

# 7

## Convergent Ways of Seeing

Is mind a radically contingent accident of life on Earth, or is it a replicable, law-like feature of any living world? This is the central question taken up in part II. The main conclusion will be that the RCT (or what survives of it from part I) does not preclude the evolution of complex cognition or even conscious minds. The rhetorical journey that will lead to this conclusion begins here, with the emergence of image-forming sensory modalities.

### 1. Image-Forming Sensory Modalities

This chapter will argue that due to the invariant physical structure of the universe, the emergence of image-forming sensory modalities (ISMs) is both robustly replicable *qua function* and highly constrained *qua form*. These constraints support detailed, law-like generalizations that transcend the peculiar features of body plans, whether on Earth or beyond. Iterated ISMs are not therefore Gouldian repetitions, despite the fact that they make use of some conserved genes and developmental machinery as might be expected for any evolutionary system that builds on itself over time. The next three chapters will extend this argument to the emergence of minds.

The chief function of sensory modalities is to tie informational inputs from a particular energy source in the environment to behavioral outputs that enhance fitness. Energy inputs are recorded by cellular receptors and then transduced into action potentials or other types of signaling cascades that culminate in behavior. Among the various forms of physical energy in the universe, only waveform energy is capable of providing the rich, real-time information necessary for organisms to construct detailed, three-dimensional “scenes” of their surrounding world—an ability that I will call “seeing.”

Life has found a way to form rich scenes using all known forms of waveform energy from which ecologically useful information about the distribution

of objects and their properties can be gleaned. Only a few types of waveform energy carry this information: light (within certain spectral bounds), sound, and electromagnetic fields. Other energy stimuli, such as chemical gradients that serve as the basis of olfaction, do not carry information that can underwrite a three-dimensional perceptual world that supports sophisticated locomotion, navigation, predation, and other cognition-mediated behaviors.

Olfaction can be informationally rich: just think of all that a dog can divine about a prey or conspecific from a single scent trail. But olfaction is informationally impoverished when it comes to forming real-time images of objects that comprise scenes. Scents remain even when the object that left them is long gone. Chemosensory systems are robustly convergent,<sup>1</sup> but they do not allow organisms to “see” in anything representationally or phenomenologically analogous to the object-based scene perception that characterizes the visual sense.<sup>2</sup> Chemosensory, tactile, and other nonimagistic senses do not support the spontaneous construction of a three-dimensional model of the world with the subject at its center—a model that updates in real time with a dynamically changing world, recording changing distributions of waveform energy and measuring them against internal states of the body and its position in space.

The cognitive side of this equation will be examined in the next three chapters. The focus here is on the sensory modalities themselves and how they tap into universal energy forms to enable organisms to see. Seeing, it will be argued later, is the evolutionary gateway to more complex forms of cognition and behavior, and even to consciousness itself. In what follows, the evolution of three distinct modes of image-formation will be explored—vision, echolocation, and electrolocation—each of which has been subject to iteration.

## 2. Vision

Of the three broad ISM types, vision is the most well-known, the most subject to iteration, the widest in scope given the ubiquity of the light stimulus and the availability of optical substrates, and the most independently derived, as indicated by the unparalleled phylogenetic breadth of its iterations. Light contains ecologically useful information in almost all habitats on Earth and is used by many organisms to fuel their metabolism. It is not surprising, therefore, that the ability to sense and respond to light is ubiquitous in the living world.

### 2.1 Phototaxis

All domains of life track sunlight, using signal transduction networks to link the perception of light with adaptive movement.<sup>3</sup> Phototrophic organisms, or those that harvest sunlight for metabolic energy, must be capable of optimally

positioning their photosynthetic apparatus in relation to the sun, taking into account both the intensity and spectral quality of light as well as current physiological states of the organism. Nonphototrophic organisms may also benefit for a variety of ecological reasons from positioning themselves at some optimum distance with respect to a light source, such as to avoid damage from ultraviolet radiation, to situate themselves within hosts or symbiotic partners, or to orient toward the shoreline or ocean surface.

The diverse set of simple motility systems that are guided by the detection of light fall under the umbrella term “phototaxis.” The three main components of phototactic systems are light detectors (which are molecularly diverse), signal transduction networks (involving proton gradients or electron transport chains), and motility mechanisms (such as the convergent flagella of bacteria and archaella of archaea). At the minuscule scale of prokaryotes, phototactic systems are generally (though not exclusively) limited to sensing and responding to light gradients, rather than to light vectors. Gradients, whether made up of light or chemicals, are rarely steep enough for prokaryotes to detect differing chemical concentrations at two ends of the cell, which are often no more than a couple of microns apart. This limits most prokaryotes to “temporal” rather than “spatial” perception of gradients—that is, to “two-dimensional” perception guided by whether concentrations are increasing or decreasing over time.<sup>4</sup>

Some of the mechanisms underlying phototaxis in prokaryotes are well established. One is the “biased random walk,” in which the cell randomly tumbles or changes direction by flagellar switching, with the probability of flagellar switch reduced during movement up or down a steep light gradient depending on whether the response is photophilic or photophobic. True three-dimensional phototactic navigation—that is, orientation along a proper light vector—evolved at least eight times independently in eukaryotes, aided by the independent evolution of eyespots (known as “stigmata”).<sup>5</sup>

In green algae, these eyespots are composed of hexagonal arrays of lipid globules—pigmented structures that provide directional shading and mirror-like focusing (refraction) of light onto adjacent photoreceptors during axial rotation; these photoreceptors trigger signaling transduction systems that link to ciliary beating, which then produce movements toward or away from the light source. All eukaryotes that independently evolved the ability to orient and move toward a light vector have done so using this same structural configuration. Some of the pigments used in stigmata, such as rhodopsins, have evolved independently as well,<sup>6</sup> increasing the “hierarchical depth” of phototactic convergence.

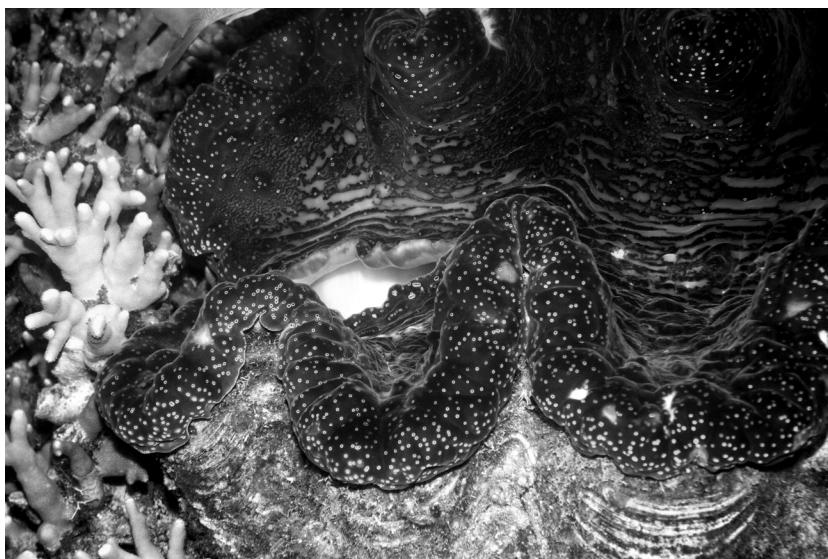
## 2.2 True Spatial Vision

The focus here is not basic phototactic systems or lensless pigmented eye cups, but rather focus-capable eyes which, in conjunction with sophisticated information processing capabilities, are capable of producing a “visual scene”: a panoramic, dynamically updating representation of spatiotemporally distributed objects in the surrounding environment. The cognitive generators of the visual scene will be examined in the next chapter. Here, the discussion will focus on the basic principles of eye construction and the informational resources they make available to organisms that can process them.

“Spatial resolution” refers to a sensory system’s ability to perceive adjacent points in space as separate points. The smaller the distance or angle between these points, the more acute the spatial resolution. Information gleaned from true spatial vision is orders of magnitude greater than that achieved in the transition from nondirectional to directional phototaxis. High-resolution spatial vision supports the effectively instantaneous detection of fine differences in light stimulus “gradients” across a sweeping spatial expanse. It does this aided by lenses, stacked photoreceptors that simultaneously monitor different elements of the scene, and screening pigments that provide directional and spectral sensitivities.

Although there are many types of eye that support some degree of spatial resolution, only variations on camera and compound eye configurations have achieved genuine panoramic vision with high angular detail that can discern and track objects over space and time. Camera-type eyes, which consist of a single optical unit with refractive elements that focus light onto a retina, have evolved in distant animal phyla, including vertebrates, mollusks, arthropods, annelids, cnidarians, and even single-cell eukaryotes. Compound eyes, which consist of numerous optical units (“ommatidia”), each with their own refractive components (lens and cornea) that focus light from a small region of space onto their corresponding photoreceptor cells, are less phylogenetically disparate than camera eyes but far more prevalent due mainly to the unparalleled success of the arthropods. The highest resolution compound eye—that of the dragonfly—consists of an array of over 30,000 individual lenses. Although simple lens eyes have evolved from lensless pigmented eye pits in numerous animal phyla, only four major animal groups (spread over three different phyla and two superphyla) have evolved high-resolution vision that supports active, dynamic life-styles: crustaceans/insects, spiders, vertebrates, and cephalopod mollusks.

The vertebrate camera eye and the paired compound eyes situated in the head of arthropods probably arose once in each respective phylum and have been conserved in these groups ever since. Homology of the arthropod eye remains contested, however; either many groups of arthropods lost compound



**Figure 7.1**

Giant clam (*Tridacna gigas*) with pinhole eyes in blue-green circles distributed throughout the mantle. Photo taken by author at Lizard Island, Great Barrier Reef, Australia.

eyes (including chelicerates, myriapods, and ancestral ostracod crustaceans), or similar compound eyes evolved in disparate arthropod groups.<sup>7</sup> What we know is that the arthropod compound eye arose first, in the base of the Cambrian, and the single-chambered eye evolved in craniate vertebrates not long thereafter, with cephalopod mollusks eventually following suit.

Probably the most striking example of convergence on the camera eye is that between coleoid cephalopod mollusks and vertebrates (see chapter 6), which have comparable visual acuities.<sup>8</sup> Vertebrates and mollusks in all likelihood share an eyeless, brainless, and relatively immobile common ancestor that lived in the Precambrian more than 550 million years ago (see chapter 9 for an extended discussion of this inference). Precisely when the cephalopod camera eye evolved is not known with certainty, though it is likely to have arisen not too long after the origin of the vertebrate eye, perhaps in the common ancestor of ammonites and belemnites in evolutionary competition with primitive fishes. Simpler eyes have evolved repeatedly in the mantles of bivalve mollusks, such as the pinhole eyes of giant clams which number in the thousands and serve as coarse-grained intruder alarms (see figure 7.1).

Spiders and other chelicerates are also an interesting case. Despite their close relation to insects and crustaceans, chelicerates generally have single-chambered

eyes like vertebrates and cephalopods. The highest level of spatial resolution achieved by any chelicerate is the principal eyes of the jumping spider (salticidae), an active diurnal hunter that exhibits some of the most cognitively sophisticated behaviors observed in arthropods (discussed in detail in chapter 10). The chelicerate camera eye may have evolved through a process of simplification on the ancestral arthropod compound eye, or it may have arisen de novo. Anyone who has scoured the East Coast beaches of the United States knows that horseshoe crabs have prominent compound eyes, like those of insects; and horseshoe crabs, as it turns out, are basal chelicerates—the clade to which spiders and other arachnids belong. It is thus possible that chelicerates tweaked the existing compound eye design once or many times to produce a single-chambered eye. Convergent reduction is a far cry from convergent origin, and thus the macroscopic eye homology between insects/crustaceans and chelicerates could undercut the “independence” (*sensu* chapter 6) of this particular eye iteration.

Complex eyes are not limited to animals or even to multicellular organisms. The remarkable camera eye of unicellular Warnowiid dinoflagellates, known as the “ocelloid,” has been fashioned out of endosymbiotic subcellular organelles (see figure 6.4 in chapter 6). For instance, in the green algae *Erythropsidinium*, interlocking plastids (which ordinarily carry out photosynthesis) form the retinal body, and mitochondria have been modified into a cornea-like structure.<sup>9</sup> Although the dinoflagellate eye has a lens and screening pigment, like the camera eyes of cubozoan cnidarians (the close relatives of jellyfish we encountered in chapter 6), the retina is placed too close to the lens to be in focus. As a result, these eyed dinoflagellates cannot direct their attention toward a detailed image or, more importantly, process any image that could be projected (they are, after all, but a single cell!).

What is the fitness advantage of such surprisingly sophisticated optics in brainless organisms like unicellular eukaryotes, cnidarians, and bivalves? The answer seems to be an increase not in resolution but in sensitivity, which suffices for the low-resolution visual tasks that these lineages face, such as coarse-grained orientation and predator evasion. The inability to resolve fine spatial details may actually be beneficial if the goal is to perceive large, stationary structures in the environment.<sup>10</sup> Greater resolution—often at the expense of sensitivity (as discussed later) and coupled to the increased processing power of brains—allowed for the richer forms of information extraction that underwrite the most active lineages of life.

Which then came first, the eye or the neurons necessary to process visual information? Unlike the infamous chicken-or-egg conundrum, this is not a pseudo problem because we have strong reasons to believe that neurons predated true eyes (see chapter 9), even though the evolution of eyes and the

complexification of nervous, proprioceptive, and motor systems is likely to have gone hand in hand. Whereas convergent eyes abound in nature, only a handful of lineages have turned these optical devices into the input sides of sophisticated information processors, triggering runaway coevolutionary feedback between brains, bodies, behavior, and ecology. Activities like the dynamic pursuit of prey and navigation by fine-grained landmarks require vision with high spatial resolution—an angle of acuity of no more than a few degrees.<sup>11</sup> Only the compound eyes of arthropods and the camera eyes of vertebrates, cephalopods, and spiders have achieved this level of angular detail.

### 2.3 Robustness, Invariance, and Evolvability

Modeling work suggests that eyes can evolve very rapidly, geologically speaking—going from a mere patch of photosensitive cells to a high-resolution camera eye in less than half-a-million years, with each step plausibly conferring a fitness benefit by increasing sensitivity and/or resolution of the eye.<sup>12</sup> The evolutionary robustness of eyes was analyzed at length in the previous chapter. The argument there was that the repeated recruitment of deep homologs (like *Pax6*) in disparate eye-bearing groups does not convert eye iterations into Gouldian repetitions or parallelisms because this conserved molecular substrate does not constrain the space of gross eye morphology in specific ways. By contrast, the macromorphological arrangements of visual ISMs are highly constrained by the properties of light: all image-forming eyes must, for example, have lenses that are made of high-refractive transparent material that operate within the laws of refraction. The lenses of organisms will, of course, be constructed out of proteins rather than glass; however, as Land and Nilsson eloquently put it, “it is chemistry rather than physics that distinguishes biology from technology.”<sup>13</sup> This bold statement is applicable not only to vision but to all three types of ISM.

Whether camera or compound, Earth-based or extraterrestrial, any visual organism must contend with a number of inexorable trade-offs imposed by the laws of optics. One such trade-off is between resolution and sensitivity: the more densely packed the photoreceptor “pixels,” the better the spatial resolution; but greater density entails smaller photoreceptors, which reduces the capacity of individual photoreceptors to detect photons, and thus reduces the overall sensitivity of the eye and its ability to discern contrast. Similarly, the more photons an eye collects thanks to a wider pupil, the greater the distorting effects of spherical and chromatic aberration, and thus the blurrier the image. The evolutionary trade-off between resolution and sensitivity has been resolved in different ways, depending on the ecological context. These optical constraints and trade-offs, like the structural configuration of eyes themselves, are not contingent on the peculiar developmental parameters of

lineages but rather are biological universals with which any visual life world must contend.

Vision is by far the most ubiquitous ISM on Earth, but it is not the only one. Two other image-forming sensory apparatuses have evolved, exploiting two additional waveform energies: echolocation (which uses sound) and electrolocation (which harnesses electromagnetism). These “alternative” modes of image-formation, which are explored in the remainder of this chapter, are far less common than vision, probably because they are harder to evolve. One reason they are harder to evolve is that they require high-intensity energy emissions. Active focused emissions, whether in the form of sound or electromagnetic fields, have significant energetic and anatomical requirements over and above those associated with passive reception and transduction. Not only must organisms have metabolic rates that could support persistent high-energy emissions, they must also have the anatomical structures necessary to produce these outputs, such as lungs capable of emitting concentrated bursts of sound, or electricity-generating organs capable of steady electromagnetic discharges.

That said, it is somewhat misleading to characterize vision as a “passive” sense. Constructing detailed visual scenes requires scanning movements of the eyes and adjustments of the perceiver’s position relative to its surrounds so as to move objects in and out of focus (regions of acute retinal resolution) and to view them from different angles. Vision is also cognitively active, in that various properties of the scene must be stitched together by dedicated regions of the brain—cognitive activities that loom just as large, as we shall see, in echolocation and electrolocation. In contrast to alternative ISMs, however, vision is accomplished through the “passive” detection of ambient light that is generated mainly by the sun and nighttime stars and reflected by objects and terrain. A hypothetical equivalent to active emission in the case of vision would be animals in the ocean midnight zone evolving a bioluminescent apparatus that produces a flashlight-like beam of light, in combination with sophisticated optical and neural apparatuses to sense, transduce, and process reflections of that beam. The closest nature has come to this is the loosejaw, a deep-sea fish with an organ that produces red light and vision attuned to the red spectrum, which is invisible to most prey in the midnight zone. This might sound like a fantastical adaptation, but it is precisely what has been achieved by alternative ISMs like echolocation and electrolocation in which active emission capabilities evolved in tandem with detection and processing apparatuses. Let us now turn to these alternative modes of seeing.

### 3. Echolocation

Echolocation is a perceptually rich discrimination and navigation technique accomplished through the use of active biosonar. Echolocation is best regarded—

functionally, representationally, and phenomenologically—not as an advanced form of hearing but as an advanced form of “seeing” (see chapter 8 on the delineation of senses). It is employed as a primary image-forming sensory modality, not as a secondary sense relied upon in circumstances in which primary sensory systems are ineffective. Echoic abilities have evolved independently at least five times in the history of life, including in three orders of mammals and two orders of birds. An even greater number of animals have evolved the capacity to generate sounds and listen to their echoes in order to locate objects or sense terrain in dark environs. For instance, blind subterranean mammals, such as moles, appear to use seismic vibrations produced by striking their heads on tunnel roofs for orientation and navigation.<sup>14</sup> Even blind humans have been shown to successfully employ rudimentary echolocation techniques, recruiting brain regions that are normally associated with the processing of visual information.<sup>15</sup>

Only in bats (microchiropterans) and whales (odontocetes), however, do echoic abilities approach vision in their imagistic resolution. Because ensonified objects tend to have higher acoustic impedance (resistance to sound waves) than air or water, they absorb virtually no sound energy, which is instead reflected and scattered in much the same way that light (electromagnetic radiation) is absorbed and scattered by objects in the environment. Water has a higher acoustic impedance than air, but animal hard parts (such as skeletal structures) and soft organ systems have a higher impedance than water, and thus return echoes that make them “visible” even when cloaked under sediment. Features of the emission echo, or what is known as the “acoustic impulse response,” contain information about the size, shape, velocity, and even the fine-grained textures of ensonified objects. Different reflecting points of a complex, irregular target will yield slightly different range values, creating a profile that is interpreted as shape.

Pioneering work by James Simmons in the 1970s showed that bat acoustic images are so precise that they can distinguish between targets separated by distances well under 1 millimeter in three-dimensional space—an angle of acuity that is comparable to the camera eye of vertebrates and cephalopods.<sup>16</sup> The acoustic impulse response is compared to the emission both in time and frequency domains in order to extract detailed spatial information about the surrounding world. Fine spatial details of objects and surrounds are resolved by detecting minuscule differences in echo return time. Bats can apparently discriminate time differences on the astonishing order of 10 nanoseconds—or 10 billionths of a second.<sup>17</sup>

Such fine-grained time perception is presumably done outside of conscious experience, with detailed acoustic images simply “presented” to echolocating bats and whales much as visual images are spontaneously presented to

organisms through “cognitively encapsulated” visual systems that do most of their imagistic processing outside of conscious awareness. Sound objects are typically experienced as temporally (but not spatially) extended.<sup>18</sup> The range differences in echolocation are so infinitesimal that they leave no time for the experience of sound. Phenomenologically and representationally speaking, echolocation is a thoroughly spatial sense. It is used by bats and dolphins not only to identify objects and avoid collisions with terrain but also for short and middle-range navigation by identifying echoic landmarks and constructing acoustic maps.

### 3.1 Echoic Object Recognition

One of the central research questions in the initial decades of biosonar research was the extent to which echoic animals are forming true, three-dimensional acoustic images of objects. In other words, it was initially unclear whether echolocating bats and whales were “seeing” with sound or simply listening to echoes and adjusting their behavior accordingly. The results of early forced choice discrimination experiments with echolocating dolphins were called into question when it became clear that humans could discriminate between objects as successfully as dolphins on the basis of echoes alone when these were played back at the frequency range of human hearing. There is a world of a difference, cognitively and phenomenologically, between listening to echoes in order to discriminate among objects or track their movement over space and time, and spontaneously seeing three-dimensional objects through the topographic analysis of their acoustic profile in a way that is functionally analogous to vision.

It is now well established that richly echoic animals, such as bats and dolphins, are not discriminating objects on the basis of the echo alone; instead, they are carefully comparing the emission with the echo in order to generate a detailed, three-dimensional image of object shapes, textures, positions, and velocities, and relations to other objects in the environment. Bat echolocation supports activities as precise and nuanced as aerial insect hawking and other prey interception techniques, navigation through cluttered foliage, and plucking small prey items from complexly textured surfaces, all at speed and in complete darkness.<sup>19</sup> Bats must be capable of intercepting flying insects that are only a few millimeters in diameter while avoiding obstacles and interference by other call-emitting bats. These prey items, moreover, do not sit idly by while being captured—many deploy sophisticated antiechoic countermeasures such as jamming techniques or the initiation of erratic flying maneuvers when high-frequency ensonifications are detected. Intercepting prey on the wing in physically and acoustically cluttered environments is a daunting task

that would be impossible if echolocation were not providing an acoustic analog to the integrated, object-structured visual scene.<sup>20</sup>

Experiments in the laboratory have shown that bats can discriminate among objects based on their detailed shapes and textures<sup>21</sup> irrespective of their size,<sup>22</sup> solely on the basis of acoustic images. Acoustic “flow fields” of reflected sound form around the bat’s head as it flies, providing real-time information about the bat’s movement and the orientations of objects and terrain in much the same way that visual flow fields guide the complex flight patterns of birds. Background objects are not mere clutter to be avoided but rather form part of the bat’s topographically rich acoustic scene. In fact, bats form multilayered acoustic maps of their spatial world: lower resolution maps for coarse-grained navigation, which allows them to beeline back to their roost after a meandering foraging translocation (sometimes over 100 kilometers), and higher resolution maps for finer-grained spatial tasks such as cave navigation or the aerial capture of insects.<sup>23</sup>

Evidence for holistic object recognition in echolocating dolphins is also compelling.<sup>24</sup> Perhaps the strongest such evidence comes from “cross-modal recognition” studies in dolphins, whereby an object initially inspected only through echolocation is immediately and globally recognized through vision, and vice versa, with very high rates of success.<sup>25</sup> This strongly suggests, and perhaps decisively shows, that what a dolphin “sees” through echolocation is comparable in crucial ways to what it sees through vision. Experimental design has excluded the confounding possibility of associative learning by requiring the immediacy of discrimination and never exposing the target objects to both senses. Given these controls, the only plausible explanation of cross-modal recognition in dolphins is that they are constructing detailed, three-dimensional object percepts that are functionally comparable across both sensory modalities.

Echolocation in bats and dolphins allows for spontaneous object recognition in much the same way that vision does—probably by comparing observed patterns with three-dimensional “templates,” “prototypes,” “categories,” or other imagistic representations stored in memory. To succeed at shape-based object recognition tasks, an organism must be capable of categorizing objects based on their shapes independently of their size, with shape representations scaled up and down to larger and smaller objects. For example, small and large triangles should be classified as the same type of object and distinguished from other shapes based on their three-dimensional profile.

Such echoic discriminations among real and virtual objects, which have been observed in laboratory experiments with bats and dolphins, cannot be explained by temporal acoustic properties alone. Rather, they implicate a complex form of neural processing that faithfully reconstructs the spatial dimensions

of stimuli, which then appear to the echolocating organism as “bound” three-dimensional representations. The fact that echoic animals recognize objects and track them over space and time shows that the paradigmatic “binding” problem of vision has been solved convergently in an ISM that exploits an entirely different waveform energy, an important result that we will revisit in chapter 8.

### 3.2 Passive Listening and Active Feeling

It is by carefully comparing the active emission with the echo that bats and dolphins are able to construct an acoustic analog to the visual scene. However, it is possible to achieve a significant degree of spatial perception on the basis of passive audition alone. Barn owls, for instance, can localize and track prey items (e.g., mice) along a three-dimensional grid by detecting small acoustic arrival delays between their left and right ears, which are then processed by a network of neurons that form a spatial map. Cells of this spatially specific cluster of neurons are organized into a topographic array that represents the horizontal and vertical positions of the target object. Because low-frequency sounds provide general information about horizontal position and high-frequency sounds provide more precise information about elevation, the owl must analyze a spectrum of frequencies in order to construct a three-dimensional “image” of the prey item’s position and velocity. It accomplishes this, as Eric Knudsen and colleagues have shown, thanks to a specialized facial anatomy: the densely packed feathers of the facial ruff that amplifies high-frequency sounds, making the right ear more sensitive to high-frequency sounds from above, and the left ear more sensitive to high-frequency sounds from below.<sup>26</sup> This enables the barn owl to locate prey along a three-dimensional grid and capture it in total darkness.

Scare quotes are placed around the word “image” above because it is unclear whether the barn owl “sees” its prey in a vision-like sense of the word. Even in its most spatially sensitive forms, the “passive” nature of audition severely limits the topographic information and textured detail that can be gleaned from the surrounding world. Living objects can generate sounds—such as a vole scurrying in the underbrush, a cricket chirping, or a songbird vocalizing—but they do not have sounds emanating from across their three-dimensional surfaces unless they are actively ensonified at high frequencies. When the world is dark and quiet, animals with even the most sophisticated audioreception are blind.

Moreover, pinpointing a sound made by an object in space—even the beating heart at its center—does not entail a representation of the object itself. We can infer the location of a particular object from the precise localization of a particular sound. But the individuation conditions for three-dimensional material objects are different from the individuation conditions for the objects of sound,

even if the latter can be spatially located. Passive audition cannot, therefore, support imagistic object recognition, let alone the creation of a visual scene.

A certain degree of spatial imaging can be achieved by tactile senses as well. The elaborate nasal organ of the star-nosed mole, for instance, is less of a “nose” and more of a sophisticated tactile device; upon making active contact with objects, it is capable of the near-spontaneous discerning of shape and movement, which allows the mole to locate prey items in the complete absence of light.<sup>27</sup> The somatosensory systems of many fossorial (subterranean) mammals, some of which are blind, are connected up with brain regions that process visual information, suggesting that a degree of spatial information is afforded by the tactile sense.

The transduction of mechanical signals, such as pressure and vibration, has produced spatial cognition in other groups of mammals as well, such as in pinnipeds (seals, sea lions, and walruses). The harbor seal, for example, has highly specialized facial vibrissae (whiskers) that compliment or even substitute for vision during foraging in dark or turbid environments, in a way that is analogous to the “lateral line” of fishes and some amphibians. Unlike the lateral line, however, the vibrissae of pinnipeds are so sensitive to water disturbances that they can discern the three-dimensional vortices left by hydrodynamic fish trails, which contain information about the body shape, size, and swimming style of the organism that left them.<sup>28</sup> It is unclear if the somatosensory images that pinnipeds form are properly imagistic and if so how detailed these images may be. Are pinnipeds gleaning information from fish trails in the same way that dogs glean information from scent trails? In any case, tactile sensory systems are restricted in the informational detail they provide and the speed at which images can be formed, updated, and acted upon. The spatial details of hydrodynamic vortices left by moving objects can prove useful for tracking, but they do not support real-time scene reconstruction.

### 3.3 Universal Constraints on Acoustic ISMs

The shape of acoustic ISMs is determined by general laws that transcend the body plans of the particular lineages in which these ISMs are found. Active sonar, whether biological or human-made, is governed by universal physical constraints described by the quantitative laws and models of wave propagation (such as principles of diffraction, the inverse square law, Doppler effects, atmospheric attenuation, etc.). Biosonar is subject to trade-offs between frequency, resolution, intensity, and range that are resolved in ways that are amenable to prediction given adequate knowledge of local ecology. Bat echolocation dynamics and call structures can be predicted on the basis of niche information alone, irrespective of bat phylogeny.

A classic problem for vision science concerns how a three-dimensional representation of the world is constructed from a two-dimensional retinal array and other two-dimensional neuronal maps. Vision solves this problem by using pictorial, oculomotor, binocular, and motion-related cues to reliably infer depth. Many of these solutions are convergent among visual organisms, such as between camera-eyed vertebrates and compound-eyed arthropods.<sup>29</sup> Although this three-dimensional construction problem applies as much to the acoustic images formed by bats and cetaceans as it does to visual images, echolocating animals can take advantage of unambiguous spatiotemporal information encoded in the echoes of ensonified objects (information that is not available in reflected light). This enables them to construct reliable three-dimensional representations of their surrounding world. As we saw earlier, bats and dolphins can discern differences in arrival time and frequency between their ears, giving them the location of an object in two-dimensional space; the third coordinate—distance—is computed unambiguously by measuring the time delay between the emission and echo. By monitoring changes in echo delay and other parameters over time, echoic animals can construct a dynamic, rapidly updating, topographic model of the external world: a true acoustic scene.

Because successful echolocation hinges on the universal properties of sound, solutions to echolocation problems are highly constrained by waveform physics. One constraint, for instance, is atmospheric attenuation. The energy of acoustic emissions dissipates as it travels as a sound wave through the atmosphere. As a result, echolocation will always be a relatively short-range detection system as compared with vision because light propagates much more effectively than sound in air. A decrease in bandwidth increases the range at which echolocation is effective but decreases resolution; a decrease in call frequency means that small objects will not be detected but reduces atmospheric attenuation.<sup>30</sup>

Because of these physical constraints, certain convergent call sequence structures have emerged. For example, bat echolocation begins with a prey search phase that sweeps a wider area, using long-duration calls of narrow bandwidth and low frequency, which are less subject to attenuation and thus reach farther and wider into space. Once a potential prey item is detected, the bat switches to shorter calls of broader bandwidth, which permit greater resolution and localization of the item. If the item is determined to be a suitable prey, the bat then switches to the final “feeding buzz” phase, during which call rates increase and call duration decreases to create a buzz-like sound that optimizes prey interception in three-dimensional space.<sup>31</sup> Similar call structures and distance compensation dynamics are known in echolocating whales.<sup>32</sup>

Another universal constraint on echolocation relates to Doppler effects—shifts in waveform frequency due to movement of the sound source relative

to the receiver. During flight, Doppler shifts present opportunities for overcoming certain constraints of echolocation, such as self-deafening. Because echolocation involves comparing pulse-echo pairs, it raises the specter of self-deafening, wherein an emission interferes with the return (echo) of a previous emission. This is more severe for microchiropterans than it is for other bats and echolocating animals, because they must capture small insect prey that reflect weaker echoes. One way that bats avoid self-deafening is by separating emissions in time, using the quiet interval to evaluate the echo; another method is to separate pulse emission and echo in the frequency domain, exploiting the Doppler effect to ensure that returning echoes arrive at a frequency that is optimal for hearing.

In order to avoid the deafening of conspecifics, some bats employ a jamming avoidance response, rapidly shifting frequencies or flying silent when foraging near conspecifics.<sup>33</sup> Because jamming is a problem facing any active emission sensory system, it is perhaps not surprising (though no less amazing) that similar jamming avoidance responses are deployed by weakly electric fish (see section 4). The speed of sound is so fast in water that it makes it difficult for echolocating whales to exploit similar Doppler effects. However, the fact that acoustic emissions propagate much farther and faster in the water medium means that there is less attenuation of ultrasound in water, and thus that echolocation can be used for broader-scale “visual” sweeping of the undersea environment.

These constraints and trade-offs must be resolved by all acoustic ISMs, on Earth and beyond. There are equally universal anatomical and metabolic constraints on the evolvability of echolocation that explain why it is “harder” to evolve than vision. First, as noted earlier, a powerful sound-production capacity, such as the lungs of tetrapods, is required to produce high-frequency emissions capable of supporting high-resolution acoustic imaging. Second, the costs of echolocation are high, which may limit acoustic imaging to organisms with high-metabolisms, such as mammals and birds. The metabolic rates of bats during echolocation, for instance, are up to five times greater than they are at rest. These costs have been offset in bats through the evolutionarily ingenious coupling of sound emission to wing-beat cycle, which function as a single unit of biomechanical and metabolic efficiency.<sup>34</sup> Sound emission is coupled with the upstroke phase of the wing-beat cycle, coinciding with contraction of abdominal muscles and pressure on the diaphragm. This significantly reduces the price of high-intensity pulse emission, making it nearly costless.<sup>35</sup> It is also why, as any careful crepuscular observer may have noticed, bats spend hardly any time gliding (which is otherwise a more efficient means of flight).

Universal constraints and internal limiting conditions can help explain cases of nonrepetition. For example, low-metabolic rates may explain why biosonar

does not seem to have evolved in fully aquatic marine reptiles such as ichthyosaurs, which had active foraging strategies that parallel those of toothed whales. But why have pinnipeds not developed the ability to echolocate, given that they have the requisite anatomic and metabolic capacities and tend to forage in dark and murky coastal waters, hunting prey items that are similar to those sought by dolphins? The absence of echolocation in pinnipeds is likely due less to internal constraints and more to the physics of echolocation itself.

Pinnipeds have an obligatory amphibious lifestyle, which requires that they spend a significant portion of their time on land. Ear structures evolve to match impedance, and secondarily aquatic mammals have to modify ears adapted to air in order to match the impedance of water. Because toothed whales are fully adapted for aquatic life, they have undergone ear restructuring for a wholly maritime existence. Pinnipeds, on the other hand, have retained the ability to hear well in the air medium; they give birth terrestrially (on land or ice) and thus must spend significant portions of their time on land. These demands have constrained their ability to evolve the exceptional underwater hearing apparatus necessary for sophisticated biosonar. This could change at some time in the evolutionary future should pinnipeds take fully to the sea.

## 4. Electrolocation

Biosonar is not an evolutionary option for animals living or foraging in darkness if they lack either sound emission capability or the metabolic rates necessary to sustain high energy acoustic emissions. This is the case for certain fish living in dark and turbid waters, which have evolved an alternative mode of image-formation: electrolocation.

### 4.1 Passive Electoreception versus Active Electrolocation

The passive perception of electromagnetic fields is common in the animal world. Many animals use information about the Earth's magnetic field for migration, orientation, and navigation. Magnetoreception has evolved convergently in all groups of vertebrates as well as in arthropods and mollusks, though the precise mechanisms of magnetoinduction are poorly understood and probably variable across animal phyla. Geomagnetic fields provide coarse-grained navigational information that affords organisms with something akin to a magnetic compass. And because geomagnetic fields vary over space, they can also provide finer-grained (though still low-resolution) information about position relative a specific place or object. However, geomagnetic fields contain very limited spatial information. Although magnetoreception can support directional

navigation, it does not allow for the imaging of objects and terrain in an organism's environment because geomagnetic fields (unlike light and sound waves) do not interact energetically with ecologically relevant objects.

True ISM-grade perception in the electromagnetic domain has been achieved by electrolocation. Unlike magnetoreception, electrolocation involves the active production of electric currents that generate magnetic fields. The rudimentary production of electricity has been achieved in several fish groups, some of which are capable of emitting high-voltage electric discharges to stun prey or deter predators. Passive electroreceptive organs, on the other hand, enable a wide range of taxa, from fish and birds to insects, to detect weak electric signals that enable them to zero in on visually hidden foraging items. These simple uses of electric fields are informationally depauperate and do not support the construction of "electric scenes."

Just as achieving a full-fledged acoustic ISM requires active sound emissions, achieving a full-blown ISM in the domain of electromagnetism requires active electromagnetic emissions. Electrolocation has arisen at least twice in populations of weakly electric freshwater fish—in African mormyrids and South American gymnotiformes—in each case produced by structures that were co-opted from muscle tissue. We will consider later whether this co-optation of a conserved tissue substrate undermines the independence of these iterations.

Electrical image-formation and scene reconstruction in weakly electric fish is achieved through active electric organ discharges from the tail region that create a stable, three-dimensional electromagnetic field around the signaling fish. The fish sense their own electric discharge with electroreceptor cells that specialize in detecting either the frequency or amplitude of electric signals. These electroreceptive cells, called mormyromasts, are distributed over much of their skin, and an "electric fovea" is positioned in the front of their head. An "electric flow field" around the fish's head and body provides information about movement and relative position, much as it does in vision and echolocation.

Mormyrids, which are the better studied of the two groups of weakly electric fish, actually have two specialized electric foveae: the first is in the nasal region, and is dedicated to longer-range electrolocation for navigation and object detection; the second is in the *schnauzenorgan*, a long and flexible chin-like appendage covered with densely packed mormyromast electroreceptive cells, which is used primarily for shorter-range prey detection and discrimination. This is akin to the dual visual foveae of certain birds (e.g., pigeons) which are specialized for flight control and foraging, respectively.

Electric organ discharges are not used solely for the purpose of imaging objects and the environment. They also serve to coordinate social behavior, from mating in gymnotiformes<sup>36</sup> to pack hunting in cichlid-eating mormyrids.<sup>37</sup>

Weakly electric fish are commonly kept as shy inhabitants of tropical fish aquariums. Though the electric outputs of these fish are well known, few aquarium enthusiasts appreciate just how remarkable are their sensory worlds.

## 4.2 Electric Object Recognition

How does electrolocation work, and what can we say about its representational and phenomenological qualities? Constant electric organ discharges emanating from the caudal region maintain a stable spatial voltage pattern over the skin surface. This voltage pattern changes when objects that have a resistance different from the surrounding water come within range of the signal and distort the field, resulting in changes of local electric voltages at particular skin loci. Objects can alter the stable electric discharge field in waveform and/or in amplitude, and weakly electric fish can detect both types of disruptions. These changes in local transepidermal electric current flow are recorded by the skin electroreceptors, which act as a “retina” upon which an electric image of the object is projected. This image is then transduced, and the information is fed to regions of the brain that process higher-order features of objects. Whereas in humans the processing of higher-order features of objects takes place in the cerebral cortex, in electrolocating fish these cognitive tasks are carried out in their hypertrophied cerebellum. The “mormyrocerebellum” is so oversized that it accounts for the vast majority of the organism’s total oxygen consumption, with metabolic expenditures exceeding that of any vertebrate.<sup>38</sup> This, in turn, speaks to the great functional utility of electrolocation: all that brain stuff must be doing something computationally demanding and ecologically important.

Two decades of pioneering work by Gerhard von der Emde and his collaborators has shown that weakly electric fish can perceive numerous properties of objects, including the passive and resistive components of the object’s electrical impedance, as well as its size, shape, depth, spatial orientation, and distance.<sup>39</sup> Like bats and cetaceans, weakly electric fish use active electrolocation not only to detect and classify objects, but also to evaluate and memorize distances, shapes, textures, patterns, and configurations—allowing them to solve object recognition tasks and construct electrical scenes.

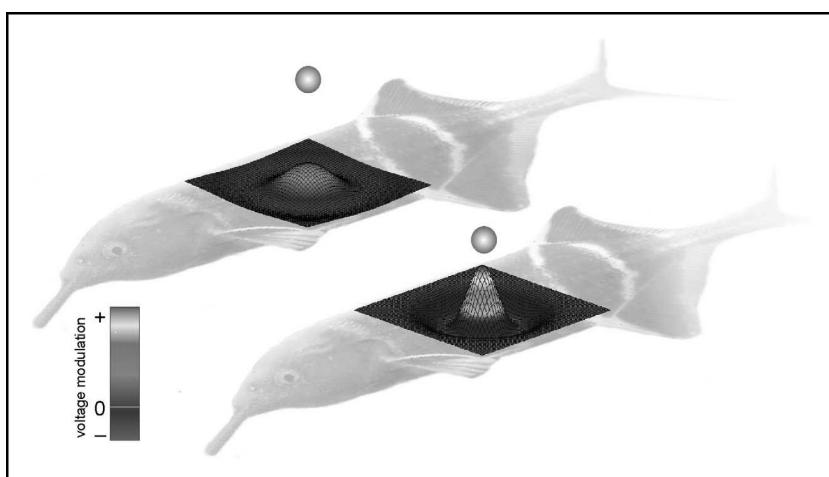
Mormyrids can, solely with their electrical sense, discriminate between real and virtual electrical objects based on their three-dimensional orientations and configurations. Like the virtual acoustic stimuli that are used in bat and dolphin echolocation research, virtual electrical objects allow experimenters to control for visual, chemical, and mechanical means of perception. Mormyrids have been shown to electrically discriminate between objects on the basis of their shape, which is apparently more salient for them than other properties such as material composition, size, or spatial orientation.

For instance, where electrolocating fish were trained to receive a positive reward (in the form of a conspecific electric organ discharge) by learning and remembering to choose a metal cube, they later preferred a plastic cube to a metal cylinder, despite the difference in material; similar results have been obtained with spheres, pyramids, ellipsoids, and crosses.<sup>40</sup> Electrical object identification and discrimination is immediate and global, suggesting that the various parametric assessments that go into these discriminations are “bound” into a single percept. In this respect, electrolocation is similar to acoustic image-formation in echolocating animals. Although electrolocation is not focus-capable, when inspecting an object weakly electric fish engage in various “motor probing” behaviors to achieve better three-dimensional electrical resolution. By bending its body and factoring in proprioceptive information, the weakly electric fish alters field direction and skin receptor position, allowing it to glean additional information about its surrounding landscape—movements that are analogous to the “acoustic angling” behaviors observed in echolocating cetaceans.

Mormyrids have also been taught to discriminate between objects on the basis of distance irrespective of the size, shape, or electrical properties of the targets, showing that they have a true sense of depth perception.<sup>41</sup> They work out distance up to a maximum of 10 centimeters, apparently by measuring the normalized maximal slope of the electric image, which is affected only by object distance and not by object size. By comparing the intensity at the electrical image center with the “fuzziness” of the image boundaries, weakly electric fish can compute distance unequivocally: the farther away an object is, the more “out of focus” or fuzzy the image becomes (see figure 7.2).<sup>42</sup>

Thus, like echolocation, electrolocation does not need to rely on probabilistic cues to establish the depth dimension, as vision is forced to do. The degree of focus, in conjunction with maximum amplitude, is the primary means by which mormyrids compute depth. Once distance is computed, the size of the object can be ascertained unambiguously by comparing image width with the distance measure. And once distance and size are determined, the fish can then use this information to calibrate waveform and amplitude measures and hence determine the object’s complex impedance. As with echolocation, this is all performed outside of experiential awareness, with holistic bound object percepts spontaneously “presented” to the fish for comparison with affectively valenced representations stored in memory, thus guiding behavior.

Even more incredibly, just as color vision measures the wavelengths of light reflected by an object, so too are electrolocating fish able to detect an object’s capacitance properties through waveform distortion.<sup>43</sup> Just as color perception measures the color (frequency) and brightness (amplitude) of light independently of one another, electrolocating fish perceive resistance and capacitance

**Figure 7.2**

Electric image of a spherical object projected onto the skin of a mormyrid (*Gnathonemus petersii*). Amplitude changes correspond to shape and distance of the object (which is closer in the bottom image), with an increase in distance resulting in a larger electric image. By comparing the fuzziness at the image borders to the maximum amplitude change at the image center, the fish is able to compute distance unequivocally. From G. von der Emde and S. Schwarz, “Imaging of Objects through Active Electrolocation in *Gnathonemus petersii*,” *Journal of Physiology-Paris* 96, no. 5–6 (2002): 431–444.

properties independently of one another. In this way, the perception of capacitance adds “color” to an otherwise “black-and-white” electrical world.<sup>44</sup> Because animate objects (such as plants and animals) tend to have capacitance properties, living things brilliantly pop out as colorful objects against a “gray” inanimate background. The result is the perceptual construction of a rich three-dimensional scene that permits sophisticated foraging and navigation in the complete absence of light and sound.

Just as striking, von der Emde and his collaborators have recently demonstrated that weakly electric fish are capable of the spontaneous recognition of objects across the visual and electric domains and vice versa.<sup>45</sup> In cross-modal recognition, an object is recognized as the same object based on one or more of its global features (such as shape), even though information about these global features is provided by distinct sensory modalities exploiting different energy forms in each case. To do this, the representation of one ISM (e.g., vision) must be encoded, stored in memory, and accessible for comparison against the information flowing in from an entirely different ISM (e.g., electrolocation or echolocation). It seems likely that this stored representation is encoded in a generic “imagistic” format that allows cross-modal recognition to occur spontaneously without associative learning. Mormyrid fish trained to

discriminate objects using vision performed identically during modal transfer tests in which only the electric sense was available for discrimination, and they did so without any previous training.

As with cross-modal task transfer in echolocating dolphins (discussed earlier), spontaneous cross-modal recognition in weakly electric fish strongly suggests that electrolocated objects are being perceived holistically in three dimensions with a representational format and/or phenomenological quality that is analogous in fundamental ways to vision. Object recognition across ISMs is thus a robustly replicable phenomenon and is indicative of both the common representational formats of ISM percepts and their global access. Further, as von der Emde and colleagues point out, cross-modal recognition is not a quirk of experimental artifice. Rather, it is a crucial adaptive functionality that ensures reliable perception in complex environments in which information flowing in from different senses must be weighted and adjusted in accordance with fluctuating conditions, such as changes in turbidity, lighting conditions, and so forth.

### 4.3 Phenomenology of the Electric Sense

What is it like to “see” as the weakly electric fish does? Some would deny that there is anything “it is like” to be a weakly electric fish at all, and thus maintain that there is no electric phenomenology to speak of. Let us set aside questions of phenomenal consciousness until the next chapter and instead use representation as a fill-in for phenomenology. Electrolocation is a sensory modality that is qualitatively alien to us, and it has some notably weird representational properties that have no analog in vision or echolocation. For instance, larger electric images may be nonsymmetrical even if the objects themselves are symmetrical (such as cubes). This is because the electric discharge organ is located in the tail, and larger objects will cross areas of the electric field that bend at different angles along the fish’s body and head, resulting in “cubist”-esque percepts.

Furthermore, the electric image of a multiobject scene is not the mere addition or patchwork of individual object images (as in vision), because the presence of proximate objects may distort the image of others and the images of different objects can superimpose. Given all of these differences, are we still justified in calling electrolocation a mode of “seeing”? We will return to the relevance of physics, representation, and phenomenology to the delineation of sensory modalities in the next chapter. The main lesson for now is that electrolocation is a full-fledged ISM that draws upon an entirely different energy source than vision and echolocation.

## 5. Evolutionary Robustness of the ISM

Let us bring the various lines of discussion together in considering the evolutionary replicability of the ISM. The visual ISM is likely to evolve earliest and most often, given its high evolvability and wide ecological applicability. This is precisely what we have seen: vision arose repeatedly in the earliest phases of animal evolution, and as visual acuity increased, so did the perceptual and cognitive machinery necessary to capitalize on the rich flow of information that it provided, resulting in the convergent evolution of brains with minds (see chapters 9 and 10). The emergence of visual ISMs—not one, but several—so soon after complex multicellular life arose speaks strongly in favor of its law-like replicability (see chapter 1), at least taking animal-grade multicellularity among the initial conditions of the replay.

But if it is so evolutionarily robust, then why did true vision and its neural-cognitive accoutrements never arise again in more than 400 million years, even while optics have evolved myriad times? The answer is unclear. We can surmise that it has to do with a combination of packed visual niches (with incumbent advantage precluding additional transitions to visual morphologies and lifeways) and the “congealing” of body plans that became recalcitrant to the developmental transformations that would be entailed by later origins of the visual ISM.

Alternative ISMs arose well after all major origins of the visual ISM, which is consistent with their narrower evolvability conditions. Once the neural structures dedicated to spatial perception, and their accompanying motor and proprioceptive systems, were up and running, the evolution of alternative ISMs would be likely to draw upon preexisting neurocognitive architecture even while they exploit a different waveform energy through a new transduction apparatus. Echolocating bats and dolphins, for instance, have co-opted regions of the mammalian brain that subserve spatial perception, such as place cells of the hippocampus. And recent work on dolphin brains shows that auditory information is fed into both the temporal lobe (associated with audition) and primary visual regions that process imagistic information.<sup>46</sup> In contrast, in electrolocating fish, the higher order features of images are processed in nonhomologous regions of the cerebellum.

We should also expect to see the recruitment of common molecular substrates in the evolutionary development of alternative ISMs. Convergence on echoic abilities between bats and dolphins, for example, has occurred not only at the level of functional morphology but also at the genomic level, with numerous convergent amino acid substitutions in genes linked to hearing.<sup>47</sup> As with the role of *Pax6* in the evolution of vision (see chapters 5 and 6), the fact that generic gene-regulatory resources associated with hearing are implicated

in the evolution of echolocation in distant mammalian groups does not convert these iterations into parallelisms or Gouldian repetitions. This would only be so if we had reason to believe that the coopted resources were themselves radically contingent—which we do not, given the robust replicability of vision, hearing, and the brains that process their inputs (see chapter 9).

In sum, we have identified convergent ISMs in many living lineages, and it seems likely that many more of them occurred in the deep evolutionary past without leaving any trace of their existence in the fossil record (or leaving traces that have yet to be found). The iterated evolution of three distinct classes of ISMs shows that complex forms of perception have been realized through multiple, externally constrained routes by exploiting the limited set of energy forms that can support image-formation—presenting numerous evolutionary gateways to mind. Each type of ISM has its own distinct advantages, limitations, and associated phenomenology. Common to each of them, however, is that they enable organisms to construct a panorama of identifiable objects, bundled with properties, that are distributed in space and time relative to the perceiver—a phenomenal world that is fundamentally familiar to the one that we experience. The next three chapters will examine these convergent solutions to the problem of mind and its evolution.



# Contingency and Convergence

## Toward a Cosmic Biology of Body and Mind

By: Russell Powell

### Citation:

*Contingency and Convergence: Toward a Cosmic Biology of Body and Mind*

By: Russell Powell

DOI: [10.7551/mitpress/11182.001.0001](https://doi.org/10.7551/mitpress/11182.001.0001)

ISBN (electronic): 9780262356596

Publisher: The MIT Press

Published: 2020



The MIT Press

© 2020 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times Roman by Westchester Publishing Services, Danbury, CT.

Library of Congress Cataloging-in-Publication Data

Names: Powell, Russell, author.

Title: Contingency and convergence : toward a cosmic biology of body and mind / Russell Powell.

Description: Cambridge, MA : MIT Press, [2019] | Series: Vienna series in theoretical biology | Includes bibliographical references and index.

Identifiers: LCCN 2019009760 | ISBN 9780262043397 (hardcover : alk. paper)

Subjects: LCSH: Evolution (Biology)--Philosophy. | Convergence (Biology)--Philosophy.

Classification: LCC QH360.5 .P68 2019 | DDC 576.8--dc23

LC record available at <https://lccn.loc.gov/2019009760>