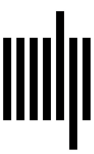
Gradient Expectations

Structure, Origins, and Synthesis of Predictive Neural Networks

By: Keith L. Downing

Citation:

Gradient Expectations: Structure, Origins, and Synthesis of Predictive Neural Networks
By: Keith L. Downing
DOI: 10.7551/mitpress/14723.001.0001
ISBN (electronic): 9780262374675
Publisher: The MIT Press
Published: 2023



The MIT Press

6 Emergence of Predictive Networks

Emergence involves processes running at several timescales: evolutionary, developmental, and epigenetic (aka lifetime learning). Their interactions are so powerful and prevalent that viewing each in isolation can yield a faulty understanding and oversimplified computational models. Evolution produces genomes that serve not as blueprints for brain circuits but as recipes for brain growth and self-assembly during development, and one key determinant of the developing neuronal connective topology is the coincident firing of neurons on both ends of a synapse: the same process, spike-timing-dependent plasticity (STDP), that governs learning. Hence, it is hard to talk about brain evolution without invoking development, which overlaps with learning.

This chapter investigates the origins of predictive networks with respect to all three of these emergent levels. The weight of the discussion involves evolution, but development and learning get drawn in quite often.

The fossil record preserves skulls, not brains, so discovering the evolutionary history of neural systems demands the highest level of scientific sleuthing. Experts in this area have several tools at their disposal to help codify at least some of the key pieces of the puzzle. The old adage that *ontogeny recapitulates phylogeny*—as proposed in 1866 by Ernst Haeckel and thoroughly examined a century later by Stephen Jay Gould (1970)—provides ample justification for examining the embryology of modern animals in search of clues to the evolution of complex life, and intelligence.[1] Comparative anatomy and physiology of extant species also provide convincing evidence of transitions in cerebral structure and function across the millennia, since many of these species have inhabited the earth for many millions of years longer than humans. In addition, genetic analyses of those same species point toward particular mechanisms, such as genetically controlled timing patterns during development, that may underlie those transitions. These and other investigative tools are the basis of several excellent books on the evolution of brains (Allman 1999; Schneider 2014; Striedter 2005; Deacon 1998; Edelman 1992).

In *Intelligence Emerging* (Downing 2015), I weave together several of these accounts, starting with the replication with modification of Hox genes that produce ever more heterogeneous segmented morphologies, with brain segments multiplying and diversifying in the same way as body segments, leading to the *ice cream cone* model (Allman 1999) of the evolving brain. More formally, each scoop of ice cream corresponds to a *neuromere*,

a ringed segment that functions as a modular zone of high cell division (producing neurons) and radial migration (of neurons to their proper layer) during brain development, as discovered by Bergquist and Kallen (1953) and further refined by Puelles and Rubenstein (1993) into what is officially known as the *neuromeric model* (Striedter 2005).

The stacked neuromeres gradually evolve into a hierarchical controller, with the spinal cord, medulla, cerebellum, and basal ganglia growing out of the lower part of the stack and handling many of the faster, subconscious sensorimotor activities, while the cortex and hippocampus sit atop the stack and handle slower, higher-level processing, including conscious reasoning and memory retrieval.

The view of brains as predictive machines does little to change these plausible evolutionary and developmental accounts of emergent neural circuits. However, it does entail a selective pressure for an ability to generate and evaluate expectations, and this interpretive bias would surely shift the anatomical and physiological focus in any of the noteworthy accounts cited above. This chapter takes that selective pressure for granted, so instead of providing a long list of examples in which different species gain a selective advantage via predictive machinery (i.e., the *why* of prediction), it focuses on the *how* and *what* regarding processes that have enabled anticipatory neural circuits to arise.

6.1 Facilitated Variation

Evolution involves three key processes: the production of genetic variation, the inheritance of genomes, and natural selection of the phenotypes produced by genomes. Kirschner and Gerhart's *theory of facilitated variation* (2005) addresses the first pillar, variation, that neither Darwin (selection) nor Mendel (inheritance) was fully equipped to explain. This goes well beyond DNA, crossover, inversion, and mutation, beyond the basic mechanisms of *how*, to the level of *how within the short span of a few billion years*. Because despite all of that time, the odds of purely random genetic change producing earth's levels of biological complexity are astoundingly long, even with the help of selection and inheritance. Kirschner and Gerhart have shown that the process is far from random and leverages an elegant hierarchy of emergent scaffolding.

The gist of facilitated variation is providing the grist for selection: over the course of evolution, genotypic variants become biased toward those with which selection can work to produce well-adapted populations. Thus, most genetic change is not lethal, and at least neutral; and the odds of producing design improvements become surmountable. This capacity to produce viable, inheritable grist for selection is termed *evolvability*, and it relies on two key characteristics: robustness and adaptability. Robust systems are those that can buffer the effects of external perturbations without creating much internal disturbance. For instance, in arctic mammals, a thick layer of subcutaneous fat provides a robust defense against extreme temperatures. Adaptability entails a more active role of internal mechanisms to combat external factors. For example, a dog adapts to hot weather by panting, to expel more internal heat; and from a longer time perspective, a canine species may evolve the ability to produce a thicker coat of hair if it gradually migrates to colder climates. In this sense, the species adapts (i.e., changes its genotype) to produce a more robust phenotype.

Facilitated variation involves three key interlinked mechanisms for attaining robustness and adaptability: modularity, exploratory growth, and functional coupling via *weak linkage*.

Together, these make a species robust to both environmental and genetic perturbations—most mutations are neutral—and adaptive, by making transitions to new, improved genotypes and/or phenotypes possible, if not highly probable. Thus, the canines that migrate to Siberia do not spend hundreds of generations shivering in the cold and producing small litters of weak pups. Instead, their genetic material and/or current phenotypes stand *poised and waiting* to handle many types of disturbances. Getting to that poised state requires modularity, exploratory growth, and weak linkage, combined and improved over long paths through the evolutionary tree.

Kirschner and Gerhart identify many *core processes* that exhibit or contribute to these three key properties. These have arisen at different points in the evolution of life on earth and are *conserved*, that is, remain vital to this day. The older inventions, aged three billion years or more, include energy metabolism, membrane formation, and DNA replication, while more recent biological breakthroughs (only a billion or less years old) are intercellular signaling pathways and body-axis formation (anterior-posterior, dorsal-ventral, and such).

Systems with high evolvability have reached a state in which most components have complex internal activity but reduced and simplified interactions with other components. This preponderance of internal over external interactions constitutes *modularity*, while the presence of many simple signals between components manifests *weak linkage*. This pertains to both phenotypes (in which organs are modular with weak linkage via chemicals in the blood stream) and genotypes, with modular substrings of base pairs comprising structural genes that produce complex protein molecules, but with regulatory signals that turn these genes on and off being relatively simple chemical compounds.

Modularity and weak linkage support a key corollary of facilitated variation: duplication and differentiation. Once the complexities of a unit are largely self-contained, with only weak connections to other units, the possibilities increase for copying it (with or without small variations) and successfully integrating the daughter unit(s) into the existing structure. This applies at both the phenotypic and genotypic levels. Neurons themselves, a phenotypic unit, can easily duplicate, with progeny cells easily linking into the growing network. Underlying the modular phenotypic body and brain segments (e.g., the ice-cream-cone brain model) are modular genetic components, such as the Hox genes, that have been duplicated and differentiated over the ages.

From an evolutionary perspective, the poised state is one in which the complex components are highly robust to genetic change, while the system as a whole achieves adaptability via a flexible network of weak linkages. The core processes only need to change the signals that they send or receive (via receptors) to alter the system's dynamic topology; and since these signals tend to be simple, abundant chemicals, such as calcium, no evolutionary time is wasted trying to invent them. In fact, the DNA contains many structural genes that already have the ability to create particular proteins but are inhibited by the nascent chemical network. Making a significant phenotypic jump from this state requires nothing more than a disinhibitor, for example, a chemical that binds to the active sites of the inhibitor. Weakly linked component topologies are easy to modify, but without messing with the finely tuned recipes for creating the complex components, recipes that took millions of years to evolve.

The third property, *exploratory growth*, typically refers to the ability of components to structurally connect to one another during development. For example, a neuron grows an axonal cone that follows chemical gradients in search of target dendrites. The type of

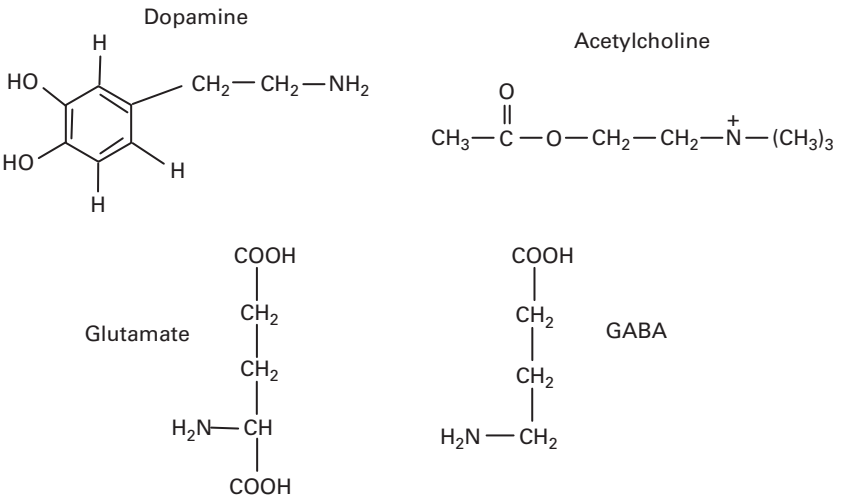


Figure 6.1
Molecular structure of four main neurotransmitters. The hexagonal ring in dopamine consists of six carbon atoms. Note that glutamate is simply CO_2 plus GABA (aka gamma-aminobutyric acid).

dendrite that it seeks may be encoded in the DNA, but its exact location is certainly not: the DNA encodes an algorithm or recipe, not a static blueprint. This ability to explore supports robustness in that one environmental (or genetic) change that alters the location of a developing module (M) in three-dimensional space does not require additional changes for partnering modules to find and connect to M; the exploratory growth process will eventually make its way to M and form the essential structural link. This also enables adaptability, since any *needed* change to a phenotype, such as a longer neck for giraffes, may require only a few genetic changes, with the rest of the body *growing to fit* the altered phenotypic trait. A longer neck does not require additional mutations for longer motor-nerve fibers when those nerves have already evolved the ability to grow to find their target muscles. Thus, the adaptive phenotypic change requires fewer genomic modifications and thus becomes a more accessible point on the evolutionary landscape.

Brains are classic examples of modular, weakly linked systems formed by exploratory growth, where the seven main neurotransmitters (all relatively small molecules) transfer signals across most synapses. In addition, the majority of interneuron signals involve either traveling depolarization waves or direct electrical stimulation (via gap junctions). Similarly, there are four or five main neuromodulators, all of which are very simple chemical compounds compared to proteins. In short, neurons communicate using a currency of only a few simple coins.

Figure 6.1 shows the chemical simplicity of four key neurotransmitters, three of which (dopamine, glutamate, and GABA) are amino acids or direct derivatives of them. In comparison, the simplest protein, ribonuclease, consists of 124 amino acids. Glutamate is the most common excitatory neurotransmitter in the brain, while the most prevalent inhibitor is GABA. Dopamine functions primarily as a neuromodulator; it is broadcast regionally by the basal ganglia. Acetylcholine (ACh) is the main transmitter employed by the spinal cord and is released at all vertebrate neuromuscular junctions (Kandel, Schwartz, and Jessell 2000).

The striking similarity between glutamate and GABA indicates that a neuron that emits one of them would require little genetic retooling to become an emitter of the other. Although the receptors for glutamate and GABA have different origins, most neurons contain receptors for both excitatory and inhibitory compounds. Thus, in nascent species, the key cellular adaptation that could significantly alter the overall neural signaling topology is a change in emitted neurotransmitter, with little or no change needed on the receiving end. The evolutionary emergence of predictive-coding circuits with their excitors, inhibitors, and comparators seems almost inevitable.

The simplicities and similarities of these neural signaling compounds, conserved across the animal kingdom, serve as a preliminary indicator of facilitated variation's role in brain evolution. This chapter digs deeper, in search of other indicators that predictive mechanisms were a natural corollary to emergent intelligence, not only because of their selective advantage, but also because of structural transitions enabled by a system poised and waiting for adaptive opportunities.

6.2 Origins of Sensorimotor Activity

As with many stories of evolutionary origins, the following is only speculative. If we view multicellular organisms as scaled-up versions of bacteria, for example, then their earliest forms should include both sensors and actuators, with the common view that all organisms sense first, then act according to the sensory inputs. However, this sense-and-act (aka outside-in) bias has been strongly contested by a host of prominent cognitive scientists (Buzsaki 2019; Clark 2016; Ballard 2015; Llinas 2001), many of whom view action as primary. Buzsaki (2019) argues that an organism that can only sense but not act would have no survival advantage, while an acting but non-sensing agent could still move (all or part of itself) and fortuitously acquire needed resources.

This argument is well worth noting in the context of prediction, since any system in which expectations play a major role can be one in which the dominance of sensory reality over internal goals or predictions comes into question. The inside-out view of cognition pervades much of the discussion in this and all remaining chapters, but we can nonetheless begin with primitive organisms that have documented abilities to both sense and act, with no necessary priority of one over the other in their cooperative synergy.

An early precursor module to nervous systems, as postulated by many neuroscientists, is a hybrid sensorimotor cell similar to that found in coelenterates such as hydra and sponges. As detailed by Schneider (2014), these cells detect external chemical and physical disturbances via cilia, which can then send electrical signals to a contractile region of the same cell. As shown in figure 6.2, when physically coupled together, the reactive behavior of individual cells can alter the shape of the colony in various selectively advantageous ways. For example, when one cell contracts, its neighbor cells might stretch so as to form a cavity.

In a subsequent evolutionary stage, the sensory and contractile regions might separate into two distinct cells that still communicate electrically, but now via a gap junction, as shown in figure 6.3 (upper left). Sensory cells could then congregate on the periphery, forming a dense ciliary mat, while muscle cells (aka myocytes) underneath would have an expanded sensory field via input from several sensory units. This represents a disjunctive contractile behavior: if *any* of the sensory units activate, contraction will follow.

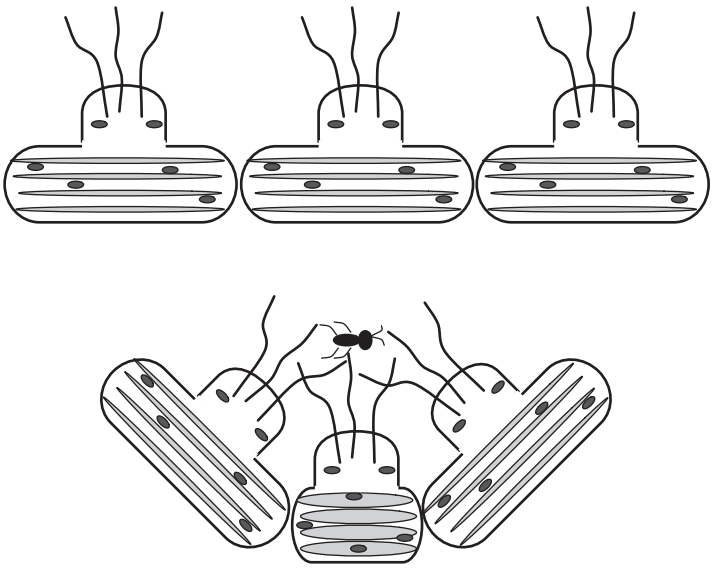


Figure 6.2
Caricature of early hybrid cells (based on those of contemporary coelenterates) that combine sensory (vertical hairs) and motor (horizontal strands) capabilities, as described by Mackie (1970). In response to sensory stimulation, the muscle fibers of the first-stimulated cell might contract, while follow-up inhibitory signals would prevent neighboring cells from contracting, thus creating invaginations to serve purposes such as capturing prey. Note that for larger prey, more cells would sense its presence, thus resulting in more widespread contraction and a larger feeding pocket.

Further spatial separation of sensors and myocytes permits a deeper embedding of motor activity in the organism's core, such that muscle contractions could cause more widespread physical deformation and/or regulate internal liquid or gas flows. However, this distance precludes direct, targeted electrical communication via gap junctions, thus requiring a *cable*, such as a neural filament (figure 6.3, upper right).

The introduction of intermediate neural cells (figure 6.3, lower right) ushers in a new era in which myocytes and sensors communicate only indirectly, via way stations that can accumulate signals from several sensors before stimulating a myocyte. This essentially adds conjunctive functionality to the system: certain myocytes may only contract when *several* sensors register a disturbance within the same short time window.

Finally, the proliferation of interneuron layers yields additional intricacy (i.e., more complex logic) to sensorimotor relationships (figure 6.3, lower left). Also, the emergence of chemical synapses permits filament behavior to change over time, implementing a primitive form of learning.

6.2.1 Origins of Oscillations

Oscillatory dynamics also appear to have evolved from external (muscle) to internal (cortical) structures, another classic example of *encephalization*, as defined by Llinas (2001): the incorporation of peripheral mechanisms into the nervous system. Basically, when cells housing contractile elements are strung together, the electrical coupling between each myocyte forms an active substrate capable of producing undulating waves, which can serve

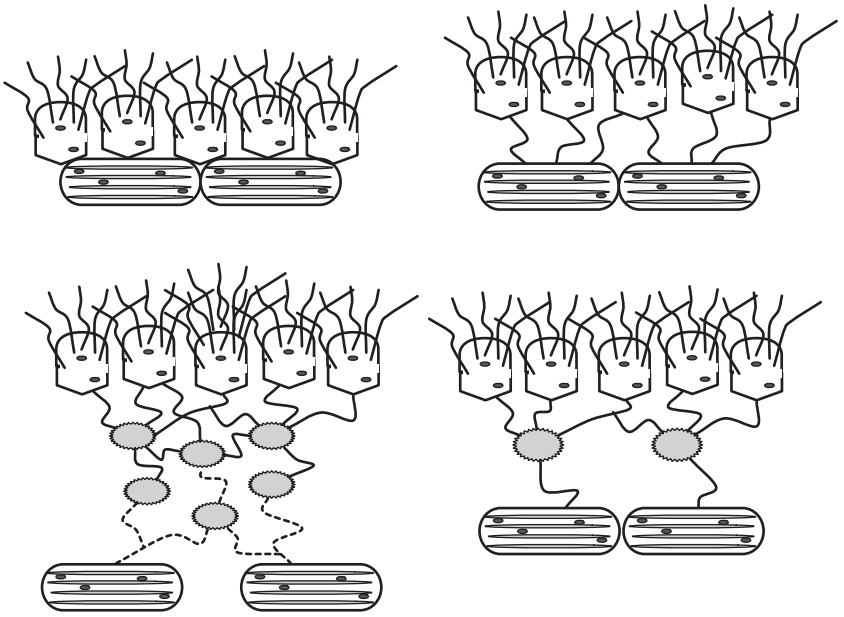


Figure 6.3
The evolutionary complexification of the neural sensorimotor network as envisioned by several neuroscientists (Llinas 2001; Schneider 2014; Mackie 1970). (Top left) The hybrid cell from figure 6.2 evolves into separate sensory and motor cells, coupled electrically via gap junctions. (Top right) Sensory and motor cells separate further and communicate along simple ionic channels with chemical synapses, but with gap junctions still possible between pairs of epithelial or contractile cells. (Bottom right) Interneurons arise as transitions between sensory and motor cells, with cable-like connections between all heterogeneous cells. (Bottom left) Proliferation of intermediate cells enables more intricate relationships between sensing and contracting, less reliance on gap junctions, and more adaptivity due to filaments containing modifiable synapses (dotted lines).

a wide variety of purposes: locomotion, digestion, excretion, circulation, and so on. Figure 6.4 illustrates this basic effect for sheets of coupled muscle fibers. The top diagram shows a simple traveling wave caused by sequential contractions, while the lower caricature of a heart muscle shows how a folded sheet of such fibers can act as a primitive pump.

The extant lamprey serves as a window into the early evolution of oscillations and their role in movement. As shown in figure 6.5, the lamprey's peripheral nervous system includes spinal-cord modules dedicated to body segments, with each module containing identical left and right neural groups. Each group receives input from a corresponding brain-stem area (via connections between B and E), and these inputs excite the motor neurons on that side while inhibiting all activity on the opposing side. Thus, contraction on the left forces relaxation on the right. However, this one signal may be enough to initiate a *fixed action pattern* (FAP) (in a particular segment) that requires no further reticular prodding.

After neuron E fires, thus stimulating M and its corresponding left-side myocyte, it also activates neuron I, which inhibits the opposite (right) side. However, E also excites the slower-charging neuron L, which eventually fires and inhibits the inhibitor, I. This disinhibition of the opposite (right) side initiates a phenomena known as *postinhibitory rebound*, wherein a strongly inhibited area depolarizes, despite the lack of excitatory input. Thus, the right side activates its corresponding motor neuron and myocyte, and inhibits the left side.

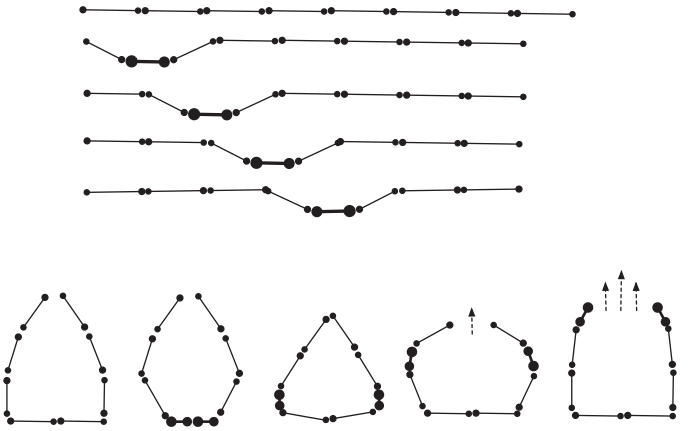


Figure 6.4
Activity waves in coupled muscle fibers (line segments), with thick (thin) segments representing contracted (relaxed) fibers. (Top) A moving invagination formed by sequential contractions. (Bottom) Pumping of liquid from a chamber via parallel chains of sequentially firing units.

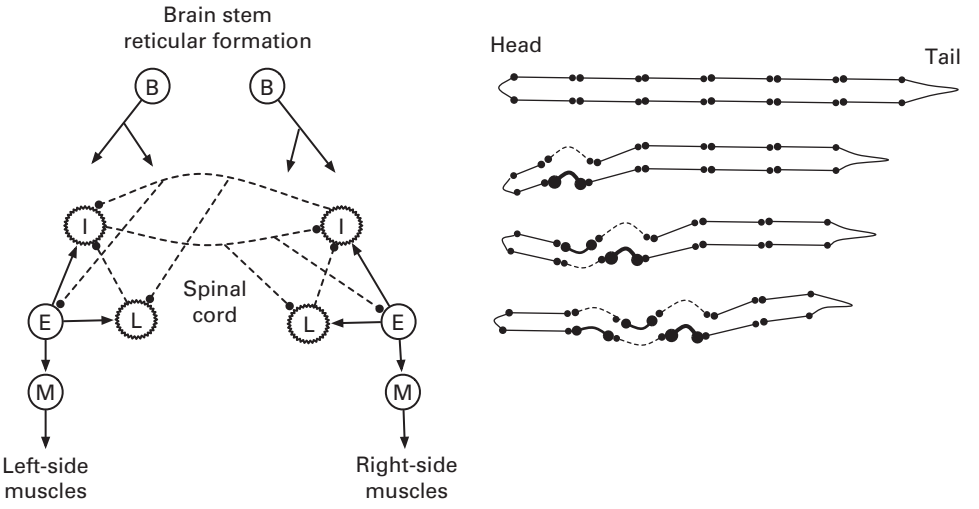


Figure 6.5
Swimming motion in the lamprey and its neural basis. (Left) Neural controller of a body segment. Solid lines are excitatory, while dashed lines are inhibitory. The four main neurons in each pair are I, fast distant inhibitor; L, slow local inhibitor; E, fast excitatory; M, motor. (Right) A wave of paired contraction (thick solid segments) and relaxation (dotted segments) combined with rebound contraction-relaxation produces slithering movement. See Kandel, Schwartz, and Jessell (2000, 745) for more detailed diagrams and descriptions.

When the right-side L neuron activates and blocks the right-side I neuron, a postinhibitory rebound occurs on the left side, and the cycle continues.

As portrayed in figure 6.5 (right), this cycle of paired excitation and inhibition produces the behavioral contraction-relaxation pair, with the former causing its side of the body to press inward and the latter producing an outward bend. Each pair connects to a neighboring pair (going from head to tail in the animal) such that activity in one pair spreads backward. As the main activity wave travels caudally (tailward) in the diagram, the oscillating rebound

deformations move rostrally (headward), thus yielding the characteristic slithering motion of lampreys, eels, and snakes.

Notably, any pair of mutually inhibiting neural groups can form an oscillator if both groups exhibit postinhibitory rebound; and this cycle-producing module can reappear in higher brain regions in the course of evolution. Hence, a primitive motor mechanism, essential for many basic bodily functions, constitutes a *core process*, in the terminology of facilitated variation, which can then ascend the neural hierarchy to become the generator of an internal brain rhythm that may control motion as well as visceral and cognitive functions.

In what looks to be a textbook example of ontogeny recapitulating phylogeny, Llinas (2001) explains how direct electrical coupling between myocytes during early development of the shark produces *motor tremor*, which drives essential embryonic movement and fluid circulation in its prenatal environment. After several weeks, the myocytes lose their interconnections and become governed by electrically coupled motor neurons, thus ushering in the *neurogenic motricity* stage. The adult lamprey exhibits precisely this stage, with the neural controllers of each segment forming a chain of coupled oscillators, whereas in the adult shark, the oscillators have moved from the spinal cord up to the brain stem—and in mammals, they reside at all levels, from the brain stem, cerebellum, and basal ganglia all the way up to the hippocampus, thalamus, and neocortex. The earlier example (in chapter 3) of circuitry for recognition versus action generation illustrates how oscillators could operate in the premotor cortex.

Llinas (2001, 63) sums this up as follows:

> How is this embedding of motricity actually accomplished and why should we want to know? Perhaps in understanding how this was/is accomplished, we will understand something about our very own nature, as the mechanism for internalization must be very closely related to how we process our own thinking, and to the nature of mind and of learning by experience. The short answer to the question, as we have discussed, is that we do this by activation and transfer of intrinsic, oscillatory electrical properties. Basically, motor neurons fire intrinsically, muscles contract rhythmically, and the receptors in muscles and joints respond to the movement and inform motor neurons about their success in producing movement and the direction of such movement in body coordinates. In other words, when I activate motor neurons, I get a sensory echo—and the echo somehow seems to be related to my body's response to the motor order.

The relationships between motor activity and this *sensory echo* is crucial to understanding the predictive brain. In its simplest form, the echo has a very limited range: from the muscles and joints to the spinal cord, where it contributes to local feedbacks that control body position. These spinal reflex loops provide an interesting variant of predictive coding, wherein expectations from higher cortical areas (e.g., motor cortex) function as goals for the body's mechanical subsystems. Roughly speaking, these goals are represented as desired (near future) sensory proprioceptive signals, such as those indicating muscular contraction, skin stretch, and joint rotation, as well as local pressures and temperatures.

Many cognitive scientists (Clark 2016; Shipp, Adams, and Friston 2013; Ballard 2015) draw attention to these spinal reflex loops as key evidence of predictive coding and *active inference* (Friston 2010) (i.e., modifying the sensory system to achieve states compatible with the mind's current expectations). Figure 6.6 illustrates the basic neural circuitry and regulatory process. Crucially, many of the body's motor control tasks require very little intervention from higher cortical areas, such as the premotor and motor cortices. They may

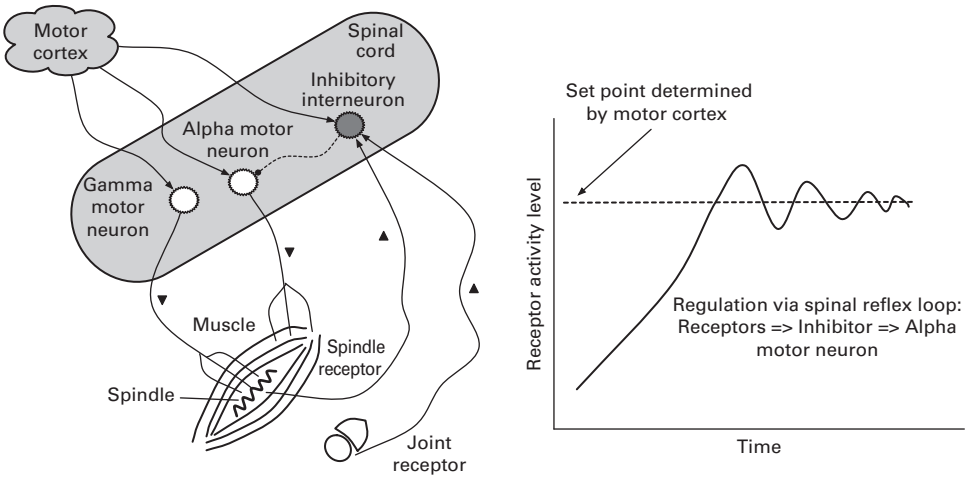


Figure 6.6
Basic mechanisms of muscle control. (Left) High-level sketch of corticospinal and spinomuscular relationships based on similar diagrams in (Kandel, Schwartz, and Jessell 2000) and discussions in (Clark 2016). (Right) Simple graph of the progression of activity in muscle receptors (in spindles, joints, and so on) toward a target level determined by signals initiated in the motor cortex and channeled through the gamma motor neurons, which adjust the spinal cord's sensitivity to the muscle receptors by effectively determining the gain / strength of the receptor pathways back to the spine.

send down a few initiating signals, but after that, feedback loops between the spinal cord (or other lower-level areas such as the medulla or cerebellum) and muscle units get the job done. The figure shows that signals from the motor cortex impinge on several spinal neurons—three of the main types are shown. These signals determine a baseline level of depolarization for each neuron, thus influencing how *jittery* or *relaxed* they will be, and this essentially determines the dynamic parameters and set point(s) for the regulator. For example, the firing rate of the gamma neuron affects physical properties of the spindle, which, in turn, determines the gain on the feedback from the spindle receptor to spinal inhibitory interneurons. These inhibitors form the key negative link in the feedback loop by adjusting activity of the alpha motor neurons, which directly control contraction of the muscle fiber.

The popular predictive-coding interpretation of this physiology is then straightforward. The higher cortical areas have predictions / goals, which they send to the spinal reflex systems, adjusting their parameters to implicitly form set points for the proprioceptive *reality* to be sampled. The negative feedback loops then ensure (under normal conditions) that sensory reality will be manipulated to match the goals, thus reducing the prediction error to zero. When the goal state consists primarily of proprioceptive readings, the concept of active inference seems wholly plausible, since an agent does have considerable control over its own body such that sensory reality can be molded to fit the goal.

As pointed out by Shipp and colleagues, the lower areas (i.e., closer to the spinal tract) of the motor cortex are poorly suited for receiving bottom-up signals from the spinal cord—due to a paucity of neurons in layer 4, the standard avenue of bottom-up projections in all cortical columns (Shipp, Adams, and Friston 2013). They also point out that these connections exist in the human fetus but largely disappear in infants. This indicates that the spinal system in

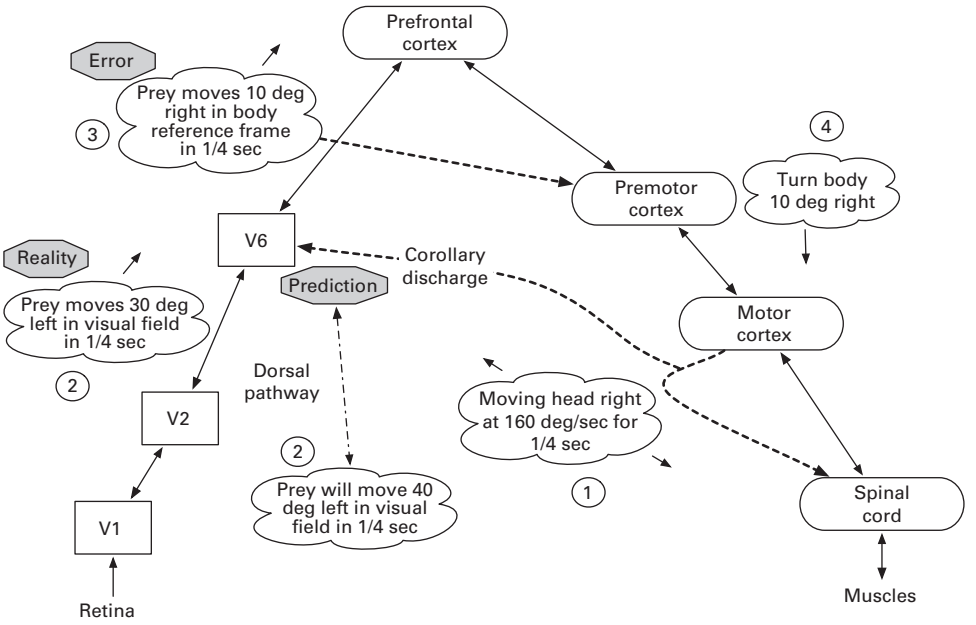


Figure 6.7
Illustrating the role of corollary discharge in predictive coding. V1, V2, and V6 are visual-processing areas. Circled numbers delineate the general sequence of activities. (1) The motor cortex signals the spinal cord to rotate the head 40 degrees in the next 1/4 second. That same information is sent to area V6 on the dorsal sensory pathway as the corollary discharge (aka efference copy). (2) This allows V6 to *predict* the prey's movement in the visual field, which it can then compare to the *sensed* movement, (3) with the difference being the prediction error, which also embodies the *perceived* movement, which is then sent back across lateral connections to the premotor cortex, (4) which computes a body rotation (to keep the prey centered in the body reference frame) and sends it to the motor cortex.

the fully formed organism has no need to send prediction errors upward in the hierarchy, since the low-level regulator will handle the problem (error), and the higher level has all the information that it needs: *the goal state is (or will soon be) achieved.*

When sensation goes beyond proprioception, to signals generated outside the body, the brain cannot assume such unmitigated success in achieving goal states, which now include the telereceptors, for example, those in the eyes and ears that detect signals at a distance. In this case, the prediction and sensory input may interact at higher levels, where telereceptor inputs get preprocessed.

The situation in figure 6.7 is one of the classic examples of predictive coding: the use of corollary discharge to form a prediction of upcoming sensory input, which is then compared to the (only partially preprocessed) sensory reality to yield a prediction error that constitutes the brain's actual perception. This basic process explains a wide range of human skills such as catching a flying object while running and rotating one's head, and oddities, for example, failing to tickle oneself. It also explains a bat's ability to differentiate self signals from those of potential prey during echolocation, and similar capabilities of fish that hunt using self-generated electric fields. Many optical and auditory illusions also stem from corollary discharge.

The contribution of oscillation in figure 6.7 is also important to note. Several of the (presumably neurally encoded) messages include the timestep, 1/4 second. Since cerebral

signaling typically occurs in discrete chunks determined by the primary oscillations used between different areas at particular times, the timestep should be implicit. For example, if the motor cortex and V6 exchange signals on a 4 Hz cycle, then each velocity signal will implicitly entail *over the past 1/4 second*. If this is the standard frequency for interactions between these regions, then all neural circuitry in those regions will become calibrated for that fixed timestep, thus saving both signaling (sending of a timestep) and calculating (multiplying a timestep by a velocity) resources.

6.2.2 Activity Regulation in the Brain

Brain oscillations proliferate during embryonic development, even in the absence of external stimulation, as reviewed and computationally modeled by Abbott and coworkers (2003). These steady firing patterns are both regulated by and regulators of the growing organism in a complex self-organizing process. The rhythmic signal seems crucial for proper wiring of the neural circuitry and then persists into adulthood. Another seminal modeling study, by Ooyen and colleagues (2003), posits homeostatic mechanisms for maintaining activity set points by adjusting dendritic growth, thereby modifying the total input signal.

As another example of *weak linkage*, both studies highlight calcium ion Ca^{++} concentration as the main signaling factor, since it is typically positively correlated with neuronal excitation. In Abbott's group, which studied rhythmic activity in models of the digestive system of crustaceans, Ca^{++} deviations from a set point promote (a) ion-channel addition, removal, and alteration, and (b) changes of synaptic strength. Both regulatory actions stabilize Ca^{++} at the set point. Ooyen's team acknowledges the role of these ion-channel and synaptic mechanisms while examining more significant structural change as a factor in firing-level homeostasis. By allowing dendritic trees to grow and retract in response to Ca^{++}, their models attain stability in networks consisting of either all excitatory neurons or mixtures of excitatory and inhibitory units, with more intense dendritic growth in the mixed networks as some overinhibited neurons search for additional excitatory inputs. Their model includes the known principle that inhibitory neurons mature more slowly than excitatory cells, and this indirectly entails that they need not expand their dendritic tree when understimulated. The weak activation of an inhibitor stems, at least partially, from weak activity in afferent excitatory units, which adapt to increase their own activity by expanding dendritic scope; and this, in turn, increases stimulation to the inhibitors. Thus, the inhibitors need not expend a lot of energy growing dendrites if they can *be patient* and wait for neighboring promoters to do the job for them. Once again, self-organization works things out.

In general, these modeling studies (and considerable additional work cited by the two labs) indicates the ubiquitous control of neural firing rates beginning prenatally—in the simpler, preliminary versions of the brain—and continuing throughout life. Many other structural and functional properties of mature brains presumably stem from this homeostasis as well.

Predictive coding may be one of them. By comparing predictions to reality throughout the brain—not only at the output end—and working to reduce their differences wherever possible, the brain regulates the activity of many neurons, such as those representing prediction error, around low set points. Although the relationship between prediction and regulation and their evolutionary origins may seem like a chicken-and-egg problem, the selective advantage of energy minimization probably trumps that of prediction at the lowest levels of

biological complexity. Thus, neural circuits that restricted neuronal activity levels and the density of neural group activity may have initially served the dual purposes of energy reduction and information maximization, with detailed thermodynamic and information-theoretic analyses (Stone 2018; Sterling and Laughlin 2015) pointing to metabolic efficiency as the absolute highest priority. Then, predictive coding could have emerged as an exaptation (i.e., alternate usage) of this regulatory machinery.

Peter Stone concludes his book on neural information theory with the following insights (Stone 2018, 166), which may apply equally well to predictive coding as to color opponency:

> In the context of neural computation, the evidence collated from a wide variety of research papers suggests that a plausible candidate regarding why physiological mechanisms work in the particular way they do is because they maximize the amount of information gained for each Joule of energy expended (i.e., to maximize metabolic efficiency). In the process of testing this hypothesis, we have found that most functions (like color opponency) which were *once perceived as mysterious* [my emphasis] are actually implemented in ways that are *inevitable* [my emphasis], and even obvious, when considered *in the light of evolution*: specifically, in the context of neural information theory.

Regardless of primary, secondary, and tertiary purposes, the tendency for the brain to maintain optimal activity levels via both short-term dynamics and longer-term structural modifications plays a significant role in learning and development; these homeostatic processes begin in the early embryo and continue throughout life. By modifying genes that determine basic properties of ion channels, synapses, axons, and dendrites, evolution has gradually crafted set points and mechanisms for maintaining them. From all of this arises a complex neural controller that we can accurately call a *prediction machine*.

6.2.3 Competition and Cooperation in Brain Development

Neurons partake in an endless quest to influence neurons. Their own activity has little value if it cannot affect others. However, the brain, as a whole, cannot afford to have all neurons activating one another all the time, since such dense and widespread activity requires considerable energy and greatly reduces the information content of the resulting aggregate signal. Hence, inhibition is omnipresent in the brain: all regions have one or several types of inhibitory interneurons that arise during development by cellular duplication, differentiation, and migration, along with the (more abundant) excitatory neurons.

Inhibitory interneurons often create a competitive environment for neighboring excitatory neurons (e.g., granular or pyramidal cells), such that when one or a small group of these activates, it stimulates nearby inhibitors which then reduce the activity in other neighboring excitatory neurons. This produces *sparse codes* in terms of small groups of active neurons, which generally have higher information content while requiring less energy (Sterling and Laughlin 2015; Stone 2018). A small cluster of active neurons in one area can then transmit its sparse information signal to another area via excitatory connections; inhibitors tend to send signals only locally, so the active granulars or pyramidals constitute the transmitted *message*.

Basically, many neural areas are built for competition between nearby neurons for the right to cooperate with downstream neurons (by activating them). Conversely, there are times when nearby neurons must cooperate (i.e., fire together) in order to achieve downstream firing and thus transmit their message. Many of the brain's neurons have synapses from tens of thousands of other neurons and require hundreds or more inputs to provide

enough post-synaptic stimulation to fire. Competition and cooperation need to achieve a delicate balance in the brain.

Nobel Laureate Gerald Edelman (1987, 1992, 2000) invokes neural competition as a central tenet in his theory of neuronal group selection (TNGS), also known as *neural Darwinism*. This is essentially a *survival of the best networkers*. During development, neurons migrate to their proper locations and then grow axons that search for dendritic targets. In order to survive, the neuron must form these connections and also achieve cooperative firing with downstream neurons. More neurons and connections are formed during development than ultimately survive in the adult brain; the losing (and thus disappearing) synapses are those experiencing few episodes of concurrent pre- and postsynaptic somatic activity, and the dying neurons are those with too few successful synapses. This pressure to hook up and cooperate is particularly intense during development, though a *use-it-or-lose-it* principle also applies to the mature brain.

Competition arises during development because of limited dendritic targets for the growing axons. If region R1 is much larger than region R2, and if both regions are genetically predisposed to target region R3, then R1 will tend to have a competitive advantage based on the sheer number of its axons. Even when some of R2's axons connect to R3, they will have less chance of cooperative firing, since R1 has more neurons and thus more chances for coincident firing of enough R1 neurons to produce activity in a given R3 neuron. All advantages go to the larger region: its enhanced chances of internal cooperation give it a leg up for external competition. Of course, any oscillations within a region that help synchronize internal firing can also provide a competitive advantage in the contest to innervate other regions. Hence, synchrony could help level the playing field if exhibited by R2 but not R1.

Terrence Deacon's displacement theory (1998) leverages neural Darwinism to show how interregional brain topology emerges from exploratory growth and competitive networking, with a minimum of genetic instruction. The DNA needs to encode only general timing constraints regarding the onset and duration for the neuron-generating processes of each region. The onset affects the context into which the new neurons arise, while the duration governs the size of each neural group. The context and size of a group then determine its competition and its relative competitiveness, respectively. The general result is that *large equals well-connected* and thus influential. Deacon (1998) sums this up as follows:

> So although genetic tinkering may not go on in any significant degree at the connection-by-connection level, genetic biasing at the level of whole populations of cells can result in reliable shifts in connection patterns. . . . Relative increases in certain neuron populations will tend to translate into the more effective recruitment of afferent and efferent connections in the competition for axons and synapses. So, a genetic variation that increases or decreases the relative sizes of competing source populations of growing axons will tend to *displace* [my emphasis] or divert connections from the smaller to favor persistence of connections from the larger.

Related work by Finlay and Darlington (1995) promotes the adage that *late equals large*: regions that form later in development tend to be larger. Joaquin Fuster's (2003) seminal research on the cortex and its development supports this principle, since, in primates, the posterior sensory cortices and intermediate motor cortices mature sooner than the frontal cortical areas, which are larger and play a dominant role in monkey and human perception and motor control.

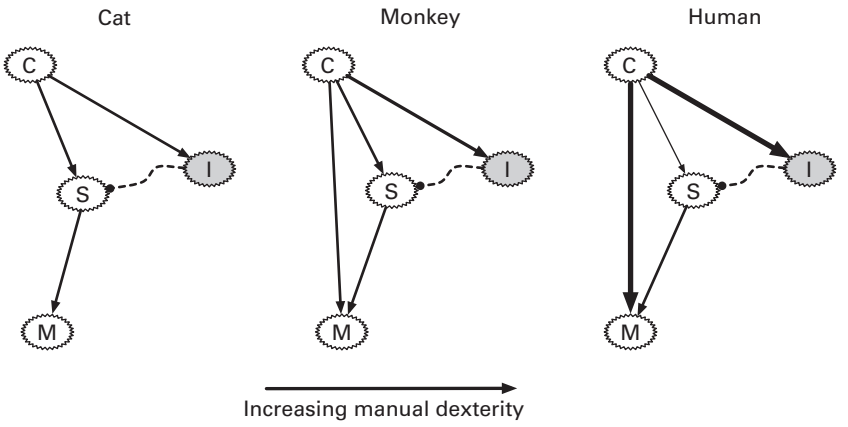


Figure 6.8
Gradual invasion of cortical regions into the motor neurons of the lower spinal cord across mammalian species based on text and diagrams in (Striedter 2005). C = excitatory cortical neurons, S = excitatory spinal neurons, I = inhibitory spinal neurons, M = motor neuron. Arrow thickness denotes strength of influence. Solid arrows are excitatory, dotted links inhibitory.

Striedter (2005) combines these results to show how small changes in developmental timing have led to later maturation in frontal regions, which enables them to grow larger and achieve greater control over a large expanse of lower cortical areas, which, in turn, seems to enhance behavioral sophistication. To wit, in studies of higher mammals, Nakajima et al. (2000) show a strong correlation between manual dexterity and the degree to which frontal cortices control motor areas.

Details of the neural circuitry investigated by Nakajima and coworkers reveal interesting parallels to predictive coding. Figure 6.8 sketches the general progression from cats to monkeys to humans in terms of the amount of cortical control of motor neurons. In the cat, cortical outputs primarily influence the upper spinal cord, which then controls the muscles via the motor neurons (in the lower spinal cord). Thus, most motor actions involve fixed-action patterns (FAPs) produced by central pattern generators (CPGs) in the the spinal circuits; the cortex merely sends enabling and disabling signals to the CPGs. In monkeys, the cortex exerts some direct control on the motor neurons while exhibiting more control over the inhibitory spinal neurons. Finally, in humans, motor neurons experience strong control from the cortex and comparatively less from the spinal cord. In addition, the cortex exerts a strong inhibitory effect on the excitatory spinal neurons.

This work highlights a competition between the cortex and spinal cord for motor control. It was originally believed that the cortex won that conflict via Deacon's displacement: as the cortex expanded from cats to primates, it simply out-fought the spinal cord for motor targets. However, Nakajima's lab found that the spinal-motor connections still exist but that the cortex increases its stimulation of spinal inhibition as a supplementary competitive *tactic*. Hence, the inhibition of one layer by another might simply emerge as an evolutionary consequence of the interregional battle for postsynaptic targets.

Combining this interlayer competitive inhibition with the previous principles of late-equals-large and caudal-to-rostral (back-to-front) cortical maturation yields one plausible account of the evolutionary and developmental (aka *EvoDevo*) emergence of cortical

hierarchies configured for predictive coding. If the lower (caudal) cortical areas mature earlier than higher-level (rostral) regions, then the latter should contain more neurons and thus have a competitive advantage for innervating targets. And the greater that advantage, the more excess connections available to inhibit the lower-level competitors. Thus, a natural relationship between a higher and lower layer could be one of inhibition, and one easily achieved by tapping into the inhibitory interneurons already established in each region (as opposed to sending many individual inhibitory connections from the higher to lower level).

As further support for the ontological emergence of predictive-coding circuitry, consider the perspective of a single neuron (X) sending out axons. Under neural Darwinism, X needs to connect to targets and simultaneously activate with them. However, if X directly excites too many targets, then that will (a) provide a lifeline to neurons that can then compete with X for other targets, and (b) produce an over-stimulated network. One way to balance these trade-offs is to stimulate inhibitory neurons. These can co-fire with X, thereby keeping X well-connected and thus alive, while simultaneously depressing many of X's potential competitors and preventing an overheated network. In short, by exciting inhibitors, a neuron enhances its own survival while also helping to modulate overall network activity (and thus enhancing the survival of the organism as a whole).

An important detail omitted from our earlier discussion of cortical columns is that the axons of layer-5 pyramidal neurons realize both top-down links to lower columns and excitatory connections to motor neurons in the spinal cord and, in general, other motor areas, such as the superior colliculus, which controls eye movement (Mountcastle 1998; Hawkins 2004). Thus, one key function of cortical columns (at least in primates) involves motor control. Contrary to the view of the brain as a strict sense-and-act hierarchy—where all sensory signals travel upward through sensory cortex and then premotor signals travel back down along the *motor side*—these layer-5 motor projections appear in the cortical columns of all regions, whether primarily sensory or motor.

The layer-5 projections to other, nonmotor, columns may emerge as by-products of competition for the motor targets during development, especially if we accept the view that top-down links have a net inhibitory effect by exciting local interneurons in the lower columns, either directly or via more indirect means, as described earlier in Hawkins and Ahmad's (2016) model: initial, weak, top-down excitation leads to more widespread inhibition when predictive signals properly anticipate sensory reality.

This view of layer-5 outputs as general-purpose *action* signals—whether promoting or blocking physical movement or prediction—meshes with the view of each higher level functioning as a controller of its lower neighbor(s). Each regulator outputs a controlling action designed to reduce the prediction-error signal.

Neural selectionism versus constructivism

Edelman's TNGS essentially views synapses like fish eggs: many are produced, but only a few survive. This *selectionist* view has been contrasted with neural constructivism from Quartz and Sejnowski's (1997) frequently cited "Constructivist Manifesto." Constructivism views development and cognition as highly intertwined mechanisms, with significant amounts of ontological growth driven by demands of the brain-body-environment interaction. It weighs in on the side of *nurture* in the classic nature-nurture debate by emphasizing flexible neural structures molded by experience over brain topologies built primarily by

innate recipes in the DNA. In so doing, it further blurs the divisions between development and learning by proposing that both (a) the former extends well into life, and (b) the latter plays a vital role in the former. They see this flexibility as an evolutionary trend that increases in more behaviorally complex lifeforms. In their words (Quartz and Sejnowski 1997, 537),

> Neural constructivism suggests that the evolutionary emergence of neocortex in mammals is a progression toward more flexible representational structures, in contrast to the popular view of cortical evolution as an increase in innate, specialized circuits. Human cortical postnatal development is also more extensive and protracted than generally supposed, suggesting that cortex has evolved so as to maximize the capacity of environmental structure to shape its structure and function through constructive learning.

The differences between selectionism and constructivism were greatly exaggerated by early critics of the manifesto such that a fictive dichotomy arose wherein either (a) the brain produces nearly all of its synapses prenatally, or neonatally, and then gradually prunes them away (in direct or indirect response to environmental signals) during childhood, adolescence, and adulthood, or (b) the brain generates axons, dendrites, and synapses in response to environmental signaling and internal neural dynamics throughout life, though primarily during the earlier decades (for humans).

A closer reading of the paper clearly indicates more of a compromise: constructivism posits adaptive growth *and elimination* of neural processes throughout life as part of the dynamic interplay between brain, body, and environment. Experimental neuroscience has since provided strong evidence for Quartz and Sejnowski's general claim (Petanjek et al. 2011; Sanes, Reh, and Harris 2006; Andersen et al. 2007), with significant structural adaptations (i.e., neurogenesis along with axonal growth and death) occurring well into the third decade of life.

In another seminal book, *Natural-Born Cyborgs*, Andy Clark addresses constructivism (Clark 2003, 84):

> The human neocortex and prefrontal cortex, along with the extended developmental period of human childhood, allows the contemporary environment an opportunity to partially redesign aspects of our basic neural hardware itself. The designer environments . . . are thus matched, step-by-step, by dedicated designer brains, with each side of the co-adaptive equation growing, changing and evolving to better fit—and maximally exploit—the other.

This view extends the cooperative-competitive narrative from the internal, interneuron level to the bigger picture of the dynamic, brain-body-world coupling. As discussed in chapter 7, this has strong implications for biologically inspired pursuits of artificial intelligence.

6.2.4 Layers and Modules

The brain is often portrayed as a hierarchical controller that continuously loops through sense-and-act cycles, wherein raw external inputs receive preprocessing on their way up the sensory hierarchy; then, the final, veridical perceptions at the higher levels initiate a signal sequence down the action side of the hierarchy, eventually producing movement. Though this workflow surely occurs in some situations, particularly those involving conscious thought and action planning, the neuroscience reveals an alternate picture (Mountcastle

1998). First, the hierarchy is more of a heterarchy: though many neural layers reside closer to the body-world interfaces (sensors and actuators) and may be in an exclusive group that exchanges signals with some section of the periphery, there are many such low-level layers that send lateral axons to the *other side* and that receive considerable stimulation from high-level layers. In fact, the often-cited ratio of top-down to bottom-up wiring in the brain is 10:1 in favor of top-down. Basically, the static architecture and functional dynamics of the brain indicate a tangled web but with undisputable evidence of heterarchy. Hence, any attempts to cleanly match predictive-coding modules to particular brain regions or cortical columns within the neural web will surely fail.

As mentioned at the outset, many prominent scientists (Buzsaki 2019; Clark 2016; Ballard 2015; Llinas 2001) prefer an inside-out, act-and-sense interpretation of the brain (over the traditional outside-in, sense-and-act view), which profits from generating models of the world that enhance the organism's ability to act, but not from capturing extraneous details of no obvious survival advantage. There is no inherent gain in building accurate mental models, nor in generating precise forecasts. Value comes only in the service of action.

Andy Clark eloquently sums up this perspective as follows (Clark 2016, 250):

> In fact, these distinctions (between perception, cognition and action) now seem to obscure, rather than illuminate, the true flow of effect. In a certain sense, the brain is revealed not as (primarily) an engine of reason or quiet deliberation, but as an organ for the environmentally situated control of action. Cheap, fast, world-exploiting action, rather than the pursuit of truth, optimality, or deductive inference, is now the key organizing principle.

As for prediction's pivotal contribution to this situated control, he adds (Clark 2016, 250),

> What real-world prediction is all about is the selection and control of world-engaging action. Insofar as such agents do try to "guess the world" that guessing is always and everywhere inflected in ways apt to support cycles of action and intervention. At the most fundamental level, this is simply because the whole apparatus (of prediction-based processing) exists only in order to help animals achieve their goals while avoiding fatally surprising encounters with the world.

This interpretation of the brain as a tightly integrated, act-and-sense heterarchy complicates our quest to neatly enumerate the core modules and weak linkages that could paint the predictive mind as an easily achieved variation, given a few hundred million years of evolution. Although cortical layers and distinct sensory and motor pathways look nice on paper, the microscope and MRI seem to tell a different story, both architecturally and functionally. However, within that confusing network reside hints of a predictive coder's building blocks.

As shown above, the interactions among motor cortex, spinal motor neurons, and muscles portray a simple controller whose prediction / goal, provided by the motor cortex, seems readily attainable by a reflexive arc. At higher levels, when corollary discharge forms the basis of a sensory prediction, the resulting prediction error constitutes the actual perception that the brain may use for conscious reasoning and/or further iterations through an act-and-sense loop. In both cases, the predictive-coding module spans several layers of neural circuitry, and at higher levels, the scope can easily include both sensory and motor regions. Whether the actual predictions begin as action requests or expected causes of sensory data (during a pure object-recognition task), they would appear to ground out in sensory expectations, which can then be compared to proprioceptors, telereceptors, or partially preprocessed

signals in the thalamus or lower cortical areas. The central component in all of these cases is some form of comparator, which easily arises in a network replete with both excitatory and inhibitory neurons.

As the story of corollary discharge seems to indicate, prediction error serves as a perception, which, as such, consists of a sensory *reality* term minus a prediction. This paints perceptions as some form of abstraction, as opposed to concrete facts about the agent-world coupling. Abstracted representations are those with many details removed, so in this case, those details are the predicted components, the expectations. Notably, the perception does, in many cases, represent a more objective, ecocentric interpretation of the world. For example, the vestibular-ocular reflex gives the true picture of the world: it is not moving even though the head is. But from an egocentric, action-oriented view of life, which organisms have presumably evolved to honor, the *subjective truth of the matter* is that the world is moving along with the head. Still, the abstraction, a more stable world (the prediction error), is the representation needed to support the ensuing action, for example, stretching and snapping the frog's tongue to catch a fly.

In short, subjective reality (as recorded by an organism's sensory apparatus) constitutes the primary world view, with the maximum amount of detail (of any relevance whatsoever to that species). Predictive coding then filters and normalizes that information (relative to predictions) to create abstractions that more faithfully and efficiently support further action. In some cases, those abstractions provide more objective pictures of the world, while in others they are *helpful illusions* that aid survival (at least on average). The widespread rustling in the bushes may be a mountain lion, or a scurry of chipmunks, but better to be safe than sorry.

Thus, what begins at the periphery as relatively standard control—with the goal of reducing a prediction error defined as the difference between expected and experienced proprioception—becomes a regulator of more abstract information in moving up the neural heterarchy. From an action-oriented perspective, the brain is quite content if most abstractions wither away to nothingness on the way up the ladder, indicating an absence of surprise; and the activity can continue, as *planned*. Conversely, any unexpected sensory readings can create far-reaching abstractions that trigger action changes (and ensuing prediction updates) deeper in the nervous system.

As shown in figure 6.9, the neural heterarchy can, as a first pass, be modeled as a predictive-coding hierarchy of standard PID controllers, where each level works with different abstractions. Upward signaling involves the error terms (e_i), while downward messaging maps a *control variable* (u_i) to a prediction (p_{i-1}) via a transformation matrix of weights (M_{i-1}). At each level, the PID controller (C_i) converts the error (e_{i-1}) into u_i using the combination of the proportional (P), integral (I) and derivative (D) term, each sketched in the inset of figure 6.9.

At the lowest level, sensory input (s_0, predominantly from proprioceptors) is compared to target / predicted settings, producing an error sent upward to controller C_1, which produces a control response u_1, which, in turn, maps downward to possible updates of the expected result p_0 and gain g_0, which represents the attention paid to e_0.

The control decision, u_1, then becomes an aspect of *reality* to be compared to a higher-level prediction, p_1, derived from a higher-level control decision, u_2 (based on $e_1 g_1$), and so on up the hierarchy. At each level, the timescales vary, with higher-frequency updates (i.e., smaller time constants) at the lower levels. In this interpretation, the higher level tries

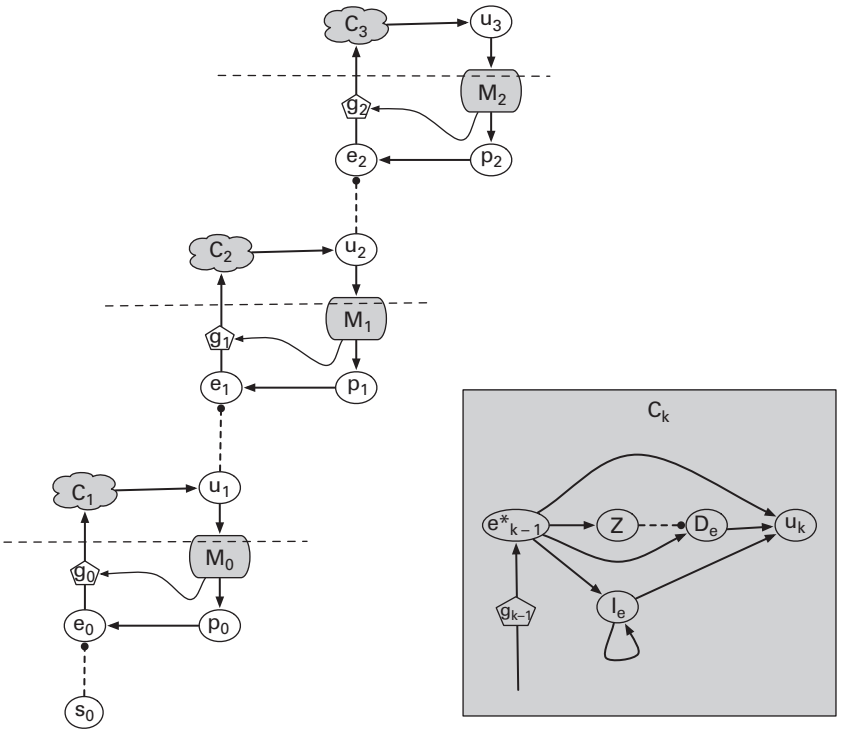


Figure 6.9
Predictive coding as a hierarchical controller, from the action-oriented perspective, where predictions constitute targets. (Left) Dotted horizontal lines separate layers according to conventional predictive-coding models. s_0 = sensory input, e_i = error, p_i = prediction, u_i = control variable produced by controller C_i, M_i = a model, a collection of weights / synapses mapping u_{i+1} to p_i, and g_i = gain applied to bottom-up signals. (Right, inset) Details of a neural PID (proportional, integral, derivative) controller circuit, with output u_k a weighted combination of the (scaled by g_{k-1}) error (from the layer below), the derivative of that error, D_e, and the error's integral I_e; neuron Z achieves the delay for computing D_e.

to predict only what its lower neighbor is *doing* (u_1), not what information it used to *decide* on that action.

From an alternate perspective, one more compatible with an act-and-sense view, p_1 indicates to level 1 what level 2 plans to do. If bottom-up information ($e_0 g_0$) presents a picture of reality that induces C_1 to produce an action (u_1) similar to p_1, then the ensuing low error (e_1) tells layer 2 that its action can be carried out, as planned. Conversely, a high error forces layer 2 to either modify its intentions or reduce its attention to e_1, by lowering g_1. In short, the higher level is telling the lower level its plans and then trying, via feedback, to reconcile them with the bottom-up reality signals—by either changing its plan or reducing the importance of the reality signal.

The primitive building blocks of the hierarchical controller also suffice to configure a value system similar to the basal ganglia (discussed in chapter 3). Figure 6.10 depicts the interplay between a predictive-coding hierarchy and a value system, which also receives control output u_3; but instead of transforming it into a prediction of a lower-level control output, it maps u_3 to an evaluation (V), thus mimicking the *critic* of a reinforcement learning (RL) system. In RL systems, these evaluations constitute predictions of cumulative future reward.

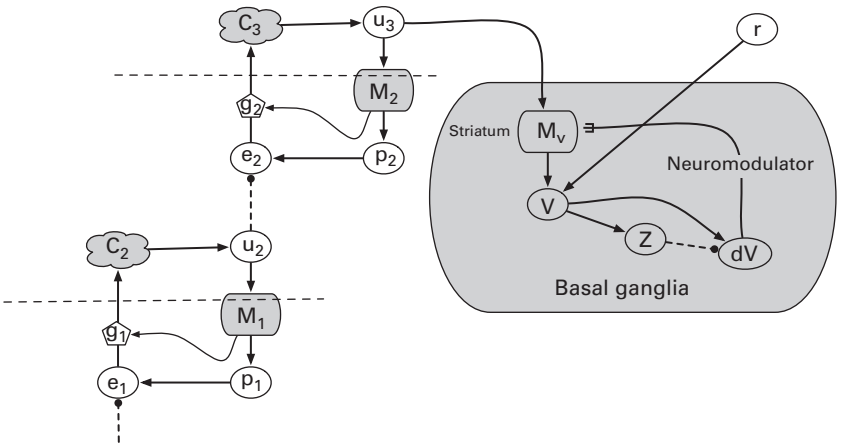


Figure 6.10
Enhancement of figure 6.9 to include reinforcement learning via an auxiliary network similar to the basal ganglia, where model M_v functions similar to the striatum and neighboring areas by mapping states to evaluations (V), while r denotes a reinforcing reward or penalty. The inverted delay connection between V and dV computes the temporal derivative of V, the *temporal difference error* described in chapter 3, which then governs the secretion of neuromodulator into the striatum, thus stimulating synaptic modification.

Employing a standard delayed inversion, the circuit then computes the temporal gradient of that prediction (dV), which represents RL's temporal-difference error (aka surprise). Neuromodulation (via chemicals such as dopamine) then responds to surprise, causing learning via synaptic change in the striatum and neighboring areas.

The similarity of building blocks between the hierarchical layers and the value system indicate the tractability of evolving either of them once nature had discovered the other, or the potential for coevolving these structures. The most significant difference lies in the neuromodulatory mechanism, which involves diffuse chemical signaling as opposed to targeted synaptic transmission. Yet, several common neuromodulators, such as dopamine and acetylcholine, also function as neurotransmitters, thus easing any evolutionary transitions from a signal-transmitting to a signal-spreading region (or vice versa).[2]

6.2.5 Running through the Woods on an Icy Evening

As a concrete, though clearly speculative, example of this heterarchical controller in action, consider the task of running on an icy trail. At the lowest level, the leg/feet joint and muscle proprioceptors frequently deviate from their expected values (p_0) due to omnipresent slippage and simple, high-frequency, reflex responses. Thus, e_0 cannot be completely ameliorated, and its scaled (by g_0) and accumulated (by I_e in C_1) value translates into the need for a higher-level adjustment (u_1), such as a greater angle of foot contact (landing on the heel and thus creating a higher pressure point to dig into the ice-coated dirt). However, this altered landing angle conflicts with the predicted angle (p_1) and this error (e_1) will, with sufficient gain (g_2) trigger C_2.

The predicted landing angle (p_1) derives from u_2, which represents a general concept of *stride*, a combination of step distance and body elevation (among other things). For example, if stride consists of a long step distance but a low elevation, the proper (predicted) landing angle will probably be high, whereas increased height facilitates a flat-footed or toe landing.

Thus, changes to stride (u_2) can modify the predicted landing angle (p_1) to better match that chosen by C_1. On a mildly icy trail, digging in with the heels works fine, but when surface friction decreases and hardness increases, the better response is to land flat-footed for maximum surface contact. Either way, the change to landing angle will affect the stride, and vice versa. Once again, the scaled and accumulated landing-angle error will cause adjustments (via controller C_2) to the stride. Signal flow back and forth between these two lower levels will continuously regulate the running motion.

Now throw an obstacle into the mix: rounding a sharp bend, a large branch comes into view, five meters ahead. The telereceptive (visual) system notices this surprise, after several levels of processing, and feeds that information to motor control at a high level, for example, C_3, whose output (u_3) represents a particular stride sequence, for example, a combination of bounding and stutter steps to set up the proper foot plant that immediately precedes the hurdling motion. Thus, u_3 translates into predicted / goal strides, which must broker some sort of compromise in terms of landing angles so as to avoid extreme slippage while properly preparing for hurdling (which is difficult from a heel-first landing and normally requires some pushoff with the toes).

These continuous top-down, bottom-up compromises cloud the issue of whether behavior consists of iterated sense-and-act or act-and-sense sequences, but this only reflects the bigger picture of the body and environment engaged in an interactive loop with frequently murky distinctions between causes and effects. Hence, the inputs to each error unit (p_i and u_i) may never neatly represent a (relatively fixed) target (e.g., a set point) and a more malleable factor. Instead, the complex dynamics of the circuitry may force one variable to conform to the other in one situation while yielding a reversal of that dominance in another scenario.

Likewise, the representation of each error node as the sum of a positive and negative term is a drastic over-simplification, but if predictive coding is a viable theory of cognition, the standard output of an error region should be lower when predictions align with subjective sensory reality. Which term, the top-down prediction or the bottom-up sensory flow, is the more negative or positive in that interaction might vary among neural regions and still capture the essence of predictive coding. In general, inhibition is everywhere in the brain, so the opportunities for one set of signals to cancel out another are ubiquitous.

The point of this example, and of figure 6.9, is hardly to provide the structural and functional neural architecture of action, but merely to sketch a general heterarchy of feedback loops that should apply equally well to neural firing patterns that represent abstract concepts (e.g., the approach-to-hurdle stride sequence) as to those signaling an over-rotated joint. And each can, through a proper model, M_i, translate into firing patterns at another level that represent a different abstraction (e.g., stride or landing angle) in a different coordinate system.

In addition, the modularity of this type of system would seem to facilitate evolution in the same way that Rodney Brooks' subsumption architecture (1999) for robotic control supports a gradual enhancement of physical dexterity via behavior-unit stacking. In figure 6.9, the horizontal lines carve up the system in accordance with predictive coding, where the errors of one level feed into the next layer, which uses those errors to help compute the control value, which it then sends back down in the form of a prediction and an error gain (as determined by the transformational model, M_i). The error gain manifests *attention* in the predictive-coding framework, as emphasized by Andy Clark (2016, 59):

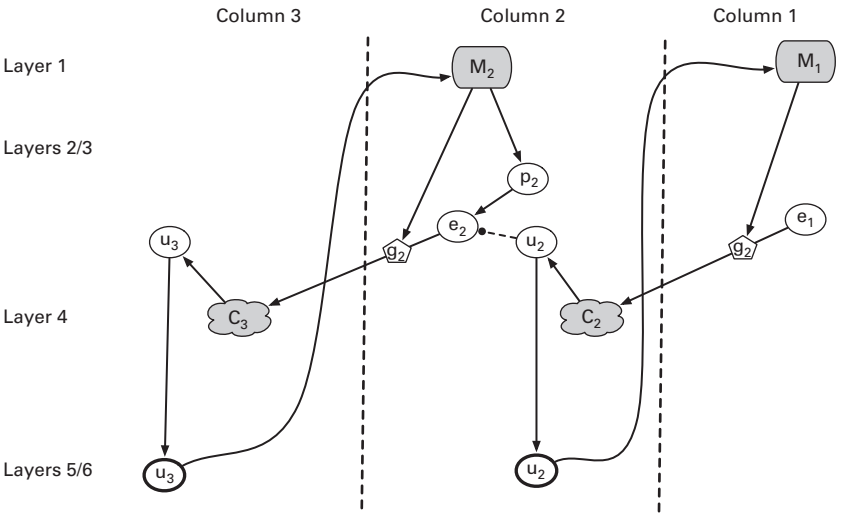


Figure 6.11
Hypothesized overlay of the predictive-coding topology of figure 6.9 onto adjacent cortical columns, separated by dotted vertical lines, with higher-level columns to the left. Transformational models (M_i) map top-down signals to predictions and gains via the highly adaptable synapses of layer 1. Layer 4 predominantly receives strong bottom-up signals, while layers 2 and 3 house a wide variety of excitatory and inhibitory neurons that could implement the basic PID controller (figure 6.9, inset). Layers 5 and 6 may amplify (thick circles) the control signal (u_i) from layer 2/3 before sending it to a lower level.

Attention, if these stories are on track, names the means or process by which an organism increases the gain (the weighting, hence the forward-flowing impact) on those prediction error units estimated to provide the most reliable sensory information relative to some current task, threat or opportunity. . . . The general idea is thus that patterns of neuronal activation (among the so-called "representation units") encode systematic guesses concerning task-relevant states of the world, while changes in the gain (i.e., changes in the weighting or "volume") on the associated error units reflect the brain's best estimate of the relative precision of the top-down "guessing" and the bottom-up sensory information.

This model maps relatively neatly onto the neural architecture of cortical columns, as shown in figure 6.11. As predictive-coding dictates, the main bottom-up signal is prediction error, and cortical models typically channel such information from layer 2/3 of the lower column to layer 4 of the higher column (Mountcastle 1998; Kandel, Schwartz, and Jessell 2000). Within layer 4, and particularly layers 2 and 3, these inputs can stimulate considerable computation that can also incorporate top-down predictive signals. Thus, layers 2–4 are likely candidates for the neural circuitry of the PID controller, including the computation of the layer's error signal. The control variable, u_i, is needed for both error computation and top-down signaling, so its value may be transferred to the large pyramidals in layers 5 and 6, which can strengthen the signal before sending it to layer 1 of a lower column.

It appears that predictive coding can piggyback off of the modular nature of cortical columns in order to satisfy a few basic requirements for evolvability: modular components with weak linkage between them. To scale up a brain composed of many such modules, a genetic mutation or two could probably generate extra columns, which could then hook to earlier columns using the same, well-tuned, developmental recipe for exploratory axonal growth. Thus, the extra columns would begin life with bottom-up synapses entering layer 4,

top-down axons exiting layers 5 and 6, and top-down synapses entering in the dendritic mat of layer 1.

At each incremental stage of evolution, the uppermost cortical columns house controllers whose output values serve as the basis for both a downward prediction and a local error term. The control value and error provide *hooks* for any newly evolved higher layer to latch onto by creating a precision weight (gain, g_i) on the error and an expectation (p_i) for the control, with both determined by a model, whose synaptic weights presumably adapt as part of the learning process. Thus, each new layer arises as a meta-level above the last: it functions to modulate the gain of and predict the value of the lower layer's main products, e_i and u_i, using an even coarser level of abstraction (relative to the subjective sensory reality).

Three main tools of facilitated variation would thus seem to support the gradual evolution of a layered (hierarchical or heterarchical) predictive brain: modular core components (the cortical columns), weak linkage (standardized synapses such as those containing many NMDA receptors in layer 1), and exploratory growth (during both neural migration and axonal extension). The hardest part of adding additional functional modules would appear to be the crafting of each new transformational model, M_i, since these would have microcircuitry and synaptic strengths tailored to the representations of the juxtaposed columns. These might have varying time windows of flexibility (aka critical periods), with lower-level models becoming fixed early in life, while higher-level models would maintain plasticity longer or indefinitely, as the neuroscience literature indicates (Sanes, Reh, and Harris 2006; Bear, Conners, and Paradiso 2001).

6.2.6 Oscillations and Learning

Assuming that the modular representation of cortical columns facilitates variation and thus enables evolution to scale up the brain by adding extra columns, which a well-refined developmental process (replete with exploratory growth) can easily interconnect, the final adaptive chore in building a predictive-coding network is the fine-tuning of synapses, a process that surely requires experience in the world. The architecture of cortical columns combined with two of the brain's most fundamental activities, oscillation and spike-time-dependent plasticity (STDP), provide interesting hints as to how this might happen.

In an attempt to explain learning of a hierarchy of abstractions, where higher-level abstractions predict those at lower levels, this section delves a bit deeper into structural and behavioral details of cortical columns. As shown in figure 6.12, the signal flow between adjacent cortical columns and within individual columns is not symmetric. There are well-established directional patterns that, although not flawlessly copied (as in an engineering device), do appear with surprising regularity across both brain regions and primate (and even many mammalian) species (Mountcastle 1998; Thomson and Bannister 2003). As described earlier, the main bottom-up flow goes from layers 2/3 in column C to layer 4 in column C + 1, then on to layers 2/3 in C + 1 and to layer 4 of column C + 2. In addition, layer 3 sends considerable excitation to layer 5 (and some to layer 6). Layers 5/6 feedback to layer 4 while also initiating the column's top-down response: those in column C send excitatory signals to the layer-1 dendritic mat of column C − 1, which has a strong modulatory effect on the activity of the entire column, but particularly layers 2/3. In short, the bottom-up feedforward path runs through layers 2–4, with the layer 3 axons of column C synapsing

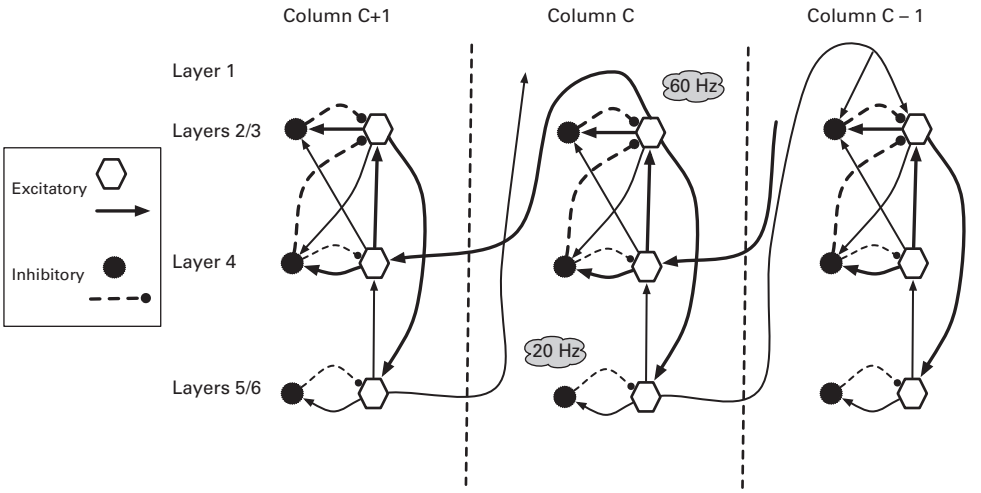


Figure 6.12
Main connections in and between cortical columns, based on diagrams and descriptions in (Bastos et al. 2012; Thomson and Bannister 2003). Each shape (hexagon or ragged circle) denotes a population of thousands of neurons; thicker (thinner) lines denote denser (sparser) connections. Many intercolumnar links are missing or partial to avoid clutter.

proximally onto column C + 1's layer-4 neurons, thus driving their activity in a relatively dominant manner. Conversely, the top-down feedback route involves the lower layers, 5 and 6, whose synapses into column C − 1 tend to be distal, and thus less dominant and more modulatory.

Several additional details in figure 6.12 deserve special attention. First, although layer 4 does house inhibitory interneurons, their effects seem minor compared to the excitatory contributions from layer 3 of the downstream column, and from its own column's layer 3 (via layer 6—shown as layer 5/6 in the diagram). Layer 4 receives no significant inhibition from any other layer. Layer 2/3 neurons display an opposite connectivity pattern: they only receive strong excitatory input from one area, their own layer 4, but exhibit dense intra- and interlayer inhibition. All of this indicates that layer 4 in column C would tend to amplify signals from layers 2/3 in column C − 1, whereas layers 2/3 in column C would sparsify the patterns feed in from layer 4, with that conversion modulated by top-down inputs from layers 5/6 of column C + 1.

Another important, though often overlooked, characteristic of cortical columns involves interlayer differences in oscillation frequencies. As emphasized by Bastos and colleagues (2012), neurons in layers 2/3 have resonant frequencies in the gamma range, with dominant values around 50–70 Hz, whereas layers 5/6 tend to oscillate more slowly, in the beta range: 20–40 Hz. The lower layers appear to integrate signals from layer 3 (over several cycles) before firing and sending their top-down messages. The higher frequencies in layers 2/3 match the inhibitor-prevalent cytoarchitectural profile, since GABA receptors, the main channels for inhibition, have typical decay times in the 10–25 millisecond range, thus inducing oscillations in the 40–100 Hz range (Buzsaki 2006).

This beta-gamma distinction may exemplify the *callup mechanism* hypothesized by Buzsaki (2006, 2019), wherein signaling between two areas begins with the eventual

receiver (Re) sending a low-frequency oscillation as a *request* to the sender (Se), which then resonates at a higher frequency. Thus, Re receives multiple waves of input from Se during each of its lower-frequency cycles: fast waves riding atop slow waves. Thus, in one of its normal cycles, Re integrates several signals from Se to create a compressed, that is, more abstract, spike train. Note how this also dovetails nicely with Buzsaki's inside-out view of the brain, since the higher (inner) areas (columns) initiate message-passing by calling up information from lower (outer) columns.

Another piece to this puzzle involves the competitive nature of neural layers containing considerable inhibition. When one neural assembly (A) activates, it can strongly block another assembly (B). However, when that depression wanes, the neurons in A may have habituated and/or those in B will exhibit a form of postinhibitory rebound (described earlier), thus giving B the upper leg in the next round of competition. Also, as discussed earlier, active cell assemblies strongly inhibit neighboring neurons and themselves for 100–150 milliseconds, thus giving neurons in other areas a chance to dominate. In this way, a high-frequency oscillation can produce a series of activity patterns, instead of many repetitions of the same pattern. The continuous strengthening and decay of synapses due to recent use and disuse entails that any widespread stimulation of an area would tend to bias this competitive, sequential pattern generation to a replay of patterns that were recently active. Thus, if column C − 1 had recently produced patterns X, Y, Z, the signal pattern Q coming down from column C could initiate gamma waves that reactivated each of X, Y, and Z in quick succession, that is, with a period of 10–25 milliseconds.

As shown in figure 6.13, this fast replay brings the temporal lag between patterns down to within the window necessary for spike-time-dependent plasticity (STDP): 0–50 milliseconds (Fregnac 2003). Hence, even if the durations and temporal gaps in X-Y-Z are much longer (i.e., seconds or more), the gamma-driven replay compresses the series into a range where Hebbian learning can link the patterns together such that, in the future, they can stimulate one another. Thus, a single signal could initiate the entire sequence: it becomes self-completing within column C − 1 when given a small push. Column C, the receiver, would learn the triggering pattern of X, Y, and Z: an integration of their signals, all of which arrive on the same active (rising) phase of its slower oscillation (driven by layers 5/6 but also affecting layer 4). This integrated signal, Q (recorded as an activation pattern in layer 4 of column C, then sparsified to pattern Q_3 in layers 2/3, and finally integrated into pattern Q_6 in layers 5/6), serves as the abstraction of column C − 1 in column C. Due to the omnipresence of Q_6 during each of X, Y, and Z, Hebbian synaptic tuning in the dendritic mat of column C − 1 should ensure that the future presence of Q (Q_6) can trigger X-Y-Z.

As part of the same or a similar task, column C − 1 may also execute series X', Y', Z', which abstracts to state R (and R_3 and R_6) in column C; and yet another series may abstract to pattern S. Then, the sequence Q_3-R_3-S_3 in layers 2/3 of column C could, via beta-gamma oscillations, get abstracted into another pattern (K) in column C+1, and so on. Building such an abstraction hierarchy using only a few columns could prove difficult, but given that the primate brain contains millions of cortical columns, it is not hard to imagine large groups of columns organized into clusters, all of which (more or less) feed forward into another such cluster, and then on to another cluster for maybe a few hundred or thousand iterations. Then, if each cluster employed this basic beta-gamma callup procedure, a gradual chunking and abstraction process would seem possible.

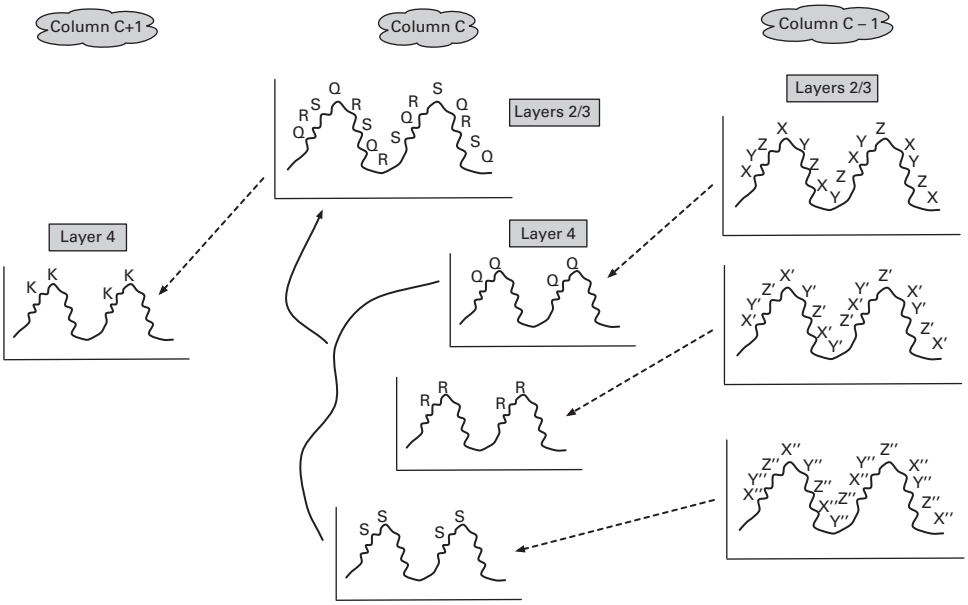


Figure 6.13
Using beta-gamma replay to chunk and abstract activation patterns. Sequences X-Y-Z, X'-Y'-Z' and X"-Y"-Z" play out in column C − 1 at different times. Beta stimulation from column C (layers 5/6) invokes beta-gamma waves in layers 2/3 of column C − 1, which replays a recent sequence, for example, X-Y-Z. In layer 4 of column C, the beta-gamma wave is processed at the slower beta frequency, thus integrating several signals before expressing a firing pattern, for example, Q, for a few gamma cycles at the top of each beta cycle. Other series in column C − 1 yield abstractions R and S in column C. Later, when layers 2/3 of column C receive beta stimulation from column C + 1, they produce a beta-gamma oscillation over patterns Q, R, and S, which, in turn, become abstracted to pattern K in layer 4 of column C + 1. To avoid notational clutter, this diagram denotes an activation pattern with a single letter, such as Q, regardless of the layer in which it appears, with the understanding that the same abstraction will invoke different patterns in each layer of a column.

This general mechanism mirrors concepts such as fixed action patterns (Llinas 2001) and hierarchical models of the brain (Ballard 2015), wherein higher brain regions send simple enabling signals to lower layers, which then perform intricate series of operations, without additional guidance. Sensory inputs predicted by the top-down signals are simply cancelled out, thus not tasking the system with unnecessary information transmission. Once again, this illustrates the inside-out view of cerebral control (Llinas 2001; Buzsaki 2019), since the high-level expectations run the show, only to be interrupted by unexpected inputs.

Buzsaki (2006, 2019) also champions beta-gamma callup as the means by which the hippocampus and neocortex exchange patterns. During exploratory activity, HC *calls up* information from the cortex by sending theta waves (4–7 Hz) to the cortex, which responds with (the higher-frequency) gamma waves (bundling many events into the active (rising) phase of one theta wave). The hippocampus then processes and (temporarily) stores these aggregates via STDP. Then, during non-REM sleep, the sleep phase with the greatest decoupling of brain and body, the process reverses and the cortex calls up HC with delta waves (0.5–3 Hz), and HC responds with ultra-fast (100–200 Hz) ripple waves, which pack several dozen patterns into the active cycle of one delta wave, thus enabling more permanent storage of the associations among these patterns in the cortex.

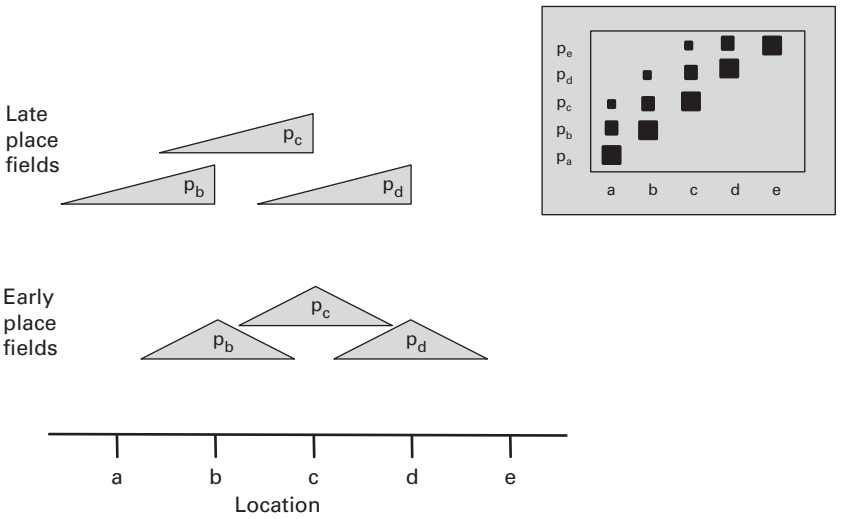


Figure 6.14
Phase precession in hippocampal place cells, $p_a - p_e$., with a–e on the lower axis denoting locations along a path. (Left) Triangles represent the place fields for their corresponding place cell, with height coarsely reflecting firing level (when the rodent is located at the spot on the horizontal axis). During early trials, the place fields are symmetric, with strong activation at the target location and gradual tapering of activity both before and afterward (in time / space), assuming the rodent moves left to right. After many trials, asymmetric fields form in which the cell fires several locations prior to the target but few (or no) locations afterward. (Right, inset) Predicted futures for well-trained place cells. Columns convey place-cell activity when rodent resides at location on the x axis, for example, when at location c, place cells p_c, p_d, and p_e are strongly, moderately, and slightly active, respectively, as reflected in the sizes of the dark rectangles.

Phase precession in the hippocampus

Prediction, oscillations, and the hippocampus come together in another, related, manner involving the phenomena of *phase precession*, as revealed by Mehta (2001). Many pyramidals in CA3 and CA1 have been identified as *place cells* (Andersen et al. 2007), meaning that their firing rates are strongly correlated with the locations of the animal: when at or near location X, the place cell(s) for X will show high activity; otherwise, they will fire infrequently. (See chapter 3 for more details on place cells.)

Mehta's neurophysiological studies show that as a rodent becomes familiar with a particular path, for example, $a \rightarrow b \rightarrow c \rightarrow d$, it will establish place cells p_a, p_b, and so on for each location. These neurons gradually become *predictive* in the sense that p_d will begin firing well before the animal reaches location d. In other words, the receptive fields (aka place fields) for the place cells extend backward in time and space, as shown in figure 6.14.

This has a relatively straightforward explanation in terms of spike-timing-dependent plasticity (STDP) and oscillations. As a rat progresses along the path, the corresponding place cells will fire, in sequence, with each place cell showing activity a bit prior to, during, and a bit after the visitation of its corresponding location, that is, the place field is symmetric about the location. The time delay between visiting locations a and b may be significantly greater than the normal window for STDP (40–50 msec), so connections between p_a and p_b cannot strengthen as a result of normal, sensory-induced activity. However, these place-cell

firing sequences are often repeated after (and before) the rodent's search process at a much higher (gamma) frequency (40–100 Hz), thus bringing the sequential firing within the STDP window. Hence, p_a may fire a mere 10 msec before p_b, thus leading to a strengthening of the $p_a \rightarrow p_b$ synapse (and similarly the $p_b \rightarrow p_c$ and $p_c \rightarrow p_d$ synapses). In addition, this should also enhance $p_a \rightarrow p_c$ and $p_a \rightarrow p_d$ although to a progressively lesser degree due to the longer (albeit compressed) delays.

Conversely, the gamma replay, by invoking p_b activity immediately after p_a, will promote depression of any $p_b \rightarrow p_a$ synapses via the other half of the STDP curve, that is, when post-synaptic firing precedes presynaptic firing. Thus, the succeeding place cells lose the ability to stimulate their predecessors, while the predecessors have gained the ability to excite their successors.

The net result, after many trips along the path and many rounds of gamma replay, is a collection of *asymmetric* place fields in the sense that p_x becomes active for many of x's predecessor locations but for few or none of its successors. As a whole, the CA3–CA1 region becomes predictive, because when sensory input indicates location x, the place cells for several successors to x will fire. When the place *fields* skew backward in time/space, the place *cells* fire predictively, indicating the locations that the animal will soon visit, with the highest activity pinned to the most immediate successors.

In Mehta's work, the strengthened and weakened synapses are those from CA3 to CA1 such that both regions initially house symmetric place cells for the same location, x, but the place fields in CA1 become asymmetric and predictive via STDP. Hence, CA1 activity represents a prediction based on excitatory connections from CA3. This can only reconcile with our earlier description (in chapter 3) of CA1 as a comparator if the predictive firings in CA1 trigger inhibition such that the anticipation is effectively subtracted from sensory reality arriving from EC.

In this scenario, note that the overall timing relationships could adapt to essentially satisfy predictive coding. For example, the sensory reality encoded by EC at time t travels to DG, then through CA3, and then to CA1, which thereby encodes a prediction for some future time $t + \triangle t$. Assuming that this chain of processing takes τ msec, then the prediction will appear in CA1 around the same time that the direct line from EC conveys the updated reality signal, for time $t + \tau$. If the hippocampal loop can calibrate such that $\tau \approx \triangle t$, then the prediction and reality signals in CA1 should correspond to the same time point, thus making the comparison and resulting prediction error more appropriate.

Evolution seems to have invented the hippocampus quite early in the game, with it always playing a very fundamental role in memory formation and receiving the most abstract information that a particular brain had to offer. Only in mammals are those HC entry signals distant, abstract remnants of the original sensory data, but in amphibians, reptiles, birds, and others, the HC can still play the roles of (a) generator of high-level, temporally extensive, prediction error; (b) detector of novelty; and (c) initiator of widespread synaptic modification. In fact, hippocampal structures appear to have arisen independently (an example of convergent evolution) in birds and mammals (Striedter 2005), which indicates a genetic tendency to produce and a selective pressure to possess exactly this type of circuitry. The HC's well-documented role in navigation (McNaughton et al. 2006; Burgess and O'Keefe 2003;

Andersen et al. 2007; Buzsaki 2006) further underscores its importance atop the functional neural hierarchy.

6.3 A Brief Evolutionary History of the Predictive Brain

Regardless of whether the earliest organisms acted with or without sensing, motricity is a crucial factor in any discussions of prediction and its origins. As discussed earlier, a moving organism gains a survival advantage from prediction, even if only of the procedural variant: it behaves *as if* it can foresee its immediate future, by, for example, various gradient-driven activities. Even bacteria have this ability, so primitive multicellular organisms probably developed similar talents as their motoric sophistry increased.

Oscillations also enter the scene quite early, since rhythmic motion among a suite of myocytes enables a structuring of the local environment (in terms of features such as controllable / predictable flow patterns) and bodily coordination. However, this typically requires inhibition of one form or another: chemical, electrical, gap junctions, interneurons, and so on. As shown above, when it comes to the neural signaling level, the most common excitatory and inhibitory neurotransmitters have nearly identical chemical structure. Thus, inhibitors probably evolved quite easily once excitors had arisen, or vice versa.

Inhibitory interneurons and the inherent delays of chained neurons also support the calculation of simple temporal gradients; and these combine with the signal-integrating features of both individual soma and recurrent subnetworks to provide the rudiments of control and prediction. So as neural circuits added even a modicum of computational distance between sensors and motors, the possibilities for prediction and control would have emerged naturally. In addition, the advent of synaptic connections between neurons paved the way for learning via pre- and post-synaptic modification.

Inhibition also facilitates comparator neurons that receive excitatory signals from one source (e.g., sensory reality) and inhibition from another (e.g., a goal or prediction). The comparator's output then represents a deviation or error, which then promotes changes to expectations, actions, or even the goals themselves. Early on, in more primitive nervous systems, any activation patterns that somehow represented imagined states (of the brain-body-world coupling) probably served as goals, to be compared to reality, with the resulting error governing behaviors that reduced the goal-reality gap. The neural machinery probably existed to support functions similar to those of a conventional PID controller.

The next evolutionary step, to something akin to an adaptive controller (as discussed in chapter 3 in relation to the cerebellum), would require a comparison of predictions to relatively raw sensory signals, with that difference then serving as a perception. Then, a second comparison would involve the goal and perception, with the resulting error governing overt behavior. For example, when running to catch its prey, the predator's raw reality signal includes considerable vibration due to its own body movement, which the predator's brain can predict. That expectation constitutes noise, which, when removed from the raw signal, yields the proper percept (of a fast but otherwise steady prey trajectory). And that percept, not the noisy raw signal, is the more reliable representation to compare to the goal state (of predator meeting prey) in the course of choosing an action: in reaching out a limb to grab prey, the predator benefits from steady, non-jerking movements (with steadiness

defined with respect to the ecocentric reference frame), despite the gyrations of the raw signals.

The basic adaptive control circuitry constitutes a module that could be duplicated and differentiated many times, quite possibly mirroring the proliferation of cortical columns in higher mammals or even the expansions of the cerebellum. This creates a hierarchy (or heterarchy) replete with predictions, reality signals, and comparators. At higher levels, the distinction between expectations and goals might blur: an activity pattern within a given module might constitute either, depending on the firing intensity of the neurons and the attention given to them, that is, the gains on their outgoing connections. Thus, goals would be patterns with higher activations and downstream influences, and thus be more likely to affect action; goals are the *envisioned states* that the brain will more stubbornly seek to achieve, while predictions have less causal force and may function as more *wishful thinking* (for positive expectations) and *traps to avoid* (for negative predictions).

Levels of this heterarchy that are somewhat distant from the peripheral sensory and motor apparatus would receive more delayed and multimodal signals, thus precluding them from low-level monitoring or control. Instead, their activation patterns would constitute abstract representations that encompass larger swathes of space and time while still playing a predictive role. The hippocampus seems to occupy the top, most abstract, tier of this organization.

Along each of these branches of the evolutionary tree, the constraints of energy conservation, firing-rate stabilization, and signaling efficiency would have incentivized proliferent inhibition and predictive coding. During development, the need to strike proper balances between competition and cooperation, and between hypo- and hyperactive firing rates, would provide selective advantages to networks with the right mixture of excitation and inhibition. Neural Darwinism (Edelman 1987) and its stipulated *survival of the best networkers* would encourage an individual neuron (N) to synchronize firing with local neurons (i.e., cooperate) to increase the chances of promoting downstream firing (more cooperation), and thus solidifying efferent synapses (and fortifying N's immediate network) via long-term potentiation. Conversely, it would also motivate N to inhibit some neural groups, to reduce competition for targets. Hence, an axon might invade some regions and synapse on excitors while breaching others to stimulate the local population of inhibitors. This *invasion with inhibitory intent*, a blatant act of competitive sabotage, would fortuitously form the key negative influence of a higher group's prediction on a lower group's comparator. Thus, predictive coding networks could emerge from the dueling forces of competition and cooperation driven by the basic needs of all neurons to survive and maintain relatively stable firing rates.

In parallel with the elevation of predictive topologies, oscillation gradually becomes encephalized. From a crucial factor in myocyte coordination, it ascends the heterarchy with each newly evolved level, providing a tool for synchronization of neural firing (and the ensuing cooperative and competitive benefits) while also promoting intraregional synaptic change and interregional information transfer, such as between the hippocampus and neocortex. This very simple mechanism has extremely far-reaching consequences in the emergence of intelligence.

Finally, facilitating each of these variations (i.e., each evolutionary branch) are the fundamental components of Kirschner and Gerhart's (2005) theory: weak linkage, modularity,

duplication and differentiation, and the grow-to-fit dynamics of neural development. Each of these features, clearly evident in the brains of most organisms, provides a robustness and flexibility to life such that random genetic mutations have more neutral and positive consequences than any statistical random-walk analysis would predict. There does seem to be a certain inevitability to the emergence of predictive machinery when one buys into the theory of facilitated variation and realizes how neural systems and their development so perfectly instantiate each of its core tenets. From a billion-year evolutionary perspective, the capacity to generate expectations is certainly expected.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Times New Roman by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Downing, Keith L., author.
Title: Gradient expectations : structure, origins, and synthesis of predictive neural networks / Keith L. Downing.
Description: [Cambridge, Massachusetts] : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022037237 (print) | LCCN 2022037238 (ebook) | ISBN 9780262545617 (paperback) | ISBN 9780262374682 (epub) | ISBN 9780262374675 (pdf)
Subjects: LCSH: Deep learning (Machine learning) | Neural networks (Computer science) | Conjugate gradient methods.
Classification: LCC Q325.73 .D88 2023 (print) | LCC Q325.73 (ebook) | DDC 006.3/2—dc23/eng20230302
LC record available at https://lccn.loc.gov/2022037237
LC ebook record available at https://lccn.loc.gov/2022037238