Dennett 2 The power OF adaptationist thinking

> *'Naked as Nature intended' was a persuasive slogan of the early Naturalist 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. BEABDSLEY 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 designer 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.

[figure diagram of the wheel-work of the Antikythera mechanism – handout]

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.

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

[figure]

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 199 la, 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 color 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 colors 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 Prison'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 morphine!) 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 Sheriock-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 post-fertilization 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 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 un-illuminating 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 with 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 tkinking 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?

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 *visual* 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 Super-bowl Game. If the Martian hits on the intentional stance—otherwise known as folk psychology—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 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 *asking* 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).