

## CHAPTER NINE

# Searching for Quality

---

### 1. THE POWER OF ADAPTATIONIST THINKING

*'Naked as Nature intended' was a persuasive slogan of the early Naturist movement. But Nature's original intention was that the skin of all primates should be un-naked.*

—ELAINE MORGAN 1990, p. 66

*Judging a poem is like judging a pudding or a machine. One demands that it work. It is only because an artifact works that we infer the intention of an artificer.*

—W. WIMSATT and M. BEARDSLEY 1954, p. 4

If you know something about the design of an artifact, you can predict its behavior without worrying yourself about the underlying physics of its parts. Even small children can readily learn to manipulate such complicated objects as VCRs without having a clue as to how they work; they know just what will happen when they press a sequence of buttons, because they know what is designed to happen. They are operating from what I call the *design stance*. The VCR repairer knows a great deal more about the design of the VCR, and knows, roughly, how all the interior parts interact to produce both proper functioning and pathological functioning, but may also be quite oblivious of the underlying physics of the processes. Only the designers of the VCR had to understand the physics; they are the ones who must descend to what I call the *physical stance* in order to figure out what sorts of design revisions might enhance picture quality, or diminish wear and tear on the tape, or reduce the electricity consumption of the product. But when they engage in *reverse engineering*—of some other manufacturer's VCR, for instance—they avail themselves not only of the physical stance, but also of what I call the *intentional stance*—they try to figure out *what the designers*

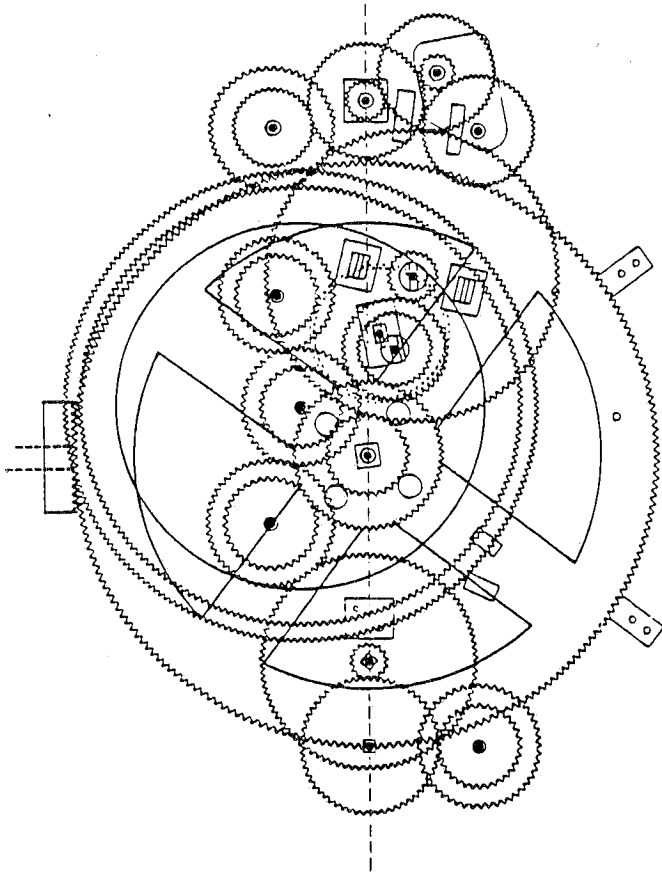


FIGURE 9.1.

Figure diagramming the wheel-work of the Antikythera mechanism by Derek de Solla Price (Yale University).

*had in mind*. They treat the artifact under examination as a product of a process of *reasoned* design development, a series of *choices* among alternatives, in which the *decisions* reached were those *deemed best* by the designers. Thinking about the postulated functions of the parts is making assumptions about the *reasons* for their presence, and this often permits one to make giant leaps of inference that finesse one's ignorance of the underlying physics, or the lower-level design elements of the object.

Archeologists and historians sometimes encounter artifacts whose meaning—whose function or purpose—is particularly obscure. It is instructive to look briefly at a few examples of such *artifact hermeneutics* to see how one reasons in such cases.<sup>1</sup>

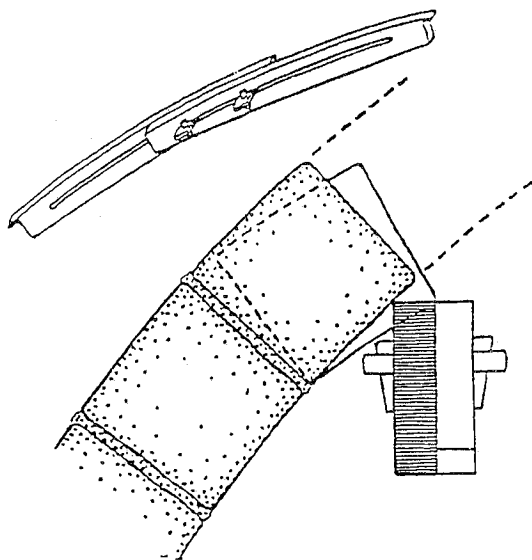
The Antikythera mechanism, discovered in 1900 in a shipwreck, and

1. For an expanded analysis of these issues, see Dennett 1990b.

dating from ancient Greece, is an astonishingly complex assembly of bronze gears. What was it for? Was it a clock? Was it the machinery for moving an automaton statue, like Vaucanson's marvels of the eighteenth century? It was—almost certainly—an orrery or a planetarium, and the proof is that it would be a *good* orrery. That is, calculations of the periods of rotation of its wheels led to an interpretation that would have made it an accurate (Ptolemaic) representation of what was then known about the motions of the planets.

The great architectural historian Viollet-le-Duc described an object called a *cerce*, used somehow in the construction of cathedral vaults.

He hypothesized that it was a movable piece of staging, used as a temporary support for incomplete web-courses, but a later interpreter, John Fitchen (1961), argued that this could not have been its function. For one thing, the *cerce* would not have been strong enough in its extended position, and, as figure 9.2 shows, its use would have created irregularities in the vault webbing which are not to be found. Fitchen's extended and elaborate



Viollet-le-Duc's *cerce* device as support for each web course during the erection of the vault. The smaller-scale drawing shows a *cerce*, based on Viollet-le-Duc's representation and description. Its extended position clearly indicates how one slotted board laps the other. Hung vertically as support for the stones of a web course, it is seen (in the detailed section) that the stones of any given course cannot line up throughout: those that lean against the far board (shown in outline) tilt much more than those that lean against the near board (shown hatched). As no such break does occur in the alignment of the web stone-coursing, it is obvious that the *cerce* device was not used in this fashion, in spite of Viollet-le-Duc's assertion that it was. [Fitchen 1961, p. 101.]

FIGURE 9.2

argument concludes that the cerce was no more than an adjustable template, a conclusion he supports by coming up with a much more elegant and versatile solution to the problem of temporary support of web courses.

The important feature in these arguments is the reliance on optimality considerations; it counts against the hypothesis that something is a cherry-pitter, for instance, if it would have been a demonstrably inferior cherry-pitter. Occasionally, an artifact loses its original function and takes on a new one. People buy old-fashioned sadirons not to iron their clothes with, but to use as bookends or doorstops; a handsome jam pot can become a pencil-holder, and lobster traps get recycled as outdoor planters. The fact is that sadirons are much better as bookends than they are at ironing clothes—when compared with the competition today. And a Dec-10 mainframe computer today makes a nifty heavy-duty anchor for a large boat-mooring. No artifact is immune from such appropriation, and however clearly its *original* purpose may be read from its current form, its new purpose may be related to that original purpose by mere historic accident—the fellow who owned the obsolete mainframe needed an anchor badly, and opportunistically pressed it into service.

The clues about such historical processes would be simply unreadable without assumptions about optimality of design. Consider the so-called dedicated word-processor—the cheap, portable, glorified typewriter that uses disk storage and an electronic display screen, but can't be used as an all-purpose computer. If you open up one of these devices, you find it is governed by an all-purpose CPU or central processing unit, such as an 8088 chip—a full-power computer vastly more powerful, swift, and versatile than the biggest computer Alan Turing ever saw—locked into menial service, performing a minuscule fraction of the tasks it *could* be harnessed to perform. Why is all this excess functionality found here? Martian reverse engineers might be baffled, but there is a simple historical explanation, of course: the genealogy of computer development gradually lowered costs of chip manufacture to the point where it was much cheaper to install a whole computer-on-a-chip in a device than to build a special-purpose control circuit. Notice that the explanation is historical but also, inescapably, proceeds from the intentional stance. It became *wise* to design dedicated word-processors this way, when the cost-benefit analysis showed that this was the *best, cheapest way to solve the problem*.

What is amazing is how powerful the intentional stance can be in reverse engineering, not only of human artifacts, but also of organisms. In chapter 6, we saw the role of practical reasoning—cost-benefit analysis in particular—in distinguishing the forced moves from what we might call the *ad lib* moves, and we saw how Mother Nature could be predicted to “discover” the forced moves again and again. The idea that we can impute such “free-floating rationales” to the mindless process of natural selection is dizzying,

but there is no denying the fruits of the strategy. In chapters 7 and 8, we saw how the engineering perspective informs research at every level from the molecules on up, and how this perspective *always* involves distinguishing the better from the worse, and the reasons Mother Nature has found for the distinction. The intentional stance is thus the crucial lever in all attempts to reconstruct the biological past. Did *Archaeopteryx*, the extinct birdlike creature that some have called a winged dinosaur, ever really get off the ground? Nothing could be more ephemeral, less likely to leave a fossil trace, than a flight through the air, but if you do an engineering analysis of its claws, they turn out to be excellent adaptations for *perching on branches*, not for *running*. An analysis of the claw curvature, supplemented by aerodynamic analysis of the archaeopteryx wing structure, makes it quite plain that the creature was *well designed* for flight (Feduccia 1993). So it almost certainly flew—or had ancestors that flew (we mustn't forget the possibility of excess functionality persisting, like the computer in the word-processor). The hypothesis that the archaeopteryx flew has not yet been fully confirmed to every expert's satisfaction, but it suggests many further questions to address to the fossil record, and when those questions are pursued, either the evidence will mount in favor of the hypothesis or it won't. The hypothesis is testable.

The lever of reverse engineering is not just for prying out secrets of history; it is even more spectacular as a predictor of unimagined secrets of the present. Why are there colors? Color-coding is generally viewed as a recent engineering innovation, but it is not. Mother Nature discovered it much earlier (for the details, see the section on why there are colors in Dennett 1991a, pp. 375–83). We know this thanks to lines of research opened up by Karl von Frisch, and, as Richard Dawkins points out, von Frisch used a bold exercise in reverse engineering to make the initial move.

Von Frisch (1967), in defiance of the prestigious orthodoxy of von Hess, conclusively demonstrated colour vision in fish and in honeybees by controlled experiments. He was driven to undertake those experiments by his refusal to believe that, for example, the colours of flowers were there for no reason, or simply to delight men's eyes. [Dawkins 1982, p. 31.]

A similar inference led to the discovery of the endorphins, the morphine-like substances that we produce in our own bodies when we are put under enough stress or pain—creating the “runner's high,” for instance. The reasoning was the reverse of von Frisch's. Scientists found receptors in the brain that are highly specific for morphine, which has a powerful painkilling effect. Reverse engineering insists that wherever there is a highly particular lock, there must be a highly particular key to fit it. *Why are these receptors here?* (Mother Nature could not have foreseen the development of mor-

phine!) There must be some molecules produced internally under some conditions, the original keys that these locks were designed to receive. Seek a molecule that fits this receptor and is produced under circumstances in which a shot of morphine might be beneficial. Eureka! Endogenously created morphine—endorphin—was discovered.

Even more devious Sherlock-Holmesian leaps of deduction have been executed. Here, for instance, is a general mystery: “Why do some genes change their pattern of expression depending on whether they are maternally or paternally inherited?” (Haig and Graham 1991, p. 1045). This phenomenon—in which the genome-reading machinery *pays more attention*, in effect, to either the paternal text or the maternal text—is known as *genomic imprinting* (for a general account, see Haig 1992), and has been confirmed to occur in special cases. What do the special cases have in common? Haig and Westoby (1989) developed a model that purports to solve the general mystery by *predicting* that genomic imprinting would be found only in organisms “in which females carry offspring by more than one male during their life span and a system of parental care in which offspring receive most of their postfertilization nutrients from one parent (usually the mother) and thus compete with offspring fathered by other males.” In such circumstances, they reasoned, there should be a conflict between maternal and paternal genes—paternal genes will tend to favor exploiting the mother’s body as much as possible, but maternal genes would “view” this as almost suicidal—and the result should be that the relevant genes will in effect choose sides in a tug-of-war, and genomic imprinting will result (Haig and Graham 1991, p. 1046).

See the model at work. There is a protein, “Insulin-like Growth Factor II” (IGF-II), which is, as its name suggests, a growth-enhancer. Not surprisingly, the genetic recipes of many species order the creation of large quantities of IGF-II during embryonic development. But, like all functioning machines, IGF-II needs the right supportive environment to do its work, and in this case it needs helper molecules known as “type 1 receptors.” So far, our story is just like the endorphin story: we have a type of key (IGF-II) and a kind of lock (type 1 receptors) in which it fits and performs an obviously important role. But in mice, for instance, there is another kind of lock (type 2 receptors) in which it also fits. What are these secondary locks for? For nothing, apparently; they are descendants of molecules that in other species (toads, for instance) play a role in cells’ “garbage-disposal” systems, but this is not what they do when they bind to IGF-II in mice. Then why are they there? Because they are “ordered” by the genetic recipe for making a mouse, of course, but here is the telltale twist: whereas both the maternal and paternal contributions to the chromosome contain recipe instructions for making them, these instructions are *preferentially expressed* from the maternal chromosome. Why? To counteract the instruction in the recipe that

calls for too much growth-enhancer. The type 2 receptors are just there to soak up—to “capture and degrade”—all the excess growth-enhancer that the paternal chromosome would pump into the fetus if it had its way. Since mice are a species in which females tend to mate with more than one male, males in effect compete to exploit the resources of each female, a competition from which females must protect themselves (and their own genetic contributions).

Haig and Westoby’s model predicts that genes would evolve in mice to protect females from this exploitation, and this imprinting has been confirmed. Moreover, their model predicts that type 2 receptors shouldn’t work this way in species in which genetic conflict of this sort can’t arise. They shouldn’t work this way in chickens, because offspring can’t influence how much yolk their eggs receive, so the tug-of-war can never get started. Sure enough, the type 2 receptors in chickens don’t bind to IGF-II. Bertrand Russell once slyly described a certain form of illicit argument as having all the advantages of theft over honest toil, and one can sympathize with the hardworking molecular biologist who reacts with a certain envy when somebody like Haig swoops in, saying, in effect, “Go look under that rock—I bet you’ll find a treasure of the following shape!”

But that is what Haig was able to do: he predicted what Mother Nature’s move would be in the hundred-million-year game of mammal design. Of all the possible moves available, he saw that there was a good reason for this move, so this is what would be discovered. We can get a sense of the magnitude of the leap that such an inference takes by comparing it with a parallel leap that we can make in the Game of Life. Recall that one of the possible denizens of the Life world is a Universal Turing machine composed of trillions of pixels. Since a Universal Turing machine can compute any computable function, it can play chess—simply by mimicking the program of any chess-playing computer you like. Suppose, then, that such an entity occupies the Life plane, playing chess against itself, in the fashion of Samuel’s computer playing checkers against itself. Looking at the configuration of dots that accomplishes this marvel would almost certainly be unilluminating to anyone who had no clue that a configuration with such powers could exist. But from the perspective of someone who *had the hypothesis* that this huge array of black dots was a chess-playing computer, enormously efficient ways of predicting the future of that configuration are made available.

Consider the savings you could achieve. At first you would be confronted by a screen on which trillions of pixels flash on and off. Since you know the single rule of Life Physics, you could laboriously calculate the behavior of each spot on the screen if you wanted, but it would take eons. As a first cost-cutting step, you could shift from thinking about individual pixels to thinking about gliders and eaters and still lifes, and so forth. Whenever you

saw a glider approaching an eater, you would just predict “consumption in four generations” without bothering with the pixel-level calculations. As a second step, you could move to thinking of the gliders as symbols on the “tape” of a gigantic Turing machine, and then, adopting this higher design stance towards the configuration, predict its future *as* a Turing machine. At this level you would be “hand-simulating” the “machine language” of a computer program that plays chess, still a tedious way of making predictions, but orders of magnitude more efficient than working out the physics. As a third and still more efficient step, you could ignore the details of the chess-playing program itself and just assume that, whatever they are, they are *good!* That is, you could assume that the chess-playing program running on the Turing machine made of gliders and eaters played not just legal chess but good legal chess—it had been well designed (perhaps it has designed itself, in the manner of Samuel’s checkers program) to find the good moves. This permits you to shift to thinking about chessboard positions, possible chess moves, and the grounds for evaluating them—to shift to reasoning about reasons.

Adopting the intentional stance towards the configuration, you could predict its future *as* a chess-player performing intentional actions—making chess moves and trying to achieve checkmate. First you would have to figure out the interpretation scheme that permits you to say which configurations of pixels count as which symbols: which glider pattern spells out “QxBch” (Queen takes Bishop; check) and the other symbols for chess moves. But then you could use the interpretation scheme to predict, for instance, that the next configuration to emerge from the galaxy would be such-and-such a glider stream—say, the symbols for “RxQ” (Rook takes Queen). There is risk involved, because the chess program being run on the Turing machine may be far from perfectly rational, and, at a different level, debris may wander onto the scene and “break” the Turing-machine configuration before it finishes the game. But if all goes well, as it normally will, if you have the right interpretation, you can astonish your friends by saying something like “I predict that the next stream of gliders to emerge in location L in this Life galaxy will have the following pattern: a singleton, followed by a group of three, followed by another singleton . . .” How on Earth were you able to predict that that particular “molecular” pattern would appear then?<sup>2</sup>

In other words, real but (potentially) noisy patterns abound in such a configuration of the Life world, there for the picking up if only you are lucky or clever enough to hit on the right perspective. They are not *vi-*

---

2. In case you wondered, I imagined “RxQ” to be spelled out in Morse code, and “R” in Morse is dot-dash-dot—the group of three gliders counts as a dash.



*sual* patterns but, you might say, *intellectual* patterns. Squinting or twisting your head in front of the computer screen is not apt to help, whereas posing fanciful interpretations (or what Quine would call “analytical hypotheses”) may uncover a gold mine. The opportunity confronting the observer of such a Life world is analogous to the opportunity confronting the cryptographer staring at a new patch of cipher text, or the opportunity confronting the Martian peering through a telescope at the Superbowl Game. If the Martian hits on the intentional stance—otherwise known as folk psychology<sup>3</sup>—as the right level to look for pattern, shapes will readily emerge through the noisy jostling of people-particles and team-molecules.

The scale of compression when one adopts the intentional stance towards the two-dimensional chess-playing computer galaxy is stupendous: it is the difference between figuring out in your head what White’s most likely (best) chess move is versus calculating the state of a few trillion pixels through a few hundred thousand generations. But the scale of the savings is really no greater in the Life world than in our own. Predicting that someone will duck if you throw a brick at him is easy from the intentional or folk-psychological stance; it is and will always be intractable if you have to trace the photons from brick to eyeball, the neurotransmitters from optic nerve to motor nerve, and so forth.

For such vast computational leverage one might be prepared to pay quite a steep price in errors, but in fact the intentional stance, used correctly, provides a description system that permits extremely reliable prediction of not only intelligent human behavior, but also the “intelligent behavior” of the process that designed organisms. All this would warm William Paley’s heart. We can put the burden of proof on the skeptics with a simple challenge argument: if there weren’t design in the biosphere, how come the intentional stance *works*? We can even get a rough measure of the design in the biosphere by comparing the cost of making predictions from the lowest-level physical stance (which assumes no design—well, almost no design, depending on how we treat the evolution of universes) with the cost of making predictions from the higher stances: the design stance and the intentional stance. The added leverage of prediction, the diminution of uncertainty, the shrinkage of the huge search space to a few optimal or

---

3. I introduced the term “folk psychology” in 1978 (Dennett 1981, 1987b) as the name for the natural, perhaps even partly innate, talent human beings have for adopting the intentional stance. See Baron-Cohen 1995 for a fascinating contribution to the current state of play. There is more agreement among philosophers and psychologists about the existence of the talent than there is about my analysis of it. See, for instance, the recent anthologies on the topic—Greenwood 1991, and Christensen and Turner 1993. See Dennett 1987b, 1990b, and 1991b for my account.

near-optimal paths, is a measure of the design that is observable in the world.

The biologists' name for this style of reasoning is *adaptationism*. It is defined by one of its most eminent critics as the "growing tendency in evolutionary biology to reconstruct or predict evolutionary events by *assuming* that all characters are established in evolution by direct natural selection of the most adapted state, that is, the state that is an optimum 'solution' to a 'problem' posed by the environment" (Lewontin 1983). These critics claim that, although adaptationism plays *some* important role in biology, it is not really all that central or ubiquitous—and, indeed, we should try to balance it with other ways of thinking. I have been showing, however, that it plays a crucial role in the analysis of every biological event at every scale from the creation of the first self-replicating macromolecule on up. If we gave up adaptationist reasoning, for instance, we would have to give up the best textbook argument for the very occurrence of evolution (I quoted Mark Ridley's version of it on page 136): the widespread existence of homologies, those suspicious similarities of design that are *not* functionally necessary.

Adaptationist reasoning is not optional; it is the heart and soul of evolutionary biology. Although it may be supplemented, and its flaws repaired, to think of *displacing* it from central position in biology is to imagine not just the downfall of Darwinism but the collapse of modern biochemistry and all the life sciences and medicine. So it is a bit surprising to discover that this is precisely the interpretation that many readers have placed on the most famous and influential critique of adaptationism, Stephen Jay Gould and Richard Lewontin's oft-cited, oft-reprinted, but massively misread classic, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme" (1979).

## 2. THE LEIBNIZIAN PARADIGM

*If, among all the possible worlds, none had been better than the rest, then God would never have created one.*

—GOTTFRIED WILHELM LEIBNIZ 1710

*The study of adaptation is not an optional preoccupation with fascinating fragments of natural history, it is the core of biological study.*

—COLIN PITTENDRIGH 1958, p. 395

Leibniz, notoriously, said that this was the best of all possible worlds, a striking suggestion that might seem preposterous from a distance, but turns out, as we have seen, to throw an interesting light on the deep questions of

what it is to be a possible world, and on what we can infer about the actual world from the fact of its actuality. In *Candide*, Voltaire created a famous caricature of Leibniz, Dr. Pangloss, the learned fool who could rationalize any calamity or deformity—from the Lisbon earthquake to venereal disease—and show how, no doubt, it was all for the best. Nothing *in principle* could prove that this was not the best of all possible worlds.

Gould and Lewontin memorably dubbed the *excesses* of adaptationism the “Panglossian Paradigm,” and strove to ridicule it off the stage of serious science. They were not the first to use “Panglossian” as a term of criticism in evolutionary theory. The evolutionary biologist J. B. S. Haldane had a famous list of three “theorems” of bad scientific argument: the Bellman’s Theorem (“What I tell you three times is true”; from “The Hunting of the Snark” by Lewis Carroll), Aunt Jobisca’s Theorem (“It’s a fact the whole world knows”; from Edward Lear, “The Pobble Who Had No Toes”), and Pangloss’s Theorem (“All is for the best in this best of all possible worlds”; from *Candide*). John Maynard Smith then used the last of these more particularly to name “the old Panglossian fallacy that natural selection favours adaptations that are good for the species as a whole, rather than acting at the level of the individual.” As he later commented, “It is ironic that the phrase ‘Pangloss’s theorem’ was first used in the debate about evolution (in print, I think, by myself, but borrowed from a remark of Haldane’s), not as a criticism of adaptive explanations, but specifically as a criticism of ‘group-selectionist’, mean-fitness-maximising arguments” (Maynard Smith 1988, p. 88). But Maynard Smith is wrong, apparently. Gould has recently drawn attention to a still earlier use of the term by a biologist, William Bateson (1909), of which he, Gould, had been unaware when he chose to use the term. As Gould (1993a, p. 312) says, “The convergence is hardly surprising, as Dr. Pangloss is a standard synecdoche for this form of ridicule.” As we saw in chapter 6, the more apt or fitting a brainchild is, the more likely it is to be born (or borrowed) independently in more than one brain.

Voltaire created Pangloss as a parody of Leibniz, and it is exaggerated and unfair to Leibniz—as all good parody is. Gould and Lewontin similarly caricatured adaptationism in their article attacking it, so parity of reasoning suggests that, if we wanted to undo the damage of that caricature, and describe adaptationism in an accurate and constructive way, we would have a title ready-made: we could call adaptationism, fairly considered, the “Leibnizian Paradigm.”

The Gould and Lewontin article has had a curious effect on the academic world. It is widely regarded by philosophers and other humanists who have heard of it or even read it as some sort of *refutation of adaptationism*. Indeed, I first learned of it from the philosopher/psychologist Jerry Fodor, a lifelong critic of my account of the intentional stance, who pointed out that

what I was saying was pure adaptationism (he was right about that), and went on to let me in on what the *cognoscenti* all knew: Gould and Lewontin's article had shown adaptationism "to be completely bankrupt." (For an instance of Fodor's view in print, see Fodor 1990, p. 70.) When I looked into it, I found out otherwise. In 1983, I published a paper in *Behavioral and Brain Sciences*, "Intentional Systems in Cognitive Ethology," and since it was unabashedly adaptationist in its reasoning, I included a coda, "The 'Panglossian Paradigm' Defended," which criticized both Gould and Lewontin's paper and—more particularly—the bizarre myth that had grown up around it.

The results were fascinating. Every article that appears in *BBS* is accompanied by several dozen commentaries by experts in the relevant fields, and my piece drew fire from evolutionary biologists, psychologists, ethologists, and philosophers, most of it friendly but some remarkably hostile. One thing was clear: it was not just some philosophers and psychologists who were uncomfortable with adaptationist reasoning. In addition to the evolutionary theorists who weighed in enthusiastically on my side (Dawkins 1983b, Maynard Smith 1983), and those who fought back (Lewontin 1983), there were those who, though they agreed with me that Gould and Lewontin had not refuted adaptationism, were eager to downplay the standard use of optimality assumptions that I claimed to be an essential ingredient in all evolutionary thinking.

Niles Eldredge (1983, p. 361) discussed the reverse engineering of functional morphologists: "You will find sober analyses of fulcra, force vectors and so forth: the understanding of anatomy as a living machine. Some of this stuff is very good. Some of it is absolutely dreadful." He went on to cite, as an example of good reverse engineering, the work of Dan Fisher (1975) comparing modern horseshoe crabs with their Jurassic ancestors:

Assuming only that Jurassic horseshoe crabs also swam on their backs, Fisher showed they must have swum at an angle of 0–10 degrees (flat on their backs) and at the somewhat greater speed of 15–20 cm/sec. Thus the 'adaptive significance' of the slight differences in anatomy between modern horseshoe crabs and their 150-million-year-old relatives is translated into an understanding of their slightly different swimming capabilities. (In all honesty, I must also report that Fisher does use optimality in his arguments: He sees the differences between the two species as a sort of trade-off, where the slightly more efficient Jurassic swimmers appear to have used the same pieces of anatomy to burrow somewhat less efficiently than their modern-day relatives). In any case, Fisher's work stands as a really good example of functional morphological analysis. The notion of adaptation is naught but conceptual filigree—one that may have played a role in motivating the research, but one that was not vital to the research itself. [Eldredge 1983, p. 362.]

But in fact the role of optimality assumptions in Fisher's work—beyond the explicit role that Eldredge conceded—is so “vital” and indeed omnipresent that Eldredge entirely overlooked it. For instance, Fisher's inference that the Jurassic crabs swam at 15–20 cm/sec has as a tacit premise that those crabs *swam at the optimal speed for their design*. (How does he know they swam at all? Perhaps they just lay there, oblivious of the excess functionality of their body shapes.) Without this tacit (and, of course, dead obvious) premise, no conclusion at all could be drawn about what the *actual* swimming speed of the Jurassic variety was.

Michael Ghiselin (1983, p. 363) was even more forthright in denying this unobvious obvious dependence:

Panglossianism is bad because it asks the wrong question, namely, What is good? . . . The alternative is to reject such teleology altogether. Instead of asking, What is good? we ask, What has happened? The new question does everything we could expect the old one to do, and a lot more besides.

He was fooling himself. There is hardly a single answer to the question “What has happened (in the biosphere)?” that doesn't depend crucially on assumptions about what is good.<sup>4</sup> As we just noted, you can't even avail yourself of the concept of a homology without taking on adaptationism, without taking the intentional stance.

So now what is the problem? It is the problem of how to tell good—irreplaceable—adaptationism from bad adaptationism, how to tell Leibniz from Pangloss.<sup>5</sup> Surely one reason for the extraordinary influence of Gould

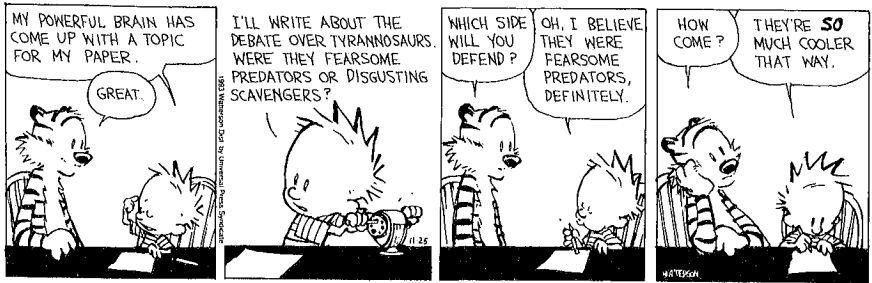
---

4. Doesn't my assertion fly in the face of the claims of those cladists who purport to deduce history from a statistical analysis of shared and unshared “characters”? (For a philosophical survey and discussion, see Sober 1988.) Yes, I guess it does, and my review of their arguments (largely via Sober's analyses) shows me that the difficulties they create for themselves are largely if not entirely due to their trying so hard to find non-adaptationist ways of drawing the sound inferences that are dead obvious to adaptationists. For instance, those cladists who abstain from adaptation talk cannot just help themselves to the obvious fact that having webbed feet is a pretty good “character” and having dirty feet (when examined) is not. Like the behaviorists who pretended to be able to explain and predict “behavior” defined in the starkly uninterpreted language of geographical trajectory of body parts, instead of using the richly functionalistic language of searching, eating, hiding, chasing, and so forth, the abstemious cladists create majestic edifices of intricate theory, which is amazing, considering they do it with one hand tied behind their backs, but strange, considering that they wouldn't have to do it at all if they didn't insist on tying one hand behind their backs. (See also Dawkins 1986a, ch. 10, and Mark Ridley 1985, ch. 6.)

5. The myth that the point of the Gould and Lewontin paper was to destroy adaptationism, not correct its excesses, was fostered by the paper's rhetoric, but in some quarters it backfired on Gould and Lewontin, since adaptationists themselves tended to pay more

## Calvin and Hobbes

by Bill Watterson



CALVIN AND HOBBS copyright 1993 Watterson. Reprinted with permission of UNIVERSAL PRESS SYNDICATE. All rights reserved.

FIGURE 9.3

and Lewontin's paper (among nonevolutionists) is that it expressed, with many fine rhetorical flourishes, what Eldredge called the "backlash" against the concept of adaptationism among biologists. What were they reacting against? In the main, they were reacting against a certain sort of laziness: the adaptationist who hits upon a truly nifty explanation for why a particular circumstance should prevail, and then never bothers to test it—because it is too good a story, presumably, not to be true. Adopting another literary label, this time from Rudyard Kipling (1912), Gould and Lewontin call such explanations "Just So Stories." It is an enticing historical curiosity that Kipling wrote his *Just So Stories* at a time when this objection to Darwinian explanation had already been swirling around for decades;<sup>6</sup> forms of it were raised by some of Darwin's earliest critics (Kitcher 1985a, p. 156). Was Kipling inspired by the controversy? In any case, calling the adaptationists' flights of imagination "Just So Stories" hardly does them credit; as delightful as I have always found Kipling's fantasies about how the elephant got its trunk, and the leopard got its spots, they are quite simple and unsurprising tales compared with the amazing hypotheses that have been concocted by adaptationists.

Consider the greater honey guide, *Indicator indicator*, an African bird that owes its name to its talent for leading human beings to wild beehives hidden in the forest. When the Boran people of Kenya want to find honey, they call for the bird by blowing on whistles made of sculpted snail shells. When a bird arrives, it flies around them, singing a special song—its "follow-

---

attention to the rhetoric than the arguments: "The critique by Gould and Lewontin had little impact on practitioners, perhaps because they were seen as hostile to the whole enterprise, and not merely to careless practise of it" (Maynard Smith 1988, p. 89).

6. Kipling began publishing the individual stories in 1897.

me" call. They follow as the bird darts ahead and waits for them to catch up, always making sure they can see where it's heading. When the bird reaches the hive, it changes its tune, giving the "here-we-are" call. When the Boran locate the beehive in the tree and break into it, they take the honey, leaving wax and larvae for the honey guide. Now, don't you ache to believe that this wonderful partnership actually exists, and has the clever functional properties described? Don't you want to believe that such a marvel could have evolved under some imagined series of selection pressures and opportunities? I certainly do. And, happily, in this case, the follow-up research is confirming the story, and even adding nifty touches as it does so. Recent controlled tests, for instance, showed that the Boran honey-hunters took much longer to find hives without the help of the birds, and 96 percent of the 186 hives found during the study were encased in trees in ways that would have made them inaccessible to the birds without human assistance (Isack and Reyer 1989).

Another fascinating story, which strikes closer to home, is the hypothesis that our species, *Homo sapiens*, descended from earlier primates via an intermediate species that was aquatic (Hardy 1960, Morgan 1982, 1990)! These aquatic apes purportedly lived on the shores of an island formed by the flooding of the area that is now in Ethiopia, during the late Miocene, about seven million years ago. Cut off by the flooding from their cousins on the African continent, and challenged by a relatively sudden change in their climate and food sources, they developed a taste for shellfish, and over a period of a million years or so they began the evolutionary process of returning to the sea that we know was undergone earlier by whales, dolphins, seals, and otters, for instance. The process was well under way, leading to the fixation of many curious characteristics that are otherwise found *only* in aquatic mammals—not in any other primate, for example—when circumstances changed once again, and these semi-seagoing apes returned to a life on the land (but typically on the shore of sea, lake, or river). There, they found that many of the adaptations they had developed for good reasons in their shell-diving days were not only not valuable but a positive hindrance. They soon turned these handicaps to good uses, however, or at least made compensations for them: their upright, bipedal posture, their subcutaneous layer of fat, their hairlessness, perspiration, tears, inability to respond to salt deprivation in standard mammalian ways, and, of course, the diving reflex—which permits even newborn human infants to survive sudden submersion in water for long periods with no ill effects. The details—and there are many, many more—are so ingenious, and the whole aquatic-ape theory is so shockingly antiestablishment, that I for one would *love* to see it vindicated. That does not make it true, of course.

The fact that its principal exponent these days is not only a woman, Elaine Morgan, but an amateur, a science writer without proper official credentials

in spite of her substantial researches, makes the prospect of vindication all the more enticing.<sup>7</sup> The establishment has responded quite ferociously to her challenges, mostly treating them as beneath notice, but occasionally subjecting them to withering rebuttal.<sup>8</sup> This is not necessarily a pathological reaction. Most uncredentialed proponents of scientific “revolutions” are kooks who really are not worth paying any attention to. There really are a lot of them besieging us, and life is too short to give each uninvited hypothesis its proper day in court. But in this case, I wonder; many of the counterarguments seem awfully thin and *ad hoc*. During the last few years, when I have found myself in the company of distinguished biologists, evolutionary theorists, paleo-anthropologists, and other experts, I have often asked them just to tell me, please, exactly why Elaine Morgan must be wrong about the aquatic-ape theory. I haven’t yet had a reply worth mentioning, aside from those who admit, with a twinkle in their eyes, that they have often wondered the same thing. There seems to be nothing *inherently* impossible about the idea; other mammals have made the plunge, after all. Why couldn’t our ancestors have started back into the ocean and then retreated, bearing some telltale scars of this history?

Morgan may be “accused” of telling a good story—she certainly has—but not of declining to try to test it. On the contrary, she has used the story as leverage to coax a host of surprising predictions out of a variety of fields, and has been willing to adjust her theory when the results have demanded it. Otherwise, she has stuck to her guns and, in fact, invited attack on her views through the vehemence of her partisanship. As so often happens in such a confrontation, the intransigence and defensiveness, on both sides, have begun to take their toll, creating one of those spectacles that then discourage anyone who just wants to know the truth from having anything more to do with the subject. Morgan’s latest book on the topic (1990)

---

7. Sir Alister Hardy, the Linacre Professor of Zoology at Oxford, who originally proposed the theory, could hardly have been a more secure member of the scientific establishment, however.

8. For instance, there is no mention at all of the aquatic-ape theory, not even to dismiss it, in two recent coffee-table books that include chapters on human evolution. Philip Whitfield’s *From So Simple a Beginning: The Book of Evolution* (1993) offers a few paragraphs on the standard savanna theory of bipedalism. “The Primates’ Progress,” by Peter Andrews and Christopher Stringer, is a much longer essay on hominid evolution, in *The Book of Life* (Stephen Jay Gould, ed., 1993b), but it, too, ignores the aquatic-ape theory—the AAT. And, adding insult to oblivion, there has also been a wickedly funny parody of it by Donald Symons (1983), exploring the radical hypothesis that our ancestors used to *fly*—“The *flying on air* theory—FLOAT, as it is acronomously (acrimoniously, among the reactionary human evolution ‘establishment’).” For an overview of the reactions, see G. Richards 1991.



responded with admirable clarity, however, to the objections that had been lodged to date, and usefully contrasted the strengths and weaknesses of the aquatic-ape theory to those of the establishment's history. And, more recently still, a book has appeared that collects essays by a variety of experts, for and against the aquatic-ape theory: Roede et al. 1991. The tentative verdict of the organizers of the 1987 conference from which that book sprang (p. 324) is that, "while there are a number of arguments favoring the AAT, they are not sufficiently convincing to counteract the arguments against it." That judicious note of mild disparagement helps ensure that the argument will continue, perhaps even with less rancor; it will be interesting to see where it all comes out.

My point in raising the aquatic-ape theory is not to defend it against the establishment view, but to use it as an illustration of a deeper worry. Many biologists would like to say, "A pox on both your houses!" Morgan (1990) deftly exposes the hand-waving and wishful thinking that have gone into the establishment's tale about how—and *why*—*Homo sapiens* developed bipedalism, sweating, and hairlessness on the savanna, not the seashore. Their stories may not be literally as fishy as hers, but some of them are pretty farfetched; they are every bit as speculative, and (I venture to say) no better confirmed. What they mainly have going for them, so far as I can see, is that they occupied the high ground in the textbooks before Hardy and Morgan tried to dislodge them. Both sides are indulging in adaptationist Just So Stories, and since *some story or other* must be true, we must not conclude we have found *the* story just because we have come up with *a* story that seems to fit the facts. To the extent that adaptationists have been less than energetic in seeking further confirmation (or dreaded disconfirmation) of their stories, this is certainly an excess that deserves criticism.<sup>9</sup>

But before leaving it at that, I want to point out that there are many adaptationist stories that *everybody* is happy to accept even though they

---

9. The geneticist Steve Jones (1993, p. 20) gives us another case in point: There are more than three hundred strikingly different species of cichlid fish in Lake Victoria. They are so different; how did they get there? "The conventional view is that Lake Victoria must once have dried up into many small lakes to allow each species to evolve. Apart from the fish themselves, there is no evidence that this ever happened." Adaptationist stories *do* get disconfirmed and abandoned, however. My favorite example is the now-discredited explanation of why certain sea turtles migrate all the way across the Atlantic between Africa and South America, spawning on one side, feeding on the other. According to this all-too-reasonable story, the habit started when Africa and South America were first beginning to split apart; at that time, the turtles were just going across the bay to spawn; the distance grew imperceptibly longer over the eons, until their descendants dutifully cross an ocean to get to where their instinct still tells them to spawn. I gather that the timing of the breakup of Gondwanaland turns out not to match the evolutionary timetable for the turtles, sad to say, but wasn't it a cute idea?

have never been "properly tested," just because they are too obviously true to be worth further testing. Does anybody seriously doubt that eyelids evolved to protect the eye? But that very obviousness may hide good research questions from us. George Williams points out that concealed behind such obvious facts may lie others that are well worth further investigation:

A human eye blink takes about 50 milliseconds. That means that we are blind about 5% of the time when we are using our eyes normally. Many events of importance can happen in 50 milliseconds, so that we might miss them entirely. A rock or spear thrown by a powerful adversary can travel more than a meter in 50 milliseconds, and it could be important to perceive such motion as accurately as possible. Why do we blink with both eyes simultaneously? Why not alternate and replace 95% visual attentiveness with 100%? I can imagine an answer in some sort of trade-off balance. A blink mechanism for both eyes at once may be much simpler and cheaper than one that regularly alternates. [G. Williams 1992, pp. 152–53.]

Williams has not himself yet attempted to confirm or disconfirm any hypothesis growing out of this exemplary piece of adaptationist problem-setting, but he has called for the research by asking the question. It would be as pure an exercise in reverse engineering as can be imagined.

Serious consideration of why natural selection permits simultaneous blinking might yield otherwise elusive insights. What change in the machinery would be needed to produce the first step towards my envisioned adaptive alternation or simple independent timing? How might the change be achieved developmentally? What other changes would be expected from a mutation that produced a slight lag in the blinking of one eye? How would selection act on such a mutation? [G. Williams 1992, p. 153.]

Gould himself has endorsed some of the most daring and delicious of adaptationist Just So Stories, such as the argument by Lloyd and Dybas (1966) explaining why cicadas (such as "seventeen-year locusts") have reproductive cycles that are prime-numbered years long—thirteen years, or seventeen, but never fifteen or sixteen, for instance. "As evolutionists," Gould says, "we seek answers to the question, why. Why, in particular, should such striking synchronicity evolve, and why should the period between episodes of sexual reproduction be so long?" (Gould 1977a, p. 99).<sup>10</sup>

---

10. Gould has recently (1993a, p. 318) described his antiadaptationism as the "zeal of the convert," and elsewhere (1991b, p. 13) confesses, "I sometimes wish that all copies of *Ever Since Darwin* would self-destruct," so perhaps he would recant these words today, which would be a pity, since they eloquently express the rationale of adaptationism.

The answer—which makes beautiful sense, in retrospect—is that, by having a large prime number of years between appearances, the cicadas minimize the likelihood of being discovered and later tracked as a predictable feast by predators who themselves show up every two years, or three years, or five years. If the cicadas had a periodicity of, say, sixteen years, then they would be a rare treat for predators who showed up every year, but a more reliable source of food for predators who showed up every two or four years, and an even-money gamble for predators that got in phase with them on an eight-year schedule. If their period is not a multiple of any lower number, however, they are a rare treat—not worth “trying” to track—for any species that isn’t lucky enough to have exactly their periodicity (or some multiple of it—the mythical Thirty-four-Year Locust-Muncher would be in fat city). I don’t know whether Lloyd and Dybas’ Just So Story has been properly confirmed yet, but I don’t think Gould is guilty of Panglossianism in treating it as established until proven otherwise. And if he really wants to ask and answer “why” questions, he has no choice but to be an adaptationist.

The problem he and Lewontin perceive is that there are no standards for when a particular bit of adaptationist reasoning is too much of a good thing. How serious, really, is this problem even if it has no principled “solution”? Darwin has taught us not to look for essences, for dividing lines between *genuine* function or *genuine* intentionality and mere *on-its-way-to-being* function or intentionality. We commit a fundamental error if we think that if we want to indulge in adaptationist thinking we need a license and the only license could be the possession of a strict definition of or criterion for a genuine adaptation. There are good rules of thumb to be followed by the prospective reverse engineer, made explicit years ago by George Williams (1966). (1) Don’t invoke adaptation when other, lower-level, explanations are available (such as physics). We don’t have to ask what advantage accrues to maple trees that explains the tendency of their leaves to fall *down*, any more than the reverse engineers at Raytheon need to hunt for a reason why GE made their widgets so that they would melt readily in blast furnaces. (2) Don’t invoke adaptation when a feature is the outcome of some general developmental requirement. We don’t need a special reason of increased fitness to explain the fact that heads are attached to bodies, or limbs come in pairs, any more than the people at Raytheon need to explain why the parts in GE’s widget have so many edges and corners with right angles. (3) Don’t invoke adaptation when a feature is a by-product of another adaptation. We don’t need to give an adaptationist explanation of the capacity of a bird’s beak to groom its feathers (since the features of the

---

Gould’s attitude towards adaptationism is not so easily discerned, however. *The Book of Life* (1993b) is packed with adaptationist reasoning that made it past his red pencil, and thus presumably has his endorsement.

bird's beak are there for more pressing reasons), any more than we need a special explanation of the capacity of the GE widget's casing to shield the innards from ultraviolet rays.

But you will already have noticed that in each case these rules of thumb can be overridden by a more ambitious inquiry. Suppose someone marveling at the brilliant autumn foliage in New England asks *why* the maple leaves are so vividly colored in October. Isn't this adaptationism run amok? Shades of Dr. Pangloss! The leaves are the colors they are simply because once the summer energy-harvest season is over, the chlorophyll vanishes from the leaves, and the residual molecules have reflective properties that happen to determine the bright colors—an explanation at the level of chemistry or physics, not biological purpose. But wait. Although this may have been the only explanation that was true up until now, today it is true that human beings so prize the autumn foliage (it brings millions of tourist dollars to northern New England each year) that they protect the trees that are brightest in autumn. You can be sure that if you are a tree competing for life in New England, there is now a selective advantage to having bright autumn foliage. It may be tiny, and in the long run it may never amount to much (in the long run, there may be no trees at all in New England, for one reason or another), but this is how all adaptations get their start, after all, as fortuitous effects that get opportunistically picked up by selective forces in the environment. And of course there is also an adaptationist explanation for why right angles predominate in manufactured goods, and why symmetry predominates in organic limb-manufacturing. These may become utterly fixed traditions, which would be almost impossible to dislodge by innovation, but the reasons why *these* are the traditions are not hard to find, or controversial.

Adaptationist research always leaves unanswered questions open for the next round. Consider the leatherback sea turtle and her eggs:

Near the end of egg laying, a variable number of small, sometimes misshapen eggs, containing neither embryo nor yolk (just albumin) are deposited. Their purpose is not well understood, but they become desiccated over the course of incubation and may moderate humidity or air volume in the incubation chamber. (It is also possible that they have no function or are a vestige of some past mechanisms not apparent to us today.) [Eckert 1992, p. 30.]

But where does it all end? Such open-endedness of adaptationist curiosity is unnerving to many theorists, apparently, who wish there could be stricter codes of conduct for this part of science. Many who have hoped to contribute to clearing up the controversy over adaptationism and its backlash have despaired of finding such codes, after much energy has been expended

in drawing up and criticizing various legislative regimes. They are just not being Darwinian enough in their thinking. Better adaptationist thinking soon drives out its rivals by normal channels, just as second-rate reverse engineering betrays itself sooner or later.

The eskimo face, once depicted as 'cold engineered' (Coon *et al.*, 1950) becomes an adaptation to generate and withstand large masticatory forces (Shea, 1977). We do not attack these newer interpretations; they may all be right. We do wonder, though, whether the failure of one adaptive explanation should always simply inspire a search for another of the same general form, rather than a consideration of alternatives to the proposition that each part is 'for' some specific purpose. [Gould and Lewontin 1979, p. 152.]

Is the rise and fall of successive adaptive explanations of various things a sign of healthy science constantly improving its vision, or is it like the pathological story-shifting of the compulsive fibber? If Gould and Lewontin had a serious alternative to adaptationism to offer, their case for the latter verdict would be more persuasive, but although they and others have hunted around energetically, and promoted their alternatives boldly, none has yet taken root.

Adaptationism, the paradigm that views organisms as complex adaptive machines whose parts have adaptive functions subsidiary to the fitness-promoting function of the whole, is today about as basic to biology as the atomic theory is to chemistry. And about as controversial. Explicitly adaptationist approaches are ascendant in the sciences of ecology, ethology, and evolution because they have proven essential to discovery; if you doubt this claim, look at the journals. Gould and Lewontin's call for an alternative paradigm has failed to impress practicing biologists both because adaptationism is successful and well-founded, and because its critics have no alternative research program to offer. Each year sees the establishment of such new journals as *Functional Biology* and *Behavioral Ecology*. Sufficient research to fill a first issue of *Dialectical Biology* has yet to materialize. [Daly 1991, p. 219.]

What particularly infuriates Gould and Lewontin, as the passage about the Eskimo face suggests, is the blithe confidence with which adaptationists go about their reverse engineering, always sure that sooner or later they will find *the reason* why things are as they are, even if it so far eludes them. Here is an instance, drawn from Richard Dawkins' discussion of the curious case of the flatfish (flounders and soles, for instance) who when they are born are vertical fish, like herring or sunfish, but whose skulls undergo a weird twist-

ing transformation, moving one eye to the other side, which then becomes the top of the bottom-dwelling fish. Why didn't they evolve like those other bottom-dwellers, skates, which are not on their side but on their belly, "like sharks that have passed under a steam roller" (Dawkins 1986a, p. 91)? Dawkins *imagines* a scenario (pp. 92–93):

... even though the skate way of being a flat fish might *ultimately* have been the best design for bony fish too, the would-be intermediates that set out along this evolutionary pathway apparently did less well in the short term than their rivals lying on their side. The rivals lying on their side were so much better, in the short term, at hugging the bottom. In genetic hyperspace, there is a smooth trajectory connecting free-swimming ancestral bony fish to flatfish lying on their side with twisted skulls. There is not a smooth trajectory connecting these bony fish ancestors to flatfish lying on their belly. There is such a trajectory in theory, but it passes through intermediates that would have been—in the short term, which is all that matters—unsuccessful if they had ever been called into existence.

Does Dawkins *know* this? Does he know that the postulated intermediates were less fit? Not because he has seen any data drawn from the fossil record. This is a purely theory-driven explanation, argued *a priori* from the assumption that natural selection tells us the true story—some true story or other—about every curious feature of the biosphere. Is that objectionable? It does "beg the question"—but what a question it begs! It assumes that Darwinism is basically on the right track. (Is it objectionable when meteorologists say, begging the question against supernatural forces, that there must be a purely physical explanation for the birth of hurricanes, even if many of the details so far elude them?) Notice that in this instance, Dawkins' explanation is almost certainly right—there is nothing especially daring about that particular speculation. Moreover, it is, of course, exactly the sort of thinking a good reverse engineer should do. "It seems so obvious that this General Electric widget casing ought to be made of two pieces, not three, but it's made of three pieces, which is wasteful and more apt to leak, so we can be damn sure that three pieces was seen as better than two in somebody's eyes, shortsighted though they may have been. Keep looking!" The philosopher of biology Kim Sterelny, in a review of *The Blind Watchmaker*, made the point this way:

Dawkins is admittedly giving only scenarios: showing that it's *conceivable* that (e.g.) wings could evolve gradually under natural selection. Even so, one could quibble. Is it really true that natural selection is so fine-grained that, for a protostick insect, looking 5% like a stick is better than looking 4% like one? (pp. 82–83). A worry like this is especially pressing because

Dawkins' adaptive scenarios make no mention of the costs of allegedly adaptive changes. Mimicry might deceive potential mates as well as potential predators. . . . Still, I do think this objection is something of a quibble because essentially I agree that natural selection is the only possible explanation of complex adaptation. So something like Dawkins' stories have got to be right. [Sterelny 1988, p. 424.]<sup>11</sup>

### 3. PLAYING WITH CONSTRAINTS

*It is just as foolish to complain that people are selfish and treacherous as it is to complain that the magnetic field does not increase unless the electric field has a curl.*

—JOHN VON NEUMANN, quoted in William Poundstone 1992, p. 235

*As a general rule today a biologist seeing one animal doing something to benefit another assumes either that it is manipulated by the other individual or that it is being subtly selfish.*

—GEORGE WILLIAMS 1988, p. 391

One may nevertheless be reasonably nervous about the size of the role of sheer, unfettered imagination in adaptationist thinking. What about butterflies with tiny machine guns for self-protection? This fantastic example is often cited as the sort of option that can be dismissed without detailed analysis by adaptationists seeking to describe the ensemble of possible butterfly adaptations from which Mother Nature has chosen the best, all things considered. It is just too distant a possibility in design space to be taken seriously. But as Richard Lewontin (1987, p. 156) aptly notes, "My guess is that if fungus-gardening ants had never been seen, the suggestion that this was a reasonable possibility for ant evolution would have been regarded as silly." Adaptationists are masters of the retrospective rationale, like the

---

11. Dawkins is not content to rest with Sterelny's dismissal of his own objections as "quibbles" since, he points out (personal communication), they raise an important point often misunderstood: "It is not up to individual humans like Sterelny to express their own commonsense scepticism of the proposition that 5% like a stick is significantly better than 4%. It is an easy rhetorical point to make: 'Come on, are you really trying to tell me that 5% like a stick really matters when compared to 4%?' This rhetoric will often convince laymen, but the population genetic calculations (e.g. by Haldane) belie common sense in a fascinating and illuminating way: because natural selection works on genes distributed over many individuals and over many millions of years, human actuarial intuitions are over-ruled."

chess-player who only notices *after* he's made the move that it forces check-mate in two moves. "How brilliant—and I almost thought of it!" But before we decide that this is a *flaw* in adaptationist character or method, we should remind ourselves that this retrospective endorsement of brilliance is the way Mother Nature herself always operates. Adaptationists should hardly be faulted for being unable to predict the brilliant moves that Mother Nature herself was oblivious of until she'd stumbled upon them.

The perspective of game-playing is ubiquitous in adaptationism, where mathematical *game theory* has played a growing role ever since its introduction into evolutionary theory by John Maynard Smith (1972, 1974).<sup>12</sup> Game theory is yet one more fundamental contribution to twentieth-century thinking from John von Neumann.<sup>13</sup> Von Neumann created game theory in collaboration with the economist Oskar Morgenstern, and it grew out of their realization that *agents* make a fundamental difference to the complexity of the world.<sup>14</sup> Whereas a lone "Robinson Crusoe" agent can view all problems as seeking stable maxima—hill-climbing on Mount Fuji, if you like—as soon as other (maxima-seeking) agents are included in the environment, strikingly different methods of analysis are required:

A guiding principle cannot be formulated by the requirement of maximizing two (or more) functions at once. . . . One would be mistaken to believe that it can be obviated . . . by a mere recourse to the devices of the theory of probability. Every participant can determine the variables which de-

---

12. Maynard Smith built his game-theory applications to evolution on the foundations already laid by R. A. Fisher (1930). One of Maynard Smith's many more recent contributions was showing Stuart Kauffman that he was, after all, a Darwinian, not an anti-Darwinian (see Lewin 1992, pp. 42–43).

13. I sometimes wonder if there is any important advance in thinking in the second half of this century that von Neumann is *not* the father of. The computer, the model of self-replication, game theory—and if that weren't enough, von Neumann also made major contributions to quantum physics. For what it is worth, however, I suspect that his formulation of the measurement problem in quantum mechanics is his one bad idea, a sleight-of-hand endorsement of a fundamentally Cartesian model of conscious observation that has bedeviled quantum mechanics ever since. My student Turhan Canli first opened this door in his (undergraduate!) term paper for me on the problem of Schrödinger's cat, in which he developed the sketch of an alternative formulation of quantum physics in which time is quantized. If I ever master the physics (a very remote prospect, sad to say), I will tackle this hunch, which might extend in wildly ambitious ways my theory of consciousness (1991a); more likely, however, is the prospect that I will be a semi-comprehending but enthusiastic spectator of this development, wherever it leads.

14. For a fascinating account of the history of game theory and its relation to nuclear disarmament, see William Poundstone's 1992 book, *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*.



scribe his own actions but not those of the others. Nevertheless those 'alien' variables cannot, from his point of view, be described by statistical assumptions. This is because the others are guided, just as he himself, by rational principles—whatever that may mean—and no *modus procedendi* can be correct which does not attempt to understand those principles and the interactions of the conflicting interests of all participants. [Von Neumann and Morgenstern 1944, p. 11.]

The fundamental insight that unites game theory and evolutionary theory is that the "rational principles—whatever that may mean" that "guide" agents in competition can exert their influence even on such unconscious, unreflective semi-agents as viruses, trees, and insects, because the stakes and payoff possibilities of competition determine which lines of play cannot help winning or losing if adopted, however mindlessly they are adopted.

The best-known example in game theory is the Prisoner's Dilemma, a simple two-person "game" which casts shadows, both obvious and surprising, into many different circumstances in our world. Here it is in basic outline (excellent detailed discussions of it are found in Poundstone 1992 and Dawkins 1989a). You and another person have been imprisoned pending trial (on a trumped-up charge, let's say), and the prosecutor offers each of you, separately, the same deal: if you both hang tough, neither confessing nor implicating the other, you will each get a short sentence (the state's evidence is not that strong); if you confess and implicate the other and he hangs tough, you go scot free and he gets life in prison; if you both confess-and-implicate, you both get medium-length sentences. Of course, if you hang tough and the other person confesses, he goes free and you get life. What should you do?

If you both could hang tough, defying the prosecutor, this would be much better for the two of you than if you both confess, so couldn't you just promise each other to hang tough? (In the standard jargon of the Prisoner's Dilemma, the hang-tough option is called *cooperating*.) You could promise, but you would each then feel the temptation—whether or not you acted on it—to *defect*, since then you would go scot free, leaving the *sucker*, sad to say, in deep trouble. Since the game is symmetrical, the other person will be just as tempted, of course, to make a sucker of you by defecting. Can you risk life in prison on the other person's keeping his promise? Probably safer to defect, isn't it? That way, you definitely avoid the worst outcome of all, and might even go free. Of course, the other fellow will figure this out, too, if it's such a bright idea, so he'll probably play it safe and defect, too, in which case you *must* defect to avoid calamity—unless you are so saintly that you don't mind spending your life in prison to save a promise-breaker!—so you'll both wind up with medium-length sentences. If only you could overcome this reasoning and cooperate!

The logical structure of the game is what matters, not this particular setting, which is a usefully vivid imagination-driver. We can replace the prison sentences with positive outcomes (it's a chance to win different amounts of cash—or, say, descendants) just so long as the payoffs are symmetrical, and ordered so that lone defection pays more than mutual cooperation, which pays each more than mutual defection does, which in turn pays more than the sucker payoff one gets when the other is a lone defector. (And in formal settings we set a further condition: the average of the sucker and mutual-defection payoffs must not be greater than the mutual-cooperation payoff.) Whenever this structure is instantiated in the world, there is a Prisoner's Dilemma.

Game-theoretic explorations have been undertaken in many fields, from philosophy and psychology to economics and biology. The most influential of the many applications of game-theoretic thinking to evolutionary theory is Maynard Smith's concept of an *evolutionarily stable strategy*, or ESS, a strategy that may not be "best" from any Olympian (or Fujian!) standpoint, but is unimprovable-upon and unsubvertible under the circumstances. Maynard Smith (1988, especially chh. 21 and 22) is an excellent introductory account of game theory in evolution. The revised edition of Richard Dawkins' *The Selfish Gene* (1989a) has a particularly good account of the development of ESS thinking in biology during the last decade or so, when large-scale computer simulations of various game-theoretic models revealed complications that had been overlooked by the earlier, less realistic versions.

I now like to express the essential idea of an ESS in the following more economical way. An ESS is a strategy that does well against copies of itself. The rationale for this is as follows. A successful strategy is one that dominates the population. Therefore it will tend to encounter copies of itself. Therefore it won't stay successful unless it does well against copies of itself. This definition is not so mathematically precise as Maynard Smith's, and it cannot replace his definition because it is actually incomplete. But it does have the virtue of encapsulating, intuitively, the basic ESS idea. [Dawkins 1989a, p. 282.]

There can be no doubt that game-theoretic analyses work in evolutionary theory. Why, for instance, are the trees in the forest so tall? For the very same reason that huge arrays of garish signs compete for our attention along commercial strips in every region of the country! Each tree is looking out for itself, and trying to get as much sunlight as possible.

If only those redwoods could get together and agree on some sensible zoning restrictions and stop competing with each other for sunlight, they

could avoid the trouble of building those ridiculous and expensive trunks, stay low and thrifty shrubs, and get just as much sunlight as before! [Dennett 1990b, p. 132.]

But they can't get together; under these circumstances, defection from any cooperative "agreement" is bound to pay off if ever or whenever it occurs, so trees would be stuck with the "tragedy of the commons" (Hardin 1968) if there weren't an essentially inexhaustible supply of sunshine. The tragedy of the commons occurs when there is a finite "public" or shared resource that individuals will be selfishly tempted to take more of than their fair share—such as the edible fish in the oceans. Unless very specific and enforceable agreements can be reached, the result will tend to be the destruction of the resource. Many species, in many regards, face various sorts of Prisoner's Dilemmas. And we human beings face them both consciously and unconsciously—sometimes in ways that we might never have imagined without the aid of adaptationist thinking.

*Homo sapiens* is not exempt from the sort of genetic conflict David Haig postulates to explain genomic imprinting; in an important new article (1993) he analyzes a variety of conflicts that exist between the genes of a pregnant woman and the genes of her embryo. It is in the embryo's interests, of course, that the mother bearing it stay strong and healthy, for its own survival depends on her not only completing her term of pregnancy but tending for her newborn. However, if the mother, in her attempt to stay healthy under trying circumstances—famine, for instance, which must have been a common circumstance in most generations of human existence—should cut down on the nutrition she provides her embryo, at some point this becomes more of a threat to the embryo's survival than the alternative, a weakened mother.

If the embryo were "given a choice" between being spontaneously aborted early in the pregnancy or being stillborn or of low birth weight on the one hand, versus being born at normal weight of a weak or even dying mother on the other, what would (selfish) reason dictate? It would dictate taking whatever steps are available to try to ensure that the mother does not cut her losses (she can always try to have another child later, when the famine is over), and this is just what the embryo does. Both embryo and mother can be entirely oblivious of this conflict—as oblivious as the trees rising competitively in the forest. The conflict plays out in the genes and their control of hormones, not in the brains of mother and embryo; it is the same sort of conflict we saw between maternal and paternal genes in the mouse. There is a flood of hormones; the embryo produces a hormone that will enhance its own growth at the expense of the mother's nutritional needs; her body responds with an antagonist hormone that attempts to undo the effect of the first; and so on, in an escalation that can produce

hormone levels many times higher than normal. This tug-of-war usually ends in a mutually semi-satisfactory standoff, but it produces a host of by-products that would be utterly baffling and senseless were they not the predictable effects of such conflict. Haig concludes with an application of the fundamental game-theoretic insight: "Maternal and fetal genes would both benefit if a given transfer of resources was achieved with a lesser production of . . . hormones and less maternal resistance, but such an agreement is evolutionarily unenforceable" (Haig 1993, p. 518).

This is not, in many regards, welcome news. Von Neumann's all-too-casual remark on the inevitability of human selfishness epitomizes the Darwinian mind-set that many people view with loathing, and it is not hard to see why. They fear that Darwinian "survival of the fittest" would *entail* that people are nasty and selfish. Isn't that just what von Neumann is saying? No. Not quite. He is saying that it is indeed entailed by Darwinism that such virtues as cooperation should be *in general* "evolutionarily unenforceable" and hence hard to come by. If cooperation and the other unselfish virtues are to exist, *they must be designed*—they do not come for free. They *can* be designed under special circumstances. (See, for instance, Eshel 1984, 1985, and Haig and Grafen 1991.) After all, the eukaryotic revolution that made multicelled organisms possible was a revolution that began when an enforceable truce was somehow engineered between certain prokaryotic cells and their bacterial invaders. They found a way of joining forces and submerging their selfish interests.

Cooperation and the other virtues are, in general, rare and special properties that can only emerge under very particular and complex R-and-D circumstances. We might contrast the Panglossian Paradigm, then, with the Pollyannian Paradigm, which cheerfully assumes, with Pollyanna, that Mother Nature is Nice.<sup>15</sup> In general, she isn't—but that isn't the end of the world. Even in the present case, we can see that there are other perspectives to adopt. Aren't we really rather fortunate, for instance, that trees are so insuperably selfish? The beautiful forests—to say nothing of the beautiful wooden sailing ships and the clean white paper on which we write our poetry—could not exist if trees weren't selfish.

There can be no doubt, as I say, that game-theoretic analyses work in evolutionary theory, but do they *always* work? Under what conditions do they apply, and how can we tell when we are overstepping? Game-theory calculations always assume that there is a certain range of "possible" moves, from which the selfish-by-definition contestants make their choices. But how realistic is this *in general*? Just because a move in a particular circumstance is the move that *reason dictates*, is it the move nature will always

---

15. For a powerful antidote to the Pollyannian Paradigm, see G. Williams 1988.

take? Isn't this Panglossian optimism? (As we have just seen, this sometimes looks more like Panglossian *pessimism*: "Darn—organisms are 'too smart' to cooperate!"<sup>16</sup>)

The standard assumption of game theory is that there will always be mutations that have the "right" phenotypic effects to rise to the occasion, but what if the right move just doesn't "occur to Mother Nature"? Is this ever or often very likely? We certainly know of cases in which Mother Nature *does* take the move—to make the forests, for instance. Are there perhaps just as many (or more) cases in which some sort of hidden constraint prevents this from happening? There may well be, but in every such case, adaptationists will want to persist by asking the next question: And *is there a reason* in this case why Mother Nature doesn't take the move, or is it just a brute, unthinking constraint on Mother Nature's rational gamesmanship?

Gould has suggested that a fundamental flaw of adaptationist reasoning is the assumption that in every fitness landscape, the way is always shown as clear to the tops of the various summits, but there might well be hidden constraints, rather like railroad tracks lying across the landscape. "The constraints of inherited form and developmental pathways may so channel any change that even though selection induces motion down permitted paths, the channel itself represents the primary determinant of evolutionary direction" (Gould 1982a, p. 383). Populations, then, do not get to spread *ad lib* across the terrain, but are forced to stay on the tracks, as in figure 9.4.

Suppose this is true. Now, how do we locate the hidden constraints? It is all very well for Gould and Lewontin to point to the possibility of hidden constraints—every adaptationist already acknowledges this as an omnipresent possibility—but we need to consider what methodology might be best for discovering them. Consider a curious variation on a standard practice in chess.

When a stronger player plays a weaker opponent in friendly matches, the stronger player often volunteers to take on a handicap, to make the game more evenly matched and exciting. The standard handicap is to give up a piece or two—to play with only one bishop or one rook, or, in a really extreme case, to play without a queen. But here is another handicapping system that might have interesting results. Before the match, the stronger player writes down on a piece of paper a hidden constraint (or constraints)

---

16. The Panglossian pessimist says, "Isn't it a shame that this is, after all, the best of all possible worlds!" Imagine a beer commercial: As the sun sets over the mountains, one of the hunks lounging around the campfire intones, "It doesn't get any better than this!"—at which point his beautiful companion bursts into tears: "Oh no! Is that really true?" It wouldn't sell much beer.

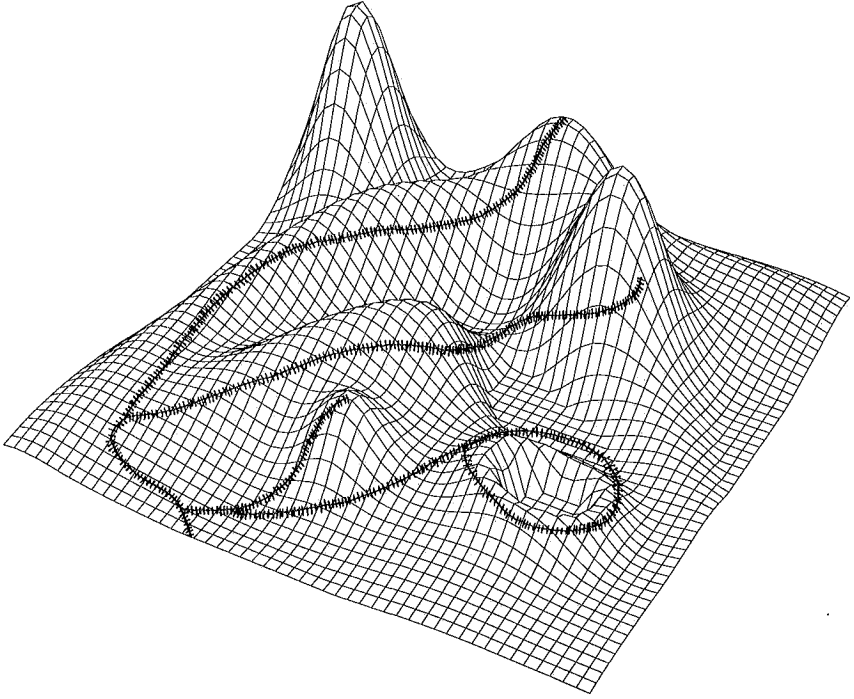


FIGURE 9.4

that she will undertake to play under, and hides the paper under the board. What is the difference between a constraint and a forced move? Reason dictates a forced move—and will always dictate it, again and again—whereas some frozen bit of history dictates a constraint, whether or not there was a reason for its birth, and whether or not there is a reason for or against it now. Here are a few of the possible constraints:

Unless I am forced by the rules to do so (because I am in check, and am obliged to play whatever legal move escapes check),

- (1) I may never move the same piece on two consecutive turns.
- (2) I may not castle.
- (3) I may capture with pawns only three times in the whole game.
- (4) My queen must move only in rook fashion, never diagonally.

Now imagine the epistemological predicament of the weaker player, who *knows* his opponent is playing with hidden constraints but doesn't know what they are. How should he proceed? The answer is quite obvious: he should play *as if* all the *apparently* possible moves—all the legal moves—

are available to her, and adjust his strategy only when evidence begins to mount that she is actually bound not to take what otherwise would be the obviously best move.

Such evidence is not at all easy to gather. If you think your opponent cannot move her queen diagonally, you might test that hypothesis by the risky tactic of offering a free capture to that queen on the diagonal. If the queen declines, that counts in favor of your hypothesis—*unless* there is a deeper reason of strategy (unimagined as yet by you) for declining the capture. (Remember Orgel's Second Rule: Evolution is cleverer than you are.)

Of course, another way of learning the hidden constraints at the chess-board is to peek at the paper, and one might think that what Gould and Lewontin are recommending is that adaptationists simply abandon their game-playing and go for the truth via a more direct examination of the molecular evidence. Unfortunately, this analogy is mistaken. You are certainly entitled to use whatever data-gathering tricks are available in the game of science, but when you peek at the molecules, all you find there is more machinery, more design (or apparent design) in need of reverse engineering. Nowhere are Mother Nature's hidden constraints *written down* in a way that can be read without the help of the interpretive rules of artifact hermeneutics (Dennett 1990b). The descent to the deeper level of the DNA, for instance, is indeed a valuable way of vastly improving one's investigative acuity—though usually at the intolerable cost of drowning in too much data—but in any case it is not an alternative to adaptationism; it is an extension of it.

The example of playing chess with hidden constraints lets us see a profound difference between Mother Nature and human chess-players that does have implications, I think, for a widespread foible in adaptationist thinking. If *you* were playing chess under hidden constraints, you would adjust your strategy accordingly. Knowing that you had secretly promised not to move your queen diagonally, you would probably forgo any campaign that put your queen at risk of capture thanks to her unusual limitation—although of course you could take a chance, hoping your weak opponent wouldn't notice the possibility. But you have knowledge of the hidden constraints, and foresight. Mother Nature does not. Mother Nature has no reason to avoid high-risk gambits; she takes them all, and shrugs when most of them lose.

Here is how the idea applies in evolutionary thinking. Suppose we notice that a particular butterfly has protective coloration on its wings that uncannily mimics the pattern of colors on the forest floor where it lives. We chalk that up as a fine adaptation, camouflage, which it undoubtedly is. This butterfly does better than its cousins *because* its coloration so perfectly reproduces the coloration of the forest floor. But there is a temptation, routinely

succumbed to, to add, implicitly or explicitly: "And what's more, if the forest floor had any other color pattern on it, the butterfly would look like *that* pattern instead!" That is uncalled for. It may well not be true. It could even be, in the limit, that this is the *only* sort of forest floor that this lineage of butterfly could mimic with much success; if the forest floor were much different, this lineage would just not be here—never forget about the importance in evolution of bait-and-switch. If the forest floor changes, what will happen? Will the butterfly automatically adapt? All we can say is that either it will adapt by changing its camouflage or it won't! If it doesn't, then either it will find some other adaptation in its limited kit of available moves, or it will soon disappear.

The limiting case, in which exactly one path was ever open to explore, is an instance of our old nemesis actualism: only the actual was possible. Such straitjacketed explorations of the space of (apparent) possibility are not ruled out, I am saying, but they must be the exception, not the rule. If they were the rule, Darwinism would be defunct, utterly incapable of explaining any of the (apparent) design in the biosphere. It would be as if you wrote a chess-playing computer program that could just play one game by rote (say, Alekhine's moves in the famous Flamberg-Alekhine match in Mannheim in 1914) and, *mirabile dictu*, it regularly won against all competition! This would be a "pre-established harmony" of miraculous proportions, and would make a mockery of the Darwinian claim to have an explanation of how the "winning" moves have been found.

But our dismissal of actualism should not tempt us to err in the other direction, supposing that the space of real possibilities is much more densely populated than it actually is. The temptation, when we think about phenotypic variation, is to adopt a sort of Identikit tactic of assuming that all the minor variations we can imagine on the themes we find in actuality are truly available. Carried to extremes, this tactic will always vastly—Vastly—overestimate what is actually possible. If the *actual* Tree of Life occupies Vanishingly narrow threads through the Library of Mendel, the *actually possible* Tree of Life is itself some rather bushier but still far from dense partial filling of the *apparently possible*. We have already seen that the Vast space of all imaginable phenotypes—Identikit Space, we might call it—no doubt includes huge regions for which there are no recipes in the Library of Mendel. But even along the paths through which the Tree of Life wanders, we are not guaranteed that the neighboring regions of Identikit Space are actually all accessible.<sup>17</sup>

---

17. Gould is fond of pointing out the mistake of looking back in time and seeing "lineages" where we should be seeing "bushes"—including all the failures that have left no descendants. I am pointing out a contrary sort of mistake: imagining dense (or even



If hidden constraints guarantee that there is a largely invisible set of maze walls—or channels or railroad tracks—in the space of apparent possibility, then “you can’t get there from here” is true much more often than we might imagine. Even if this is so, we still can do no better in our exploration of this possibility than to play out our reverse-engineering strategies at every opportunity, at every level. It is important not to overestimate the actual possibilities, but it is even more important not to underestimate them, an equally common foible, though not one that adaptationists typically manifest. Many adaptationist arguments are of the if-it’s-possible-it-will-happen variety: cheats will emerge to invade the saints; or an arms race will ensue until such-and-such a first-order adaptive stability is achieved, etc. These arguments presuppose that enough of the space of possibilities is “habitable” to ensure that the process approximates the game-theory model used. But are these assumptions always appropriate? Will these bacteria mutate into a form that is resistant to our new vaccine? Not if we’re lucky, but we’re better off assuming the worst—namely, that there are, in the space actually accessible to these bacteria, countermoves in the arms race our medical innovation has set in motion (Williams and Nesse 1991).

*CHAPTER 9: Adaptationism is both ubiquitous and powerful in biology. Like any other idea, it can be misused, but it is not a mistaken idea; it is in fact the irreplaceable core of Darwinian thinking. Gould and Lewontin’s fabled refutation of adaptationism is an illusion, but they have raised everybody’s consciousness about the risks of incautious thinking. Good adaptationistic thinking is always on the lookout for hidden constraints, and in fact is the best method for uncovering them.*

*CHAPTER 10: The view of Darwinian thinking presented so far in this book has been challenged, repeatedly, by Stephen Jay Gould, whose influential writings have contributed to a seriously distorted picture of evolutionary biology among both lay people and philosophers and scientists in other fields. Gould has announced several different “revolutionary” abridgments of orthodox Darwinism, but they all turn out to be false alarms. There is a pattern to be discerned in these campaigns: Gould, like eminent evolutionary thinkers before him, has been searching for skyhooks to limit the power of Darwin’s dangerous idea.*

---

continuous) bushes of unactualized possibility where in fact there may be rather sparse twigs creating paths to relatively isolated outposts in the huge space of apparent possibilities.