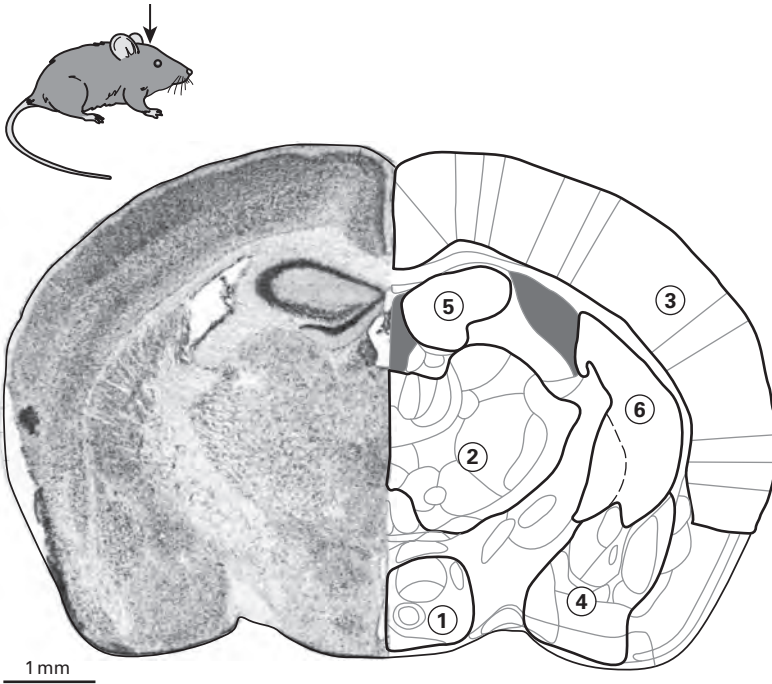


3 Why a Bigger Brain?

This chapter will explain why, despite the worm's success with 302 neurons, brains expand. The mouse cerebral cortex contains about 10^7 neurons. This seems like a lot until you consider that the cortex of the macaque monkey, a key experimental model, is larger by 100-fold, and that human cortex is 10-fold larger still (Herculano-Houzel, 2011). Despite this huge range of scales, one feels comfortable generalizing about the “mammalian brain”—because every part identified in mouse can also be identified in macaque and human (figure 3.1; Kaas, 2005).

Consider also the fly brain. It has 500-fold fewer neurons than the mouse brain, but 500-fold more neurons than the worm brain, plus a rich structure—so warranting a slot in the “large brain” category. Insect and mammal brains share many similarities. For example, both gather their neurons into clusters and their axons into cables (*tracts*). Both employ special structures to accomplish the same broad tasks: store high-level input patterns, generate low-level output patterns, and retrieve patterns using reduced instructions. Of course, there are differences, given the differences in body design and behavior. Yet, despite half a billion years of evolutionary opportunity to diverge, brain designs in insect and mammal seem to have followed the same rules.

For designs to have persisted across this immensity of time and spatial scale implies that they are neither arbitrary nor accidental. Rather, they must have emerged as responses to some broad constraint. That is what elevates the shared responses to the status of *principles*. This chapter will identify the key constraint and indicate how it leads to three principles that govern the organization of larger brains.



- ① Generate patterns for wireless signaling and appetitive behaviors.
- ② “Preprocessing” to shape signals for higher processing.
- ③ High-level processing: assemble larger patterns, choose behaviors.
- ④ “Tag” high-level patterns for emotional significance.
- ⑤ Store and recall.
- ⑥ Evaluate reward predictions.

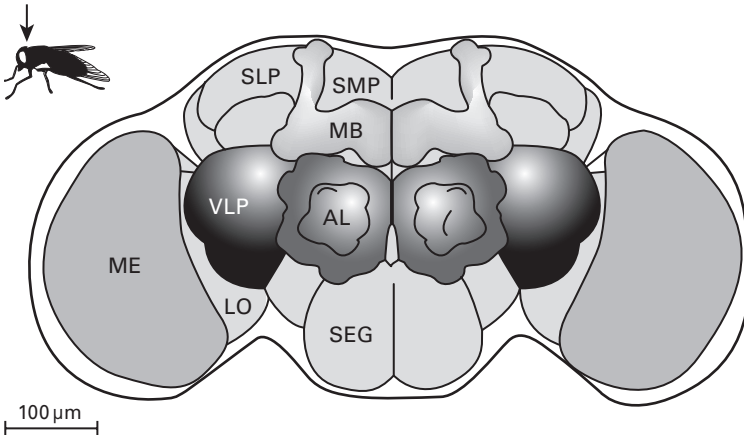


Figure 3.1

Mammalian and insect brains share many broad aspects of design. **Upper:** Cross section through mouse brain; inset indicates plane of section. **Left:** Fine dots are neurons; dark regions are neuron clusters; bright regions are myelinated tracts (chapter 4). **Right:** Numbered regions dedicated to core tasks: (1) *hypothalamus*; (2) *thalamus*; (3) *cerebral cortex*; (4) *amygdaloid complex*; (5) *hippocampus*; (6) *striatum*. Reprinted with modifications and permission from Franklin and Paxinos (1996). **Lower:** Cross section through fly brain; inset indicates plane of section. Brain is built of more than fifty clusters, each specialized for particular tasks. Depicted here are ME, medulla—detect and map local visual patterns; LO, lobula—assemble local visual patterns into larger patterns; AL, antennal lobe—preprocess olfactory signals for pattern recognition; VLP, ventrolateral protocerebrum; SLP, superior lateral protocerebrum; SMP, superior medial protocerebrum—all involved in high-level integration; MB, mushroom body—store and recall; SEG, subesophageal ganglion—integrate information for wired and wireless output to body.

A brain's core tasks

As animals emerge from the soil to a wider, less viscous world, the possibilities for foraging expand immensely. A worm explores mainly in two dimensions over an area of 0.01 m² whereas a honeybee typically covers an area of nearly 10⁷ m², and a fly somewhat less. So foraging area expands by 10⁹ (1 billionfold). Add the third dimension, and the volume to be explored becomes astronomical. Larger animals, such as fish, birds, and mammals, may migrate and thus forage over thousands of kilometers—thus millions of square kilometers.

Such gigantic territories contain immense resources and, of course, harbor innumerable dangers. For an animal to find the one and avoid the other requires it to rapidly gather vast amounts of information from the environment. To calibrate “vast” with one example, the eye sends the brain about 10 megabits per second, roughly the rate of an Ethernet connection (Koch et al., 2006). All sense data reach the brain in the form of tiny patterns—evanescent pieces of a dynamic jigsaw puzzle—and to be of any use, they require assembly to reveal a larger pattern. So if gathering information is to be at all rewarding, the brain must commit resources to assembling larger patterns on spatial and temporal scales that are relevant to behavior.

Yet, even a larger pattern might be useless until it is compared to a library of stored patterns where it can be identified: *edible/toxic*, *friend/foe*, or *search item not found*. Either outcome provides a basis for behavioral choice. A

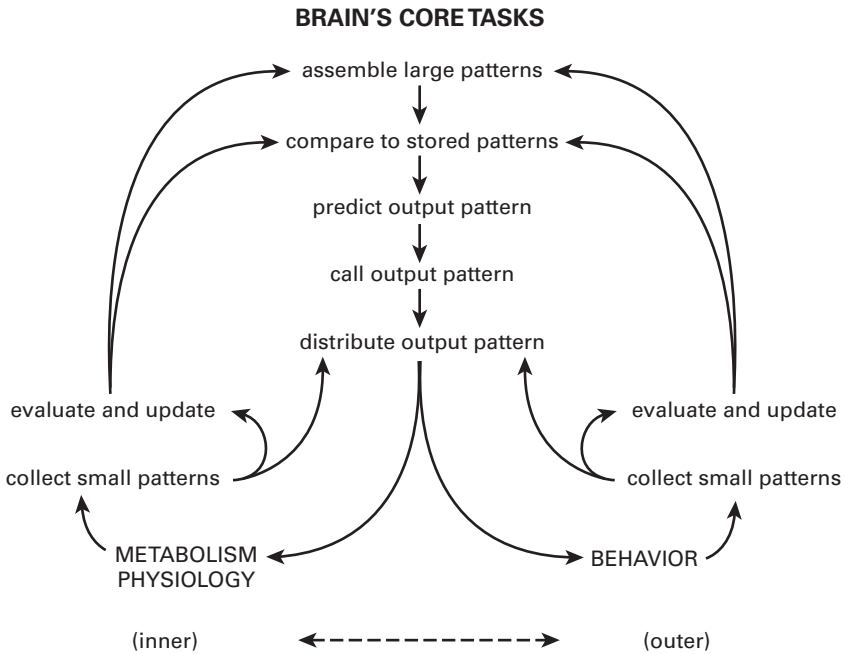


Figure 3.2
Large brains accomplish the same broad tasks. Note that inner and outer tasks couple to serve *each other* (\leftrightarrow).

world; others store metabolites and energy-rich compounds for release upon demand. Still others regulate ionic balance and cleanse the internal milieu, or distribute oxygen and metabolites to hungry tissues. Specialized organs of immunity protect against infectious agents and parasites. Organs couple to form systems, and systems cross-couple to optimize overall function.

The standard idea is that the internal systems more or less take care of themselves. Each parameter is supposed to have a set point, like a thermostat, from which deviations trigger feedback to correct the mismatch (*homeostasis*). Internal regulation also employs *autonomic nerves*—so termed because they are in some sense independent of voluntary control—thus, autonomous. We cannot “will” our heart to beat faster or our blood pressure to decrease. However, we can accomplish these shifts by recalling or imagining the appropriate scene. This implies the existence of neural pathways from pattern stores to pattern generators for autonomic circuits. Thus, although the autonomic nerves are generally supposed to serve

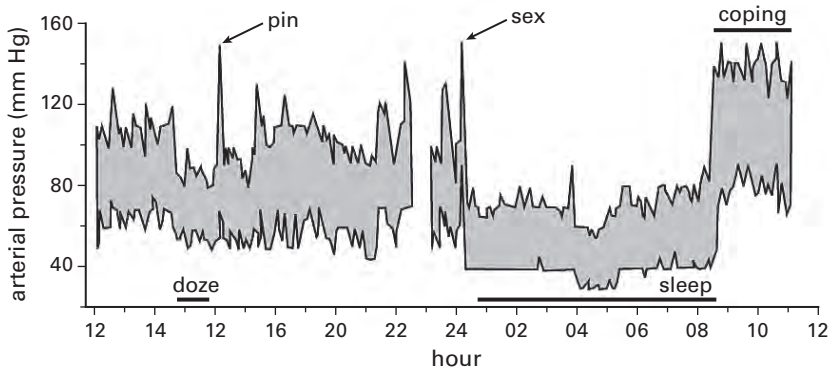


Figure 3.3

Internal systems match behavior. Arterial pressure fluctuates with demand. Each shift in pressure is accompanied by parallel shifts in hormonal and neural signaling that follow the broad catabolic and anabolic patterns. Redrawn from Bevan et al. (1969) and reprinted from Sterling (2004b).

Such anticipatory tuning requires coordinated action of multiple organs and organ systems. To raise pressure, the heart accelerates and vessels constrict. Also the kidney expands blood volume by pumping more salt water into the circulation. The kidney also signals the brain that the body will soon need more supplies of salt and water. Thus, like the gastrointestinal tract, the kidney alerts the brain well in advance of an upcoming need to resupply. Each contribution operates on a different timescale: faster for heart and vessels, slower for kidney's pumping, and still slower for the brain's rise of salt appetite and thirst. These contributions to internal regulation are all initiated simultaneously—and largely by the same signals.

In short, every move we make is matched by a corresponding cardiovascular and renal pattern. Of this we are generally unaware. Yet if the motor command (“Arise!”) slightly precedes the internal command (“Tighten vessels!”), blood flow to the head drops, and we faint. That this experience, *postural hypotension*, occurs rarely attests to the rigorous coupling between the cardiovascular pattern and muscular patterns on a 100-ms timescale. On a slower timescale “Arise!” increases by eightfold a signal to the kidney to save water.²

Note that matching blood pressure to environmental context requires all of the brain's broad tasks as diagrammed in figure 3.2—the collecting and assembling of patterns, the comparison to stores, and so forth. How else to decide if the jab is from a friend or enemy? Moreover, every high-level call

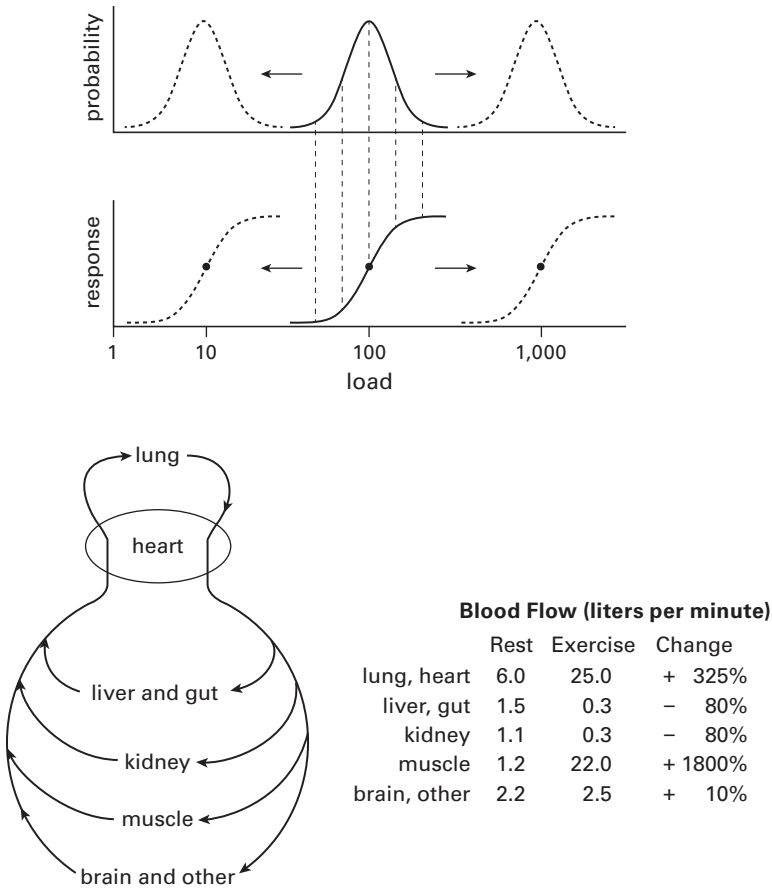


Figure 3.4

Adapt, match, trade. Upper: Adapt response capacity to load. Every system confronts some distribution of probable loads (bold). As conditions shift, so does the distribution (dashed). The response curve (bold) is typically sigmoid with its most sensitive region (steep part) matched to the most probable loads. As a sensor detects a statistically reliable change in the distribution, it prepares the effectors by shifting their response curves to match the new distribution (dashed). Each sensor also adapts its own sensitivity. Reprinted from Sterling (2004b). **Lower:** Organs and organ systems couple efficiently by matching loads to capacities. Trade-offs allow better performance while reducing unused capacity and enhancing “portability.” Blood flow pattern changes with exercise: total flow quadruples, but that is insufficient for muscle. To meet the full need, blood is routed from liver, gut, and kidney, temporarily reducing their performance but eventually benefiting from what the muscular effort has accomplished. Data from Weibel (2000).

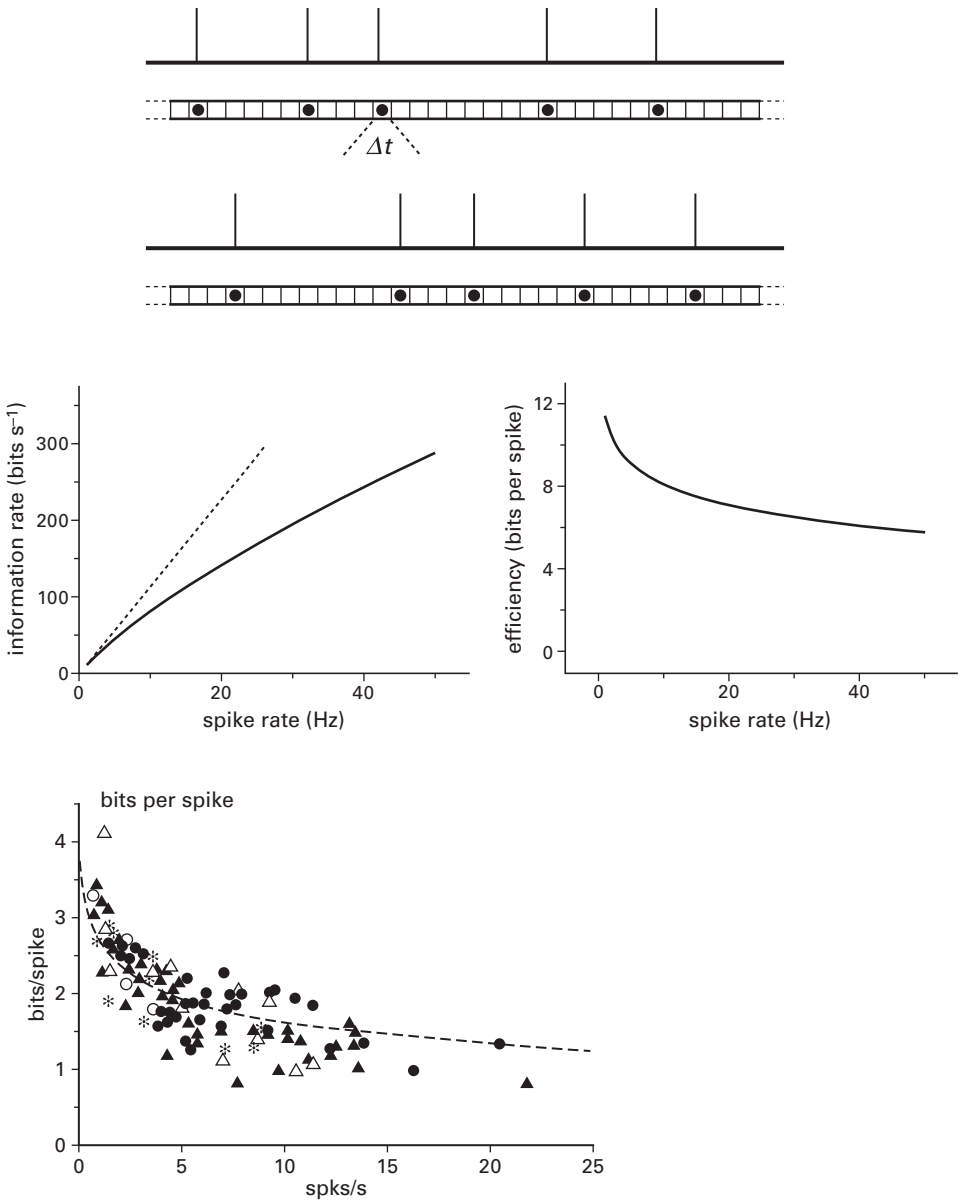


Figure 3.5

Mathematics and biophysics govern the representational capacity of signal trains.

Upper: Distinct sequences of spikes in time intervals Δt represent different inputs.

Middle left: Theory predicts information rate to increase sublinearly with spike rate, with the consequence shown at **middle right:** Increasing spike rate reduces the information transmitted per spike. These theoretical curves were calculated using the standard approximation for signal entropy at low spike rates (Rieke et al., 1997, equation 3.22). In general neurons do not achieve their theoretical capacity because of noise and redundancy; consequently, measured values of bits/spike are lower (figure 11.25). **Lower:** Measured bits per spike falls as mean spike rate increases. Data pooled from several classes of guinea pig retinal ganglion cell. Reprinted with permission from Balasubramanian & Sterling (2009).

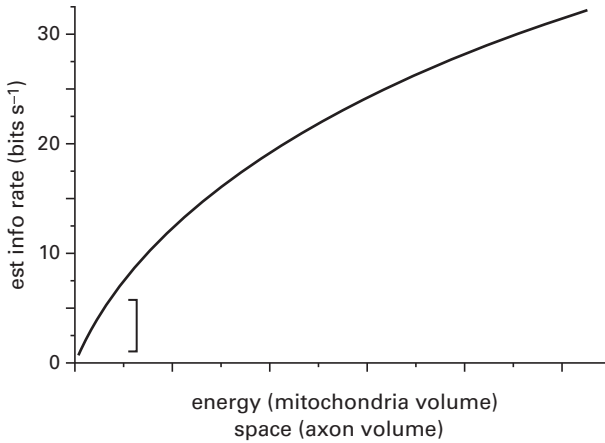


Figure 3.6

Law of diminishing returns. Doubling information rate of retinal ganglion cells more than doubles space and energy costs. Consequently, neural designs try to stay on the steep region of this empirically measured curve. Modified from Balasubramanian & Sterling (2009) and reprinted with permission.

expensive. They use about 20% of the brain’s energy (Attwell & Laughlin, 2001; Sengupta et al., 2010). A spike charges a neuron’s membrane capacitance by about 100 mV, and the membrane area is substantial due to a neuron’s local branching. Higher mean spike rates require a larger cell body with greater membrane area; this increases energy cost per spike and adds to the cost of transmitting bits at high rates. Consequently, where spikes are sent sporadically and at low mean rates, more information can be sent for the same energy—more bits per ATP. This saving in energy by low rates is compounded by a saving in space.

Higher spike rates also require thicker axons.⁴ Because axon diameter, d , increases directly with firing rate, axon volume rises as d^2 ; therefore, doubling the firing rate quadruples axon volume. The concentration of mitochondria, an indicator of energy cost, tends to be constant with axon diameter; therefore, as volume quadruples, so does the energy supply (Perge et al., 2009, 2012). In summary, there is a *law of diminishing returns*: cost per bit, both in energy and space, rises steeply with bit rate (figure 3.6).

Three principles of neural design

The inescapable cost of sending any information and the disproportionate cost of sending at higher rates lead to three design principles: *send only what is needed*; *send at the lowest acceptable rate*; *minimize wire, that is, length and*