enunciated the core principles. Nor do the citations document who did what first; rather they indicate where supporting evidence will be found—often a review.

Existing compendia often pause to explain the ancient origins of various terms, such as cerebellum or hippocampus. This might have been useful when most neuroscientists spoke a language based in Latin and Greek, but now with so many native speakers of Mandarin or Hindi the practice seems anachronistic, and we have dropped it. Certain terms may be unfamiliar to readers outside neuroscience, such as physicists and engineers. These are italicized at their first appearance to indicate that they are technical (*cation channel*). A reader unfamiliar with this term can learn by Googling in 210 ms that “*cation channels are pore-forming proteins that help establish and control the small voltage gradient across the plasma membrane of all living cells . . .*”

(Wikipedia). So rather than impede the story, we sometimes rely on you to Google.

Many friends and colleagues long aware of this project have wondered why it has taken so long to complete. Some have tried to encourage us to let it go, saying, “After all, it needn’t be perfect . . .” To which we reply, “Don’t worry, it isn’t!” It’s just that more time is needed to write a short book than a long one.

Acknowledgments

For reading and commenting on various chapters we are extremely grateful to: Larry Palmer, Philip Nelson, Dmitry Chklovskii, Ron Dror, Paul Glimcher, Jay Schulkin, Glenn Vinnicombe, Neil Krieger, Francisco Hernández-Heras, Gordon Fain, Sally Zigmond, Alan Pearlman, Brian Wandell, and Dale Purves.

For many years of fruitful exchange PS thanks colleagues at the University of Pennsylvania: Vijay Balasubramanian, Kwabena Boahen, Robert Smith, Michael Freed, Noga Vardi, Jonathan Demb, Bart Borghuis, Janos Perge, Diego Contreras, Joshua Gold, David Brainard, Yoshihiko Tsukamoto, Minghong Ma, Amita Sehgal, Jonathan Raper, and Irwin Levitan. SL thanks colleagues at the Australian National University and the University of Cambridge. Adrian Horridge, Ben Walcott, Ian Meinertzhagen, Allan Snyder, Doekele Stavenga, Martin Wilson, Srini (MV) Srinivasan, David Blest, Peter Lillywhite, Roger Hardie, Joe Howard, Barbara Blakeslee, Daniel Osorio, Rob de Ruyter van Steveninck, Matti Weckström, John Anderson, Brian Burton, David O'Carroll, Gonzalo Garcia de Polavieja, Peter Neri, David Attwell, Bob Levin, Aldo Faisal, John White, Holger Krapp, Jeremy Niven, Gordon Fain, and Biswa Sengupta.

For kindly answering various queries on specific topics and/or providing figures, we thank: Bertil Hille, Nigel Unwin (ion channels), Stuart Firestein, Minghong Ma, Minmin Luo (Olfaction); Wallace Thoreson, Richard Kramer, Steven DeVries, Peter Lukasiewicz, Gary Matthews, Henrique von Gersdorff, Charles Ratliff, Stan Schein, Jeffrey Diamond, Richard Masland, Heinz Wässle, Steve Massey, Dennis Dacey, Beth Peterson, Helga Kolb, David Williams, David Calkins, Rowland Taylor, David Vaney, (retina); Roger Hardie, Ian Meinertzhagen (insect visual system); Nick Strausfeld, Berthold Hedwig, Randolph Menzel, Jürgen Rybak (insect brain); Larry Swanson, Eric Bittman, Kelly Lambert (hypothalamus); Michael Farries, Ed Yeterian, Robert Wurtz, Marc Sommer, Rebecca Berman (striatum); Murray Sherman,

Al Humphreys, Alan Saul, Jose-Manuel Alonso, Ted Weyand, Larry Palmer, Dawei Dong (lateral geniculate nucleus); Indira Raman, Sacha du Lac, David Linden, David Attwell, John Simpson, Mitchell Glickstein, Angus Silver, Chris De Zeeuw (cerebellum); Tobias Moser, Paul Fuchs, James Saunders, Elizabeth Glowatzki, Ruth Anne Eatock (auditory and vestibular hair cells); Jonathan Horton, Kevan Martin, Deepak Pandya, Ed Callaway, Jon Kaas, Corrie Camalier, Roger Lemon, Margaret Wong-Riley (cerebral cortex).

For long encouragement and for skill and care with the manuscript and illustrations, we thank our editors at MIT: Robert Prior, Christopher Eyer, Katherine Almeida, and Mary Reilly, and copy editor Regina Gregory.

Introduction

A laptop computer resembles the human brain in volume and power use—but it is stupid. Deep Blue, the IBM supercomputer that crushed Grandmaster Garry Kasparov at chess, is 100,000 times larger and draws 100,000 times more power (figure I.1). Yet, despite Deep Blue's excellence at chess, it too is stupid, the electronic equivalent of an idiot savant. The computer operates at the speed of light whereas the brain is slow. So, wherein lies the brain's advantage? A short answer is that the brain employs a hybrid architecture of superior design. A longer answer is this book—whose purpose is to identify the sources of such computational efficiency.

The brain's inner workings have been studied scientifically for more than a century—initially by a few investigators with simple methods. In the last 20 years the field has exploded, with roughly 50,000 neuroscientists applying increasingly advanced methods. This outburst amounts to 1 million person-years of research—and facts have accumulated like a mountain. At the base are detailed descriptions: of neural connections and electrical responses, of functional images that correlate with mental states, and of molecules such as ion channels, receptors, G proteins, and so on. Higher up are key discoveries about mechanism: the action potential, transmitter release, synaptic excitation and inhibition. Summarizing this Everest of facts and mechanisms, there exist superb compendia (Kandel et al., 2012; Purves et al., 2012; Squire et al., 2008).

But what if one seeks a book to set out principles that explain how our brain, while being far smarter than a supercomputer, can also be far smaller and cheaper? Then the shelf is bare. One reason is that modern neuroscience has been “technique driven.” Whereas in the 1960s most experiments that one might conceive were technically impossible, now with methods such as patch clamping, two-photon microscopy, and functional magnetic resonance imaging (fMRI), aided by molecular biology, the situation has reversed, and it is harder to conceive of an experiment that *cannot* be done.



Figure I.1

How do neural circuits use space and power so efficiently? Computer: Image http://upload.wikimedia.org/wikipedia/commons/d/d3/IBM_Blue_Gene_P_supercomputer.jpg. Brain: Photo by UW-Madison, University Communications © Board of Regents of the University of Wisconsin System.

Consequently, the idea of pausing to distill principles from facts has lacked appeal. Moreover, to many who ferret out great new facts for a living, it has seemed like a waste of time.

Yet, we draw inspiration from Charles Darwin, who remarked, “My mind seems to have become a kind of machine for grinding general laws out of large collections of facts” (Darwin, 1881). Darwin, of course, is incomparable, but this is sort of how our minds work too. So we have written a small book—relative to the great compendia—intending to beat a rough path up “Data Mountain” in search of organizing principles.

Principles of engineering

The brain is a physical device that performs specific functions; therefore, its design must obey general principles of engineering. Chapter 1 identifies several that we have gleaned (not being engineers) from essays and books on mechanical and electrical design. These principles do not address specific questions about the brain, but they do set a context for ordering one's thoughts—especially helpful for a topic so potentially intimidating. For example, it helps to realize that neuroscience is really an exercise in “reverse engineering”—disassembling a device in order to understand it.

This insight points immediately to a standard set of questions that we suppose are a mantra for all “reverse engineers”: *What does it do? What are its specifications? What is the environmental context?* Then there are commandments, such as *Study the interfaces* and *Complicate the design*. The latter may puzzle scientists who, in explaining phenomena, customarily strive for simplicity. But engineers focus on designing effective devices, so they have good reasons to complicate.¹ This commandment, we shall see, certainly applies to the brain.

Why a brain?

To address the engineer's first question, we consider why an animal should need a brain—what fundamental purpose does it serve and at what cost to the organism? Chapter 2 begins with a tiny bacterium, *Escherichia coli* which succeeds *without* a brain, in order to evaluate what the bacterium can do and what it cannot. Then on to a protozoan, *Paramecium caudatum*, still a single cell and brainless, but so vastly larger than *E. coli* (300,000-fold) that it requires a faster type of signaling. This prefigures long-distance signaling by neurons in multicellular organisms.

The chapter closes with the tiny nematode worm, *Caenorhabditis elegans*, which does have a brain—with exactly 302 neurons. This number is small in absolute terms, but it represents nearly one third of the creature's total cells, so it is a major investment that better turn a profit, and it does. For example, it controls a multicellular system that finds, ingests, and digests bacteria and that allows the worm to recall for several hours the locations of favorable temperatures and bacterial concentrations.

Humans naturally tend to discount the computational abilities of small organisms—which seem, well . . . , mentally deficient—nearly devoid of learning or memory. But small organisms *do* learn and remember. It's just

that their memories match their life contexts: they remember only what they need to and for just long enough. Furthermore, the mechanisms that they evolved for these computations are retained in our own neurons—so we shall see them again.

The progression bacterium → protozoan → worm is accompanied by increasing computational complexity. It is rewarded by increasing capacity to inhabit richer environments and thus to move up the food chain: protozoa eat bacteria, and worms eat protozoa. As engineering, this makes perfect sense: little beasts compute only what they must; thus they pay only for what they use. This is equally true for beasts with much larger brains discussed in chapter 3.

Why a bigger brain?

The brain of a fruit fly (*Drosophila melanogaster*) is 350-fold larger than *C. elegans*’, and the brain of a human (*Homo sapiens*) is a million-fold larger than the fly’s. These larger brains emerge from the same process of natural selection as the smaller ones, so we should continue to expect from them nothing superfluous—only mechanisms that are essential and pay for themselves. We should also expect that when a feature works really well, it will be retained—like the wheel, the paper clip, the aluminum beer can, and the transistor (Petroski, 1996; Arthur, 2009). We note design features that brains have conserved (with suitable elaborations) across at least 400 million years of natural selection. These features in the human brain are often described as “primitive”—reptilian—reflecting what are considered negative aspects of our nature. But, of course, any feature that has been retained for so long must be pretty effective.

This chapter identifies the core task of all brains: it is to regulate the organism’s internal milieu—by responding to needs and, better still, by anticipating needs and preparing to satisfy them before they arise. The advantages of omniscience encourage omnipresence. Brains tend to become universal devices that tune all internal parameters to improve overall stability and economy. “Anticipatory regulation” replaces the more familiar “homeostatic regulation”—which is supposed to operate by waiting for each parameter to deviate from a “set point,” then detecting the error and correcting it by feedback. Most physiological investigations during the 20th century were based on the homeostatic model—how kidney, gut, liver, pancreas, and so on work independently, despite Pavlov’s early demonstration of the brain’s role in anticipatory regulation (Pavlov, 1904). But gradually anticipatory control has been recognized.

Anticipatory regulation offers huge advantages.² First, it matches overall response capacity to fluctuations in demand—there should always be enough but not too much. Second, it matches capacity at each stage in the system to anticipated needs downstream, thus threading an efficient path between excess capacity (costly storage) and failure from lack of supplies. Third, it resolves potential conflict between organs by setting and shifting priorities. For example, during digestion it can route more blood to the gut and less to muscle and skin, and during exercise it can reverse this priority. This allows the organism to operate with a smaller blood volume than would otherwise be needed. Finally, it minimizes errors—which are potentially lethal and also cause cumulative damage.

Anticipatory regulation includes behavior

An organ that anticipates need and regulates the internal milieu by overarching control of physiology would be especially effective if it also regulated behavior. For example, it could reduce a body's need for physiological cooling (e.g., sweating—which costs energy and resources—sodium and water) by directing an animal to find shade. Moreover, it could evoke the memory of an unpleasant heatstroke to remind the animal to take anticipatory measures (travel at night, carry water). Such anticipatory mechanisms are driven ceaselessly by *memories* of hunger, cold, drought, or predation: *Pick the beans! Chop wood! Build a reservoir! Lock the door!*

The memories of danger and bad times that shape our behavior can be our own, but often they are stored in the brains of our parents and grandparents. We are reared with *their* nightmares—the flood, the drought, the famine, the pogrom. Before written history, which spans only 6,000 years, all lessons that would help one anticipate and thus avoid a lethal situation could be transmitted only by oral tradition—the memory of a human life span. Given that the retention of memories in small brains corresponds to their useful span, and that retention has a cost, human memory for great events should remain vivid with age whereas recent memories of lesser events should fade (chapter 14).

The most persistent dangers and opportunities, those extending far beyond a few generations, eventually become part of the neural wiring. Monkeys universally fear snakes, and so do most humans—suggesting that the response was encoded into brain structure before the lines split—on the order of 35 million years. But beyond alertness for predators, primate societies reserve their most acute observations and recall for relationships within the family and the troop. The benefit is that an individual's chances

for survival and reproduction are enhanced by the *group's* ability to anticipate and regulate. The cost is that the individual must continuously sense the social structure—in its historical context—to receive aid when needed and to avoid being killed or cast out (Cheney & Seyfarth, 2007).

Consequently, primate brains have expanded significantly in parts concerned with social recognition and planning—such as prefrontal cortex and amygdala. Humans greatly expand these areas and also those for social communication, such as for language, facial expression, and music. These regions serve both the cooperative and the competitive aspects of anticipatory regulation to an awesome degree. They account for much of our brain structure and many of our difficulties.

Flies too show anticipatory behavior—to a level consonant with their life span and environmental reach. A fly need not wait for its blood sugar to fall dangerously low, nor for its temperature to soar dangerously high, before taking action. Instead its brain expresses prewired commands: *Find fruit! In a cool spot!* Anticipatory commands are often tuned to environmental regularities that predict when and where a resource is most likely to appear—or disappear. Thus, circadian rhythms govern foraging and sleep. Seasonal rhythms, which broadly affect resource availability, govern mating and reproduction. Consequently, specific brain hormones tuned to day length send orders to prewired circuits: *Court a mate! Intimidate a competitor!*

What drives behavior?

To ensure that an organism will execute these orders, there are neural mechanisms to make it “feel bad” when a job is undone and “feel good” when it has succeeded. These are circuits whose activity humans experience, respectively, as “anxiety” and “pleasure.” Of course, we cannot know what worms or flies experience—but the same neurochemicals drive similar behaviors. This is one wheel that has certainly been decorated over hundreds of millions of years, but not reinvented.

To actually accomplish a task is vastly complicated. Reconsider Deep Blue's task. Each side in chess has 16 pieces—that move one at a time, slowly (minutes), and only in two dimensions. Each piece is constrained to move only in certain ways, and some pieces repeat so that each side has only six different types of motion. This relatively simple setup generates so many possible moves that to evaluate them requires a Deep Blue.

But the organ responsible for anticipatory regulation takes continuous data from every sensory neuron in the organism—both internal and

external—plus myriad hormones and other chemicals. While doing so, it is calculating in real time—milliseconds—how to adjust every body component inside and out. It is flying the fly, finding its food, shade, and mate; it is avoiding predators and intimidating competitors—all the while tweaking every *internal* parameter to match what is about to be needed. Thus, it seems fair to say that Deep Blue is stupid even compared to a fruit fly. This defines sharply the next engineering question: what constrains the design of an effective and efficient brain?

What constrains neural design?

When Hillel was asked in the first century B.C.E. to explain the whole Torah while standing on one leg, he was ready: “That which is hateful to you, do not unto another. The rest is commentary—and now go study.”

There is a one-leg answer for neural design: “As information rate rises, costs rise disproportionately.” For example, to transmit more information by spikes requires a higher spike rate. Axon diameter rises linearly with spike rate, but axon volume and energy consumption rise as the diameter squared. Thus, the essence of neural design: “Send only information that is needed, and send it as slowly as possible” (chapter 3). This key injunction profoundly shapes the brain’s macroscopic layout, as explained in chapter 4. We hope that readers will . . . go study.

If spikes were energetically cheap, their rates would matter less. However, a 100-mV spike requires far more current than a 1-mV response evoked by one packet of chemical transmitter. Obviously then, it is cheaper to compute with the smaller currents. This exemplifies another design principle: minimize energy per *bit* of information by computing at the finest possible level. Chapter 5 identifies this level as a change in protein folding on the scale of nanometers. Such a change can capture, store, and transmit one bit at an energetic cost that approaches the thermodynamic limit. Chapter 6 explains how proteins couple to form intracellular circuits on the scale of micrometers, and chapter 7 explains how a neuron assembles such circuits into devices on a scale of micrometers to millimeters.

It emerges that to compute most efficiently in space and energy, neural circuits should *nanofy*:

1. Make each component irreducibly small: a functional unit should be a single protein molecule (a channel), or a linear polymer of protein subunits (a microtubule), or a sandwich of monomolecular layers (a membrane).

2. Combine irreducible components: a membrane to separate charge and thus permit a voltage, a protein transporter to pump ions selectively across the membrane and actually separate the charges (charge the battery), a pore for ions to flow singly across the membrane and thus create a current, a “gate” to start and stop a current, an amplifier to enlarge the current, and an adaptive mechanism to match a current to circumstance.
3. Compute with *chemistry* wherever possible: regulate gates, amplifiers, and adaptive mechanisms by binding/unbinding small molecules that are present in sufficient numbers to obey the laws of mass action. Achieve speed with chemistry by keeping the volumes small.
4. For speed over distance compute *electrically*: convert a signal computed by chemistry to a current that charges membrane capacitance to spread passively up to a millimeter. For longer distance, regenerate the current by appropriately clustered voltage-gated channels.

Design in the visual system

Having discussed protein computing and miniaturization as general routes to efficiency, we exemplify these points in an integrated system—phototransduction (chapter 8). The engineering challenge is to capture light reflected from objects in the environment in order to extract informative patterns to guide behavior. Transduction employs a biochemical cascade with about half a dozen stages to amplify the energy of individual photons by up to a million-fold while preserving the information embodied as signal-to-noise ratio (S/N) and bandwidth. We explain why so many stages are required.

The photoreceptor signal, once encoded as a graded membrane voltage, spreads passively down the axon to the synaptic terminal. There the analogue signal is digitized as a stream of synaptic vesicles. The insect brain can directly read out this message with very high efficiency because the distance is short enough for passive signaling (chapter 9). The mammal brain *cannot* directly read out this vesicle stream because the distance is too great for passive signaling. The mammal eye must transmit by action potentials, but the photoreceptor’s analogue signal contains more information than action potentials can encode. Therefore, on-site retinal processing is required (chapters 10, 11).

Principles at higher levels

The principles of neural design at finer scales and lower levels also apply at larger scales and higher levels. For example, they can explain why the first

visual area (V1) in cerebral cortex enormously expands the number and diversity of neurons. And why diverse types project in parallel from V1 to other cortical areas. And why cortex uses many specific areas and arranges them in a particular way. The answers, as explained in chapter 12, are always the same: diverse circuits allow the brain to send only information that is needed and to send it at lower information rates. This holds computation to the steep part of the benefit/cost curve.

Wiring efficiency

Silicon circuits with very large-scale integration strive for optimal layout—to achieve best performance for least space, time, and energy. Neural circuits do the same and thereby produce tremendous diversity of neuronal structure at all spatial scales. For example, cerebellar output neurons (*Purkinje cells*) use a two-dimensional dendritic arbor whereas cerebral output neurons (*pyramidal cells*) use a three-dimensional arbor. Both circuits employ a layered architecture, but the large Purkinje neurons lie *above* a layer of tiny neurons whereas the large pyramidal neurons lie *below* the smaller neurons. Cerebellar cortex folds intensely on a millimeter scale whereas cerebral cortex on this scale is smooth.

Such differences originate from a ubiquitous biophysical constraint: the irreducible electrical resistance of neuronal cytoplasm. Passive signals spread spatially and temporally only as the square root of dendritic diameter (\sqrt{d}). This causes a second law of diminishing returns: a dendrite, to double its conduction distance or halve its conduction delay, must quadruple its volume. This prevents neural wires from being any finer and prevents local circuits from being any more voluminous. In both cases conduction delays would grow too large. The constraint on volume drives efficient layout: equal lengths of dendrite and axon and an optimum proportion of wire and synapses. Chapter 13 will explain.

Designs for learning

All organisms use new information to better anticipate the future. Thus, learning is a deep principle of biological design, and therefore of neural design. Accordingly, the brain continually updates its knowledge of every internal and external parameter—which means that learning is also a brain function. As such, neural learning is subject to the same constraints as all other neural functions. It is a design principle that must obey all the others.

To conserve space, time, and energy, new information should be stored at the site where it is processed and from whence it can be recalled without

further expense. This is the synapse. Low-level synapses relay short-term changes in input, so their memories should be short, like that of a bacterium or worm. These synapses should encode at the cheapest levels, by modifying the structure and distribution of proteins. High-level synapses encode conclusions after many stages of processing, so their memories deserve to be longer and encoded more stably, by enlarging the synapse and adding new ones.

A new synapse of diameter (d) occupies area on the postsynaptic membrane as d^2 and volume as d^3 . Because adding synapses increases costs disproportionately, learning in an adult brain of fixed volume is subject to powerful space constraints. For every synapse enlarged or added, another must be shrunk or removed. Design of learning must include the principle “save only what is needed.” Chapter 14 explains how this plays out in the overall design.

Design and designer

This book proposes that many aspects of the brain’s design can be understood as adaptations to improve efficiency under resource constraints. Improvements to brain efficiency must certainly improve fitness. Darwin himself noted that “natural selection is continually trying to economize every part of the organization” and proposed that instincts, equivalent in modern terms to “genetically programmed neural circuits,” arise by natural selection (Darwin, 1859). So our hypothesis breaks no new conceptual ground.

A famous critique of this hypothesis argues that useless features might survive pruning if they were simply unavoidable accompaniments to important features (Gould & Lewontin, 1979). This possibility is undeniable, but if examples are found for neural designs, we expect them to be rare because each failure to prune what is useless would render the brain less efficient—more like Deep Blue—whereas the brain’s efficiency exceeds Deep Blue’s by at least 10^5 .

So what do we claim *is* new? The energy and space constraints have been known for a while, as have various principles, such as “minimize wire.” The present contribution seems to lie in our gathering various rules as a concise list and in systematically exemplifying them across spatial and functional scales. When a given rule was found to apply broadly with constant explanatory power, we called it a “principle.” Ten are listed as a round number. As with the Biblical Commandments and the U.S. Bill of Rights, some readers

will find too many (redundancy) and others too few. We are satisfied to simply set them out for consideration.

Some readers may object to the expression “design” because it might imply a designer, which might suggest creationism. But “design” can mean “*the arrangement of elements or details,*” also “*a scheme that governs functioning.*” These are the meanings we intend. And, of course, there *is* a designer—as noted, it is the process that biologists understand as natural selection.³

Limits to this effort

Our account rests on facts that are presently agreed upon. Where some point is controversial, we will so note, but we will not resort to imagined mechanisms. Our goal is not to explain how the brain *might* work, but rather to make sense of what is already known. Naturally what is “agreed upon” will shift with new data, so the story will evolve. We gladly acknowledge that this account is neither complete nor timeless.

We omit *so* much—many senses, many brain regions, many processes—and this will disappoint readers who study them. We concentrate on vision partly because it has dominated neuroscience during its log growth phase, so that is where knowledge goes deepest at all scales. Also we have personally concentrated on vision, so that is where our own knowledge is deepest. Finally, to apply principles across the full range of scales, but keep the book small, has required rigorous selection. We certainly hope that workers in other fields will find the principles useful. If some prove less than universal and need revision, well, that’s science. The best we can do with Data Mountain really is just to set a few pitons up the south face.

1 What Engineers Know about Design

During the Cold War, the Soviets would occasionally capture a U.S. military aircraft invading their airspace, and with comparable frequency a defecting Soviet pilot would set down a MiG aircraft in Japan or Western Europe. These planes would be instantly swarmed by engineers—like ants to a drop of honey—with one clear goal: to “reverse engineer” the craft. This is the process of discovering how a device works by disassembling and analyzing in detail its structure and function. Reverse engineering allowed Soviet engineers to rather quickly reproduce a near perfect copy of the U.S. B-29 bomber, which they renamed the Tu-4. Reverse engineering still flourishes in military settings and increasingly in civilian industries—for example, in chip and software development where rival companies compete on the basis of innovation and design.

The task in reverse engineering is accelerated immensely by prior knowledge. Soviet engineers knew the B-29’s purpose—to fly. Moreover, they knew its performance specifications: carry 10 tons of explosive at 357 mph at an altitude of 36,000 feet with a range at half-load of 3,250 miles. They also knew how various parts function: wings, rudder, engines, control devices, and so forth. So to grasp how the bomber must work was straightforward. Once the “how” of a design is captured, a deeper goal can be approached: what a reverse engineer really seeks is to understand the *why* of a design—*why* has each feature been given its particular form? And *why* are their relationships just so? This is the step that reveals principles; it is the moment of “aha!”—the thrilling reward for the long, dull period of gathering facts.

Neuroscience has fundamentally the same goal: to reverse engineer the brain (O’Connor, Huber, & Svoboda, 2009). What other reason could there be to invest 1 million person-years (so far) in describing so finely the brain’s structure, chemistry, and function? But neuroscience has been somewhat handicapped by the lack of a framework for all this data. To some degree we

resemble the isolated tribe in New Guinea that in the 1940s encountered a crashed airplane and studied it without comprehending its primary function. Nevertheless, we can learn from the engineers: we should try to state the brain's primary goal and basic performance specifications. We should try to intuit a role for each part. By placing the data in some framework, we can begin to evaluate how well our device works and begin to consider the why of its design. We will make this attempt, even though it will be incomplete, and sometimes wrong.

Designing de novo

Engineers know that they cannot create a general design for a general device—because there is no general material to embody it.^{1,2} Engineers must proceed *from the particular to the particular*. So they start with a list of questions: Precisely what is this machine supposed to accomplish? How fast must it operate and over what dynamic range? How large can it be and how heavy? How much power can it use? What error rates can be tolerated, and which type of error is most worrisome—a false alarm or a failure to respond? The answers to these questions are design specifications.

Danger lurks in every vague expression: “very fast,” “pretty small,” “power-efficient,” “error free.” Generalities raise instant concern because one person’s “very” is another’s “barely.” To a biologist, “brief” is a millisecond (10^{-3} s), but to an electronic engineer, “brief” is a nanosecond (10^{-9} s), and the difference is a millionfold. Engineers know that no device can be truly instantaneous or error free—so they know to ask how high should we set the clock rate, how low should we hold the error rate, and at what costs?

The engineer realizes that every device operates in an environment and that this profoundly affects the design. A car for urban roads can be low slung with slender springs, two-wheel drive, and a transmission geared for highway speeds. But a pickup for rough rural roads needs a higher undercarriage, stouter springs, four-wheel drive, and a transmission geared for power at low speeds. The decision regarding which use is more likely (urban or rural) suffuses the whole design. Moreover an engineer always wants to quantify the particular environment to estimate the frequencies of key features and hazards.

One assumes, for example, that before building a million pickups, someone at Nissan bothered to measure the size distribution of rocks and potholes on rural roads. Then they could calculate what height of undercarriage would clear 99.99% of these obstructions and build to that standard.

Knowing the frequencies of various parameters allows rational consideration of safety factor and robustness: how much extra clearance should be allowed for the rare giant boulder; how much thicker should the springs be for the rare overload? Such considerations immediately raise the issue of expense—for a sturdier machine can always be built, but it will cost more and could be less competitive. So design and cost are inseparable.

Of course, environments change. Roads improve—and then deteriorate—so vehicle designs must take this into account. One strategy is to design a vehicle that is cheap and disposable and then bring out new models frequently. This allows adaptations to environmental changes to appear in the next model. Another strategy is to design a more expensive vehicle and invest it with intrinsically greater adaptive capacity—for example, adjustable suspension. Both designs would operate under the same basic principles; the main difference would lie in their strategies for adaptation to changes in demand. In biology the first strategy favors small animals with short lives; the second strategy, by conserving time and effort already invested, favors larger animals with longer lives. As we will see, these complementary strategies account for many differences between the brains of tiny worms, flies, and humans.

Design evolves in the context of *competition*. Most designs are not *de novo* but rather are based upon an already existing device. The new version tries to surpass the competition: lighter, faster, cheaper, more reliable—but each advance is generally modest. To totally scrap an older model and start fresh would cost too much, take too long, and so on. However, suppose a small part could be modified slightly to improve one factor—or simply make the model prettier? The advance might pay for itself because the device would compete better with others of the same class. A backpacker need not outrun the bear—just a companion—and the same is true for improvements in design. The revolutionary Model T Ford was not the best car ever built, but it was terrific for its time: cheaper and more reliable than its competitors.

How engineers design

An engineer takes account of the laws of physics, such as mechanics and thermodynamics. For example, a turbine works most efficiently when the pressure drop is greatest, so this is where to place the dam or hydro-tunnel. Similarly, power generation from steam is most efficient at high temperatures, which requires high pressures. But using pressure to do work is most efficient when the pressure change is infinitesimally small—which takes

infinitely long. There is no “right” answer here, but the laws of physics govern the practicality of power generation and power consumption—and thus affect many industrial designs.

Similarly, a designer is aware of unalterable physical properties. Certain particles move rapidly: photons in a vacuum (3×10^8 m in a second). In contrast, other particles move slowly: an amino acid diffusing in water (~ 1 μm in a millisecond)—a difference of 10^{14} . So for a communications engineer to choose photons to send a message would seem like a “no brainer”—except that actual brains rely extensively on diffusion! This point will be developed in chapters 5 and 6.

Designers pay particular attention to the interfaces where energy is transferred from one medium to another. For example, an automobile designed for a V-8 engine needs wide tires to powerfully grip the road. This is the final interface, tire-to-road, through which the engine’s power is delivered; so to use narrow, lightly treaded tires would be worse than pointless—it would be lethal. More generally it is efficient to match components—for their operating capacities, robustness, reliability, and so on. Efficient designs will match the capacities of all parts so that none are too large or too small.

Matching may be achieved straightforwardly where the properties of the input are predictable, such as a power transformer driven by the line voltage, or a transistor switch in a digital circuit. But the engineer knows that the real world is more variable and allows for this in the design—by providing greater tolerances, or adjusting the matches with feedback. And to estimate what tolerances or what sorts of feedback are needed, the engineer—once again—must analyze the statistics of the environment. Chapters 8–12 will do this for vision.

What components?

Having identified a specific task, its context and constraints, a designer starts to sketch a device. The process draws on deep knowledge of the available components—their intrinsic properties (both advantageous and problematic), their functional relationships, robustness, modifiability, and cost. A mechanical engineer draws from a vast inventory of standard bolts, gears, and bearings and exploits the malleability and versatility of plastics and metal alloys to tailor new parts to particular functions. For example, Henry Ford, in designing his 1908 Model T, solved the mechanical problem of axles cracking on roads built for horses by choosing a tougher, lighter steel alloyed with vanadium.³ An electrical engineer solves

an electronic problem by drawing on a parts catalog or else drawing on known properties and costs to design a new chip. Consequently, as models advance, the number of parts grows explosively: the Boeing 747 comprises 6 million parts.

In these respects the genome is a parts catalog—a list of DNA sequences that can be transcribed into RNA sequences (“messengers”) that in turn can be translated into amino acid sequences—to create proteins that serve signaling in innumerable ways. This extensive genetic parts list is not the end, but rather a start—for there are vast opportunities for further innovation and tailoring (see chapter 5). An existing gene can be duplicated and then modified slightly, just as an engineer would hope, to produce an alternative function. For example, the protein (*opsin*) that is tuned to capture light at middle wavelengths (550 nm) has been duplicated and retuned in evolution by changing only a few amino acids out of several hundred to capture longer wavelengths (570 nm). This seemingly minor difference supports our ability to distinguish red from green.

At the next level, a single DNA sequence can be transcribed to produce shorter sequences of messenger RNA that can be spliced in alternative patterns to produce subtle but critical variants. For example, alternative splicing produces large families of receptor proteins with subtly different binding affinities—which give different time constants. Other variants desensitize at different rates. How these variations are exploited in neural design will be discussed, as will the capacity to further innovate and tailor the actual proteins by binding small ions and covalently adding small chemical groups (posttranslational modification). In short, with 20% of our genome devoted to coding neural signaling molecules, plus the additional variation allowed by duplication, alternative splicing, and posttranslational modification, the brain draws from a large inventory of adaptable parts. The versatility of these components, as explained further in chapter 5, is a major key to the brain’s success.

At a still higher level biological design builds on preexisting structures and processes. Where a need arises from an animal’s opportunity to exploit an abundant resource, natural selection can fashion a new organ from an old one that served a different purpose. Famously, for example, the panda’s “thumb” evolved not from the first digit that humans inherited from earlier primates but from a small bone in the hand of its ancestors that served a different purpose (Gould, 1992). Thus, efficient designs can be reached via natural selection from various directions and various developmental sequences. This was recognized a century ago by a key founder of neuroscience, Santiago Ramón y Cajal (1909):

We certainly acknowledge that developmental conditions contribute to morphological features. Nevertheless, although cellular development may reveal how a particular feature assumes its mature form, it cannot clarify the utilitarian or teleological forces that led developmental mechanisms to incorporate this new anatomical feature. (edited for brevity)

Neatening up

A designer's list of needed functions might also reveal several that could be accomplished with a single, well-placed component. This has been termed "neatening up," which Ford certainly did for the Model T. For example, rather than manufacture separate cylinders and bolt them together (the standard method), he cast the engine as a solid block with holes for the cylinders. Moreover, rather than use a separate belt to drive the magneto (which provides spark to initiate combustion), he built magnets into the engine's flywheel—thereby reducing parts and weight. The sum of his efforts to improve components (vanadium steel) and design (flexible suspension) and neatening up (engine block, magneto) produced a model that was 25% lighter and delivered 25% more horsepower per pound than the competition, such as the Buick Tourabout.

Brain design reflects this process of neatening up. For example, one synapse can simultaneously serve two different pathways: fast and slow; ON and OFF. One neuron can alternately serve two different circuits: one during daylight and another during starlight (chapter 11). But this strategy must not compromise functionality.

Complicate but do not duplicate

Scientists are constantly lashed with the strop from Occam's razor. That is, we are forcefully encouraged to keep our explanatory models and theories *simple*. So the following design principle seems, not merely surprising, but actually counterintuitive: *if one design is simple and another complicated, choose the complicated* (Glegg, 1969; Pahl et al., 2007). Here is the reasoning: when one part is forced to do two jobs, it can do neither well. An example is the two-stroke engine.

The operating cycle of a four-stroke automobile engine involves four sweeps of the piston through the cylinder. One draws in the fuel, the next compresses it, the third delivers power as combustion drives the piston outward, and the fourth sweeps out the exhaust. The two-stroke engine discharges the exhaust with the same stroke that draws in fuel at the bottom

of the combustion stroke and the beginning of the compression stroke. This serves well for a lawn mower or a power saw because the simpler design avoids the need for a valve gear to separately port fuel and exhaust, giving a better ratio of power to weight. However, the four-stroke engine's more complicated design delivers much more power per liter of fuel and runs smoother and quieter. Moreover, its more efficient combustion discharges fewer pollutants (French, 1994).

There are general advantages to providing a separate part for each task. First, each part can be independently tuned for speed, sensitivity, and so forth—without compromise. Second, each part can be regulated independently. Third, more parts provide more opportunities for further refinement, innovation, and improvement—simply because there are more starting points.

Complicate! is such an important principle for neural design that it seems justified to give one example. The vertebrate retina might have used only one type of photoreceptor but instead it uses two: rod and cone. The rod photopigment is more stable but slower to regenerate, so it serves best in dim light. The cone photopigment is less stable but faster to regenerate, so it serves best in bright light. Having complicated the retina's cellular architecture with two cell types, each type has developed its own molecular refinements—specialized versions of the transduction molecules and of *their* regulatory molecules—all tuned for different light intensities. To take full advantage of these refinements, the two cell types have developed different circuits within the retina. However, just before the retinal output, there is a neatening up: rod and cone circuits merge to share a set of excitatory synapses onto a set of common output cells (ganglion cells). Further explanation is to be found in chapters 8 and 11.

There is another way to complicate a design: include several parts that appear to serve the same function. For example, a neuron may express several enzymes that produce the same product. And neighboring cells may express different versions of a similar protein; for example, axons and astrocytes (glial cells) in optic nerve both express a sodium/potassium pump but with subtly different properties. Also one region may connect to another via multiple pathways: dorsal spinocerebellar tract, ventral spinocerebellar tract, spino-reticulo-cerebellar tract, spino-olivo-cerebellar tract, and so on. These parallel features might once have been regarded as “redundant”—to increase reliability and protect against failure. But now most biologists appreciate that multiple pathways generally serve different roles and thus are not truly redundant.

Indeed, engineers try avoid redundancy, and for good reason. A part waiting for a function occupies space, adds weight, and costs extra. So, this consideration raises suspicion that the multiplicity of intracellular phosphatases, sodium/potassium pumps, spinocerebellar tracts, and so on represent complexity of the good kind.

Choosing materials

Engineers can choose from diverse materials. However, they must try to select the least costly material that is appropriate for the task. For a sailboat mast, wood was traditional, but it is heavy. Graphite can be equally stiff for less weight, but it is brittle; titanium gives the best physical performance, but it is costly. So the choice depends on whether the boat is a dinghy for weekend sailing or a 12 meter yacht for the America's Cup.

Brain design is forced to select from a far narrower set of materials. For example, biological membranes are composed of lipids and proteins. Although mechanisms for regulating the passage of substances and ions *across* the membrane in either direction are myriad (ion channels, pumps, cotransporters, antiporters, flippases, etc.), the intrinsic properties of the membrane itself are relatively constant. In particular, the membrane's specific capacitance is fixed at around $1 \mu\text{F cm}^{-2}$. Neurons generate electrical signals by opening and closing channels in the membrane that allow ions to move down their electrochemical gradient and carry charge in and out of the cell.

The time constant of this electrical response is the product of the membrane resistance and capacitance, but capacitance is fixed. Therefore, to speed up an electrical process, a neuron of given surface area must reduce its membrane resistance by opening more channels, thus allowing more ions to cross the membrane. The cost of restoring these ions so as to maintain the electrochemical gradient is high—in fact, it is the human brain's major energy cost: more than 60% goes for pumping ions, making this a key constraint upon design. Thus, for the brain, as for 12 meter yachts and automobiles, speed comes at a premium—and the brain is forced to use it sparingly. This theme will recur.

Integrating across systems

Engineers look for trade-offs among individual components to improve overall performance. For example, because a truck's suspension reduces shock, investment in better springs and shock absorbers can be traded for

weight and strength in the axles. So designers evaluate the whole system to discover where investment in one component is more than compensated by savings in others. This integrated approach to design extends the principle of matching components to include cost.

An example relevant to neural design is the mobile telephone (Mackenzie, 2005). Like many animals, it is small and roams on limited power. Models compete fiercely, and success depends upon performance, beauty, and energy efficiency. One notable innovation provides the phone's "brain," its tiny internal computer, with a *turbo code* that extracts wireless signals from environmental noise. This code employs an algorithm for *belief propagation* that is computationally expensive. But the investment pays because the code eliminates noise so effectively that the efficiency of wireless communication approaches the theoretical limit defined by Shannon's equation (chapter 5).

Optimizing efficiency allows the phone to reduce the amplitude of its output signals. These consume the highest proportion of the phone's power because more energy is needed to transmit radio signals long distances in all directions than to send electrical pulses along short connections in a tiny computer. Consequently, the energy invested in the phone's brain for turbo coding produces much larger savings in the heavy work of signal transmission. By analogy, an animal's small brain saves energy by efficiently directing the activities of large, power-hungry muscles.

To understand the design of an integrated system requires teamwork. When no single person can grasp the details of every component and process, designers team up. Specialists integrate their detailed knowledge of each particular into an efficient whole. It was a team of specialists that reverse engineered the B-29. They needed to combine expertise in aerodynamics, structural engineering, materials science, fluid mechanics, control systems, and so on. Neuroscientists are reaching the same conclusion and forming teams that integrate specialized knowledge to reverse engineer their systems. Brains are integrated systems because they evolved to integrate, so how else can we understand them?

How to proceed and a caution

To consider brain design as a problem of reverse engineering, we must begin with an overview of its main tasks, establish some basic measures of performance, and then see how these relate to the investment of resources in particular mechanisms (chapters 2 and 3). Having established some basic principles, we select one important system—vision—and treat each stage of

processing in the framework of design. We present the environmental context, then the circuit structure and some “hows” of its functioning. Then for each stage we will point out some “whys” of the design and note where other neural systems use similar principles.

The design principles evinced here do not explain everything. In fact, principles cannot explain how anything works—not the B-29, not the Model T, and certainly not the brain. What, then, is their use? Design principles deepen our understanding of why things work the way they do, and armed with this deeper understanding, we can reverse engineer more efficiently. Of course, applying inappropriate or misguided principles would slow us down. Thus, principles derived theoretically, without real objects and mechanisms to illustrate them, are not yet of much use. So we attempt to balance the insights that come from principled explanations against the doubts that come from overdoing them.