Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

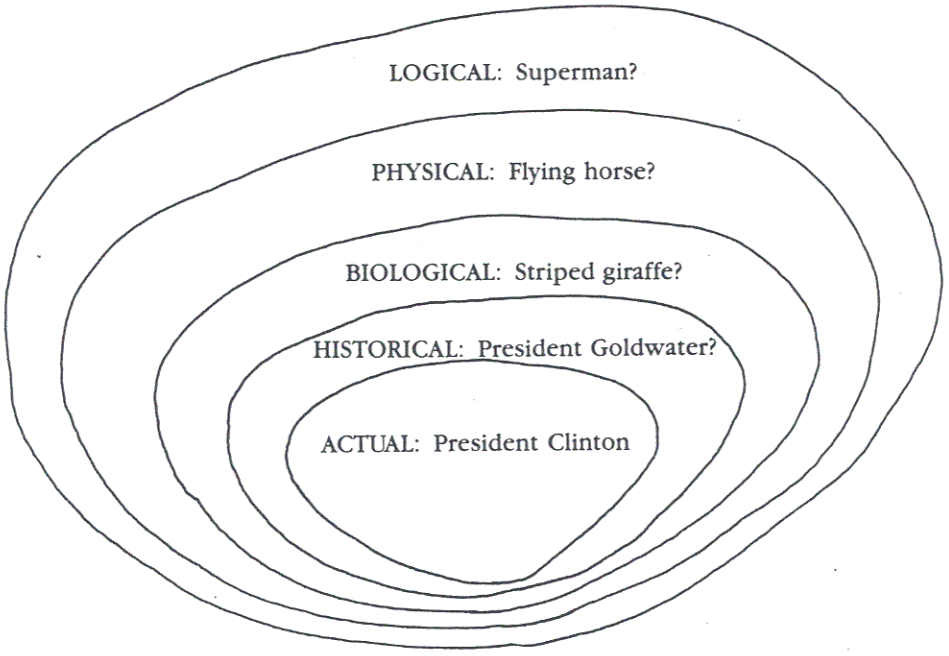

FIGURE 5.1

biological possibility, and then perhaps it will suggest some payoffs for how to make sense of the grander varieties.[1]

2. THE LIBRARY OF MENDEL

The Argentine poet Jorge Luis Borges is not typically classified as a philosopher, but in his short stories he has given philosophy some of its most valuable thought experiments, most of them gathered in the stunning collection *Labyrinths* (1962). Among the best is the fantasy—actually, it is more a philosophical reflection than a narrative—that describes the Library of Babel. For us, the Library of Babel will be an anchoring vision for helping to answer very difficult questions about the scope of biological possibility, so we will pause to explore it at some length. Borges tells of the forlorn explorations and speculations of some people who find themselves living in

---

1. Back in 1982, François Jacob, the Nobel laureate biologist, published a book entitled *The Possible and the Actual*, and I rushed to read it, expecting it to be an eye-opening essay on how biologists should think about some of these conundrums about possibility. To my disappointment, the book had very little to say on this topic. It is a fine book, and has a great title, but the two don't go together, in my humble opinion. The book I was eager to read hasn't yet been written, apparently, so I'll have to try to write part of it myself, in this chapter.

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

a vast storehouse of books, structured like a honeycomb, composed of thousands (or millions or billions) of hexagonal air shafts surrounded by balconies lined with shelves. Standing at a railing and looking up or down, one sees no top or bottom to these shafts. Nobody has ever found a shaft that isn't surrounded by six neighboring shafts. They wonder: is the warehouse infinite? Eventually, they decide that it is not, but it might as well be, for it seems that on its shelves—in no order, alas—lie *all the possible books.*

Suppose that each book is 500 pages long, and each page consists of 40 lines of 50 spaces, so there are two thousand character-spaces per page. Each space either is blank, or has a character printed on it, chosen from a set of 100 (the upper- and lowercase letters of English and other European languages, plus the blank and punctuation marks).[2] Somewhere in the Library of Babel is a volume consisting entirely of blank pages, and another volume is all question marks, but the vast majority consist of typographical gibberish; no rules of spelling or grammar, to say nothing of sense, prohibit the inclusion of a volume. Five hundred pages times 2,000 characters per page gives 1,000,000 character-spaces per book, so there are $100^{1,000,000}$ books in the Library of Babel. Since it is estimated[3] that there are only 100^{40} (give or take a few) *particles* (protons, neutrons, and electrons) in the region of the universe we can observe, the Library of Babel is not remotely a physically possible object, but, thanks to the strict rules with which Borges constructed it in his imagination, we can think about it clearly.

Is this truly the set of *all* possible books? Obviously not—since they are restricted to being printed from "only" 100 different characters, excluding, we may suppose, the characters of Greek, Russian, Chinese, Japanese, and Arabic, thereby overlooking many of the most important *actual* books. Of course, the Library does contain superb translations of all these actual books into English, French, German, Italian, . . . , as well as uncountable trillions of shoddy translations of each book. Books of more than 500 pages are there,

---

2. Borges chose slightly different figures: books 410 pages long, with 40 lines of 80 characters each. The total number of characters per book is close enough to mine (1,312,000 versus 1,000,000) to make no difference. I chose my rounder numbers for ease of handling. Borges chose a character set with only 25 members, which is enough for uppercase Spanish (with a blank, a comma, and a period as the only punctuation), but not for English. I chose the more commodious 100 to make room without any doubt for the upper- and lowercase letters and punctuation of all the Roman-alphabet languages.

3. Stephen Hawking (1988, p. 129) insists on putting it this way: "There are something like ten million (1 with eighty zeroes after it) particles in the region of the universe that we can observe." Denton (1985) provides the estimate of 10^{70} atoms in the observable universe. Eigen (1992, p. 10) calculates the volume of the universe as 10^{84} cubic centimeters.

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

beginning in one volume and continuing without a break in some other volume or volumes.

It is amusing to think about some of the volumes that must be in the Library of Babel somewhere. One of them is the best, most accurate 500-page biography of you, from the moment of your birth until the moment of your death. Locating it, however, would be all but impossible (that slippery word), since the Library also contains kazillions of volumes that are magnificently accurate biographies of you up till your tenth, twentieth, thirtieth, fortieth . . . birthday, and completely false about subsequent events of your life—in a kazillion different and diverting ways. But even finding one readable volume in this huge storehouse is unlikely in the extreme.

We need some terms for the quantities involved. The Library of Babel is not infinite, so the chance of finding anything interesting in it is not literally infinitesimal.[4] These words exaggerate in a familiar way—we caught Darwin doing it in his summary, where he helped himself to an illicit "infinitely"—but we should avoid them. Unfortunately, all the standard metaphors—"astronomically large," "a needle in a haystack," "a drop in the ocean"—fall comically short. No *actual* astronomical quantity (such as the number of elementary particles in the universe, or the time since the Big Bang measured in nanoseconds) is even visible against the backdrop of these huge but finite numbers. If a readable volume in the Library were as easy to find as a particular drop in the ocean, we'd be in business! If you were dropped at random into the Library, your chance of ever encountering a volume with so much as a grammatical sentence in it would be so vanishingly small that we might do well to capitalize the term—"Vanishingly" small—and give it a mate, "Vastly," short for "Very-much-more-than-astronomically."[5]

Moby Dick is in the Library of Babel, of course, but so are 100,000,000 mutant impostors that differ from the canonical *Moby Dick* by a *single*

4. The Library of Babel is finite, but, curiously enough, it contains all the grammatical sentences of English within its walls. But that's an infinite set, and the library is finite! Still, any sentence of English, of whatever length, can be broken down into 500-page chunks, each of which is somewhere in the library! How is this possible? Some books may get used more than once. The most profligate case is the easiest to understand: since there are volumes that each contain a single character and are otherwise blank, repeated use of these 100 volumes will create any text of any length. As Quine points out in his informative and amusing essay "Universal Library" (in Quine 1987), if you avail yourself of this strategy of re-using volumes, and translate everything into the ASCII code your word-processor uses, you can store the whole Library of Babel in two extremely slender volumes, in one of which is printed a 0 and in the other of which appears a 1! (Quine also points out that the psychologist Theodor Fechner propounded the fantasy of the universal library long before Borges.)

5. Quine (1987) coins the term "hyperastronomic" for the same purpose.

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

typographical error. That's not yet a Vast number, but the total rises swiftly when we add the variants that differ by 2 or 10 or 1,000 typos. Even a volume with 1,000 typos—2 per page on average—would be unmistakably recognizable as *Moby Dick*, and there are Vastly many of those volumes. It wouldn't matter which of these volumes you found, if you could only find one of them. They would almost all be just about equally wonderful reading, and all tell the same story, except for truly negligible—almost indiscriminable—differences. Not quite all of them, however. Sometimes a single typo, in a crucial position, can be fatal. Peter De Vries, another philosophically delicious writer of fiction, once published a novel[6] that began:

"Call me, Ishmael."

Oh, what a single comma can do! Or consider the many mutants that begin: "Ball me Ishmael. . . ."

In Borges' story, the books are not shelved in any order, but even if we found them scrupulously alphabetized, we would have insoluble problems finding *the* book we were looking for (for instance, the "essential" version of *Moby Dick*). Imagine traveling by spaceship through the *Moby Dick* galaxy of the Library of Babel. This galaxy is in itself Vastly larger than the whole physical universe, so, no matter what direction you go in, for centuries on end, even if you travel at the speed of light, all you see are virtually indistinguishable copies of *Moby Dick*—you will never ever reach anything that looks like anything else. *David Copperfield* is unimaginably distant in this space, even though we know that there is a path—a shortest path, ignoring the kazillions of others—leading from one great book to the other by single typographical changes. (If you found yourself on this path, you would find it almost impossible to tell, by local inspection, which direction to go to move towards *David Copperfield*, even if you had texts of both target books in hand.)

In other words, this *logical* space is so Vast that many of our usual ideas about location, about searching and finding and other such mundane and practical activities, have no straightforward application. Borges put the books on the shelves in random order, a nice touch from which he drew several delectable reflections, but look at the problems he would have

---

6. *The Vale of Laughter* (1953). (It goes on: "Feel absolutely free to. Call me any hour of the day or night. . . .") De Vries also may have invented the game of seeing how large an effect (deleterious or not) you can achieve with a single typographical change. One of the best: "Whose woods are these, I think I know; his house is in the *V*illage though. . . ." Others have taken up the game: in the state of nature, mutant-Hobbes tells us, one finds "the *w*ife of man, solitary, poore, nasty, brutish, and short." Or consider the question: "Am I my brothe*l*'s keeper?"

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

created for himself if he'd tried to arrange them in alphabetical order in his honeycomb. Since there are only a hundred different alphabetic characters (in our version), we can treat some specific sequence of them as Alphabetical Order—e.g., a, A, b, B, c, C . . . z, Z, ?, ;, „, ., !,), (, %, . . . à, â, è, ê, é, . . . Then we can put all the books beginning with the same character on the same *floor*. Now our library is only 100 stories high, shorter than the World Trade Center. We can divide each floor into 100 *corridors*, each of which we line with the books whose second character is the same, one corridor for each character, in alphabetical order. On each corridor, we can place 100 *shelves*, one for each third-slot. Thus all the books that begin with "aardvarks love Mozart"—and how many there are!—are shelved on the same shelf (the "r" shelf) in the first corridor on the first floor. But that's a mighty long shelf, so perhaps we had better stack the books in file drawers at right angles to the shelf, one drawer for each fourth-letter position. That way, each shelf can be only, say, 100 feet long. But now the file drawers are awfully deep, and will run into the backs of the file drawers in the neighboring corridor, so . . . but we've run out of dimensions in which to line up the books. We need a million-dimensional space to store all the books neatly, and all we have is three dimensions: up-down, left-right, and front-back. So we will just have to pretend we can imagine a multidimensional space, each dimension running "at right angles" to all the others. We can conceive of such hyperspaces, as they are called, even if we can't visualize them. Scientists use them all the time to organize the expression of their theories. The geometry of such spaces (whether or not they count as only imaginary) is well behaved and well explored by mathematicians. We can confidently speak about locations, paths, trajectories, volumes (hypervolumes), distances, and directions in these logical spaces.

We are now prepared to consider a variation on Borges' theme, which I will call the *Library of Mendel*. This Library contains "all possible genomes"—DNA sequences. Richard Dawkins describes a similar space, which he calls "Biomorph Land," in *The Blind Watchmaker* (1986a). His discussion is the inspiration for mine, and our two accounts are entirely compatible, but I want to stress some points he chose to pass over lightly.

If we consider the Library of Mendel to be composed of *descriptions* of genomes, then it is already just a proper part of the Library of Babel. The standard code for describing DNA consists of only four characters, A, C, G, and T (standing for Adenine, Cytosine, Guanine, and Thymine, the four kinds of nucleotides that compose the letters of the DNA alphabet). All the 500-page permutations of these four letters are already in the Library of Babel. Typical genomes are much longer than ordinary books, however. Taking the current estimate of 3×10^9 nucleotides in the human genome, the exhaustive description of a single human genome—such as your own—would take approximately 3,000 of the 500-page volumes in the Library of

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

Babel (keeping print size the same).[7] The description of the genome for a horse (flying or not) or a cabbage or an octopus would be composed of the same letters, A, C, G, and T, and certainly not much longer, so we can suppose, arbitrarily, that the Library of Mendel consists of all the DNA strings described in all the 3,000-volume boxed sets consisting entirely of those four characters. This will capture enough of the "possible" genomes to serve any serious theoretical purpose.

I overstated the case in describing the Library of Mendel as containing "all possible" genomes, of course. Just as the Library of Babel ignored the Russian and Chinese languages, so the Library of Mendel ignores the (apparent) possibility of alternative genetic alphabets—based on different chemical constituents, for instance. We are *still* beginning in the middle, making sure we understand today's local, earthly circumstances before casting our nets wider. So any conclusions we come to regarding what is possible relative to *this* Library of Mendel may have to be reconsidered when we try to apply them to some broader notion of possibility. This is actually a strength rather than a weakness of our tactic, since we can keep close tabs on exactly what sort of modest, circumscribed possibility we are talking about.

One of the important features of DNA is that all the permutations of sequences of Adenine, Cytosine, Guanine, and Thymine are about equally stable, chemically. All could be constructed, in principle, in the gene-

---

7. The comparison of a human genome with the volumes in the galaxy of *Moby Dick* readily explains something that occasionally baffles people about the Human Genome Project. How can scientists speak of sequencing (copying down) *the* human genome if every human genome is different from every other in not just one but hundreds or thousands of places (*loci*, in the language of genetics)? Like the proverbial snowflakes, or fingerprints, no two actual human genomes are exactly alike, even those of identical twins (the chance of typos creeping in is always present, even in the cells of a single individual). Human DNA is readily distinguishable from the DNA of any other species, even that of the chimpanzee, which is over 90 percent the same at every locus. Every actual human genome that has ever existed is contained within a galaxy of possible human genomes that is Vastly distant from the galaxies of other species' genomes, yet within the galaxy there is plenty of room for no two human genomes to be alike. You have two versions of each of your genes, one from your mother and one from your father. They passed on to you exactly half of their own genes, randomly selected from those they received from their parents, your grandparents, but since your grandparents were all members of *Homo sapiens*, their genomes agree at almost all loci, so it makes no difference most of the time which grandparent provides either of your genes. But their genomes nevertheless differ at many thousands of loci, and in those slots, which genes you get is a matter of chance—a coin-toss built into the machinery for forming your parents' contributions to your DNA. Moreover, mutations accumulate at the rate of about 100 per genome per generation in mammals. "That is, your children will have one hundred differences from you and your spouse in their genes as a result of random copying errors by your enzymes or as a result of mutations in your ovaries or testicles caused by cosmic rays" (Matt Ridley 1993, p. 45).

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

splicing laboratory, and, once constructed, would have an indefinite shelf life, like a book in a library. But not every such sequence in the Library of Mendel corresponds to a viable organism. Most DNA sequences—the Vast majority—are surely gibberish, recipes for nothing living at all. That is what Dawkins means, of course, when he says there are many more ways of being dead (or not alive) than ways of being alive. But what kind of a fact is this, and why should it be so?

3. The Complex Relation Between Genome and Organism

If we are going to try to make progress by boldly oversimplifying, we should at least alert ourselves to some of the complications we are temporarily setting aside. I see three main sorts of complexity we should acknowledge and keep an eye on as we proceed, even if we are once again postponing their full discussion.

The first concerns the "reading" of the "recipe." The Library of Babel presupposed readers: the people who inhabited the Library. Without them, the very idea of the collection of volumes would make no sense at all; their pages might as well be smeared with jam or worse. If we are to make any sense of the Library of Mendel, we must also presuppose something analogous to readers, for without readers DNA sequences don't *specify* anything at all—not blue eyes or wings or anything else. Deconstructionists will tell you that no two readers of a text will come up with the same reading, and something similar is undoubtedly true when we consider the relationship between a genome and the embryonic environment—the chemical microenvironment as well as the surrounding support conditions—in which it has its informational effects. The immediate effect of the "reading" of DNA during the creation of a new organism is the fabrication of many different proteins out of amino acids (which have to be on hand in the vicinity, of course, ready to be linked together). There are Vastly many possible proteins, but which become actual depends on the DNA text. These proteins get created in strict sequence, and in amounts determined by the "words"—triplets of nucleotides—as they are "read." So, for a DNA sequence to specify what it is supposed to specify, there must be an elaborate reader-constructor, well stocked with amino-acid building blocks.[8] But that is just a small part of the process. Once the proteins get created, they have to be

8. This is an oversimplification, leaving out the role of messenger RNA and other complications.

Darwins' Dangerous Idea, Daniel C. Dennett, Touchtone, NY, 1995.

brought into the right relations with each other. The process begins with a single fertilized cell, which then divides into two daughter cells, which divide again, and so forth (each with its own duplicate copy of all the DNA that is being read, of course). These newly formed cells, of many different varieties (depending on which proteins are jiggled into which places in which order), must in turn migrate to the right locations in the embryo, which grows by dividing and dividing, building, rebuilding, revising, extending, repeating, and so forth.

This is a process that is only partly controlled by the DNA, which in effect *presupposes* (and hence does not itself *specify*) the reader and the reading process. Compare genomes to musical scores. Does a written score of Beethoven's Fifth Symphony *specify* that piece of music? Not to Martians, it wouldn't, because it presupposes the existence of violins, violas, clarinets, trumpets. Suppose we take the score and attach a sheaf of directions and blueprints for making (and playing) all the instruments, and send the whole package to Mars. Now we are getting closer to a package that could in principle be used to re-create Beethoven's music on Mars. But the Martians would still have to be able to decipher the recipe, make the instruments, and then play them as the score directed.

This is what makes the story of Michael Crichton's novel *Jurassic Park* (1990)—and the Steven Spielberg movie made of it—a fantasy: even completely intact dinosaur DNA would be powerless to re-create a dinosaur without the aid of a dinosaur-DNA-reader, and those are just as extinct as dinosaurs (they are, after all, the ovaries of dinosaurs). If you *have* a (living) dinosaur ovary, then it, together with dinosaur DNA, can specify *another* dinosaur, another dinosaur ovary, and so forth indefinitely, but dinosaur DNA by itself, even complete dinosaur DNA, is only half (or, depending on how you count, maybe less than half) the equation. We might say that every species that has ever existed on this planet has had its own dialect of DNA-reading. Still, these dialects have had a lot in common with each other. The principles of DNA-reading are apparently uniform across all species, after all. That is what makes genetic engineering possible; the organismic effect of a particular permutation in DNA can often be predicted in practice. So the idea of bootstrapping our way back to a dinosaur-DNA-reader is a coherent idea, however improbable. With a helping of poetic license, the film-makers might pretend that acceptable substitute readers could be found (introduce the dinosaur-DNA text to the DNA-reader in a frog, and hope for the best).[9]

---

9. The film-makers never really address the problem of the DNA-*reader* at all, and use frog DNA just to patch the missing parts of the dinosaur DNA. David Haig has pointed out to me that this choice of a frog by the film-makers manifests an interesting error—an