Foreword

Daniel C. Dennett

This joyful book tells the story of how meaning came into existence, and how we ourselves came to be able to make sense of our world. It blithely ignores hostile boundaries and unites philosophy and science, poetry and biochemistry, Shannon's mathematical theory of information and good old-fashioned literary scholarship. What could you possibly learn from Aristotle or Francis Bacon about the dynamics of gene regulation, and what could you possibly learn about literary interpretation from the role of retroviruses in rewiring placental regulatory networks? Do scientists have anything to gain by learning the history of their own fields, and can scholars in the humanities and social scientists strengthen their grip on *their* own fields by learning some of the fascinating details of microbiology? Having aroused the passions of partisans on both side of this great divide, I want now to invite them on a roller-coaster ride of ideas, dexterously drawn by David Haig from an astonishing range of sources, from Laurence Sterne's curious novel *Tristram Shandy* (1759) and Immanuel Kant's *Critique of the Power of Judgment* (1790–1793) through the work of Charles Sanders Peirce and Jacques Derrida to the latest peer-reviewed articles on the intricate machinery of gene transmission, duplication, and expression. The ride is all

the more exciting because it is in the form of elegant arguments conducted with calm confidence, unfolding a novel vision of life without fanfare or special pleading, letting the implications speak for themselves.

Philosophers, according to a threadbare stereotype, concern themselves with the old question:

What is the meaning of life?

And "hard" scientists, according to another threadbare stereotype, set that question aside, postponing it indefinitely, while they work on a more respectable set of questions, ultimately *physical* or *mechanical* questions about *how* things work, not *why*:

What is matter?

What is time?

How do molecules do what they do?

How did life arise and how do living things stay alive?

These stereotypes have been baked into powerful presumptions, separating the arts and humanities, with their goals and methods, from the sciences, with their differing (*and opposing*) goals and methods. *Geisteswissenschaften* (literally, sciences of the mind) versus *Naturwissenschaften* (literally, sciences of nature—in the broadest sense, which includes moons and mountains, oxygen and oceans). Lawless narratives to the left, laws of nature to the right.

These traditional dividing lines have been taught for centuries, and are still taught today, but they *have* to be taught, since we all know better! We all know that there are *reasons why* the intricate organs of living things are the way they are—if only we can discern them—and we all know that the meanings we find in our arts and humanities are real phenomena, certainly as real

as the meanings scientists find and transmit in their equations. There *must be* a way of putting matter and meaning, mechanism and purpose, causation and information, into a unified perspective.

And for over a century and a half we have understood, at least dimly, that the key to this unification lies in Darwin's dangerous idea of evolution by natural selection. In very different ways this has become common ground for many deep thinkers from Darwin's own day to a growing assembly of biologists, physicists, psychologists, and philosophers in the twenty-first century. Aristotle's Final Cause—the *telos* or *what for?* question—has been absorbed, somehow, by Darwin's proposal that life is designed by a blind, *purposeless* process of differential replication, from which emerge products—living things and their artifacts—which themselves have purposes. But how is this possible? There is still an unresolved tension, epitomized by the oft-quoted exultation by Karl Marx when he discovered Darwin's *Origin of Species*:

> Not only is a death blow dealt here for the first time to "teleology" in the natural sciences but their rational meaning is empirically explained. (Marx, letter to LaSalle, 1861)

Does Darwin's idea explain teleology, or explain it away? It does both; it shows how real purposes, real functions can arise out of phenomena that themselves have no purpose, no function. But in the eyes of many, this is a dubious assertion, wishful thinking at best, or self-deception. Aristotle's *telos* is a seductive idea but it smells . . . *divine* and hence *anathema*. Should we puritanically abstain from teleology when doing science (but not when doing literature, history, philosophy, psychology), or has Darwin shown us how we can tame Aristotle's shrew, and make an honest woman of her? (This obstreperous or maybe even offensive way of putting the issue is itself an issue to be considered; the

submerged distastes and unarticulated biases of thinkers across the spectrum is exposed in the bright beam of Haig's widely informed attention.)

The details matter, more than one might think—more than I had thought until I got swept up in Haig's consummate explanation of how the pieces all fit together. Many of us think we understand evolutionary theory, if not in every minute particular at least in handy if sometimes sketchy outline, and we have tended to think that this is enough understanding to secure a firm foundation for our extensions of evolutionary thinking beyond gene pools and organisms to minds, cultures, and societies. Haig shows that there is much more of value to be gleaned from the intricacies, if we adopt his careful *adaptationist* perspective, treating all the phenomena of life as susceptible to reverse engineering, seeking the underlying reasons for all the patterns we find in nature. Stephen Jay Gould and Richard Lewontin (1979) famously warned us about the polluting influence of this "Panglossian Paradigm" (where we supposedly live in the best of all possible worlds) but we must cast aside our residual worries about the risks of this "Darwinian paranoia" as some have called it, and let Haig lead us into the strange and enchanting (but also disconcerting) world of strategic genes, mindless competitions between renegades and impostors, team players and sentries, robots made of robots made of robots managing to pilot their enormous vehicles into the future without ever having to recognize or appreciate the reasons why they act as they do. This is the world of selfish genes, brilliantly introduced by Richard Dawkins (1976), explored in even more detail by Haig.

Two informal rules can help us place Haig's project on a useful, if provisional, foundation: Braitenberg's Law and the Warden's Rule. Valentino Braitenberg, a Swiss neuroscientist, published

an elegant little book, *Vehicles* (1984), in which he described an increasingly complex collection of imaginary organisms, simpler than bacteria (which are far from simple!) and described how they would behave, singly and in groups. The justification for this methodology was Braitenberg's *Law of Uphill Analysis and Downhill Synthesis*. It is easier, he noted, to predict the behavior of a complex entity you synthesize out of simpler entities—downhill synthesis, a constructivist approach—than to analyze the inner workings of a complex entity whose behavior you observe—uphill analysis. He was right, in the opinion of many, and his little book set in motion a considerable research tradition in robotics and other computer-aided inquiries. But something like Braitenberg's Law is also at work in Haig's analysis of the many ways that information gets captured and moved around in the process of natural selection: start small, and build up.

The *Warden's Rule*, if it *can* happen, it *will* happen (discussed by me in *Freedom Evolves*, 2003, 160–161), is something I picked up from somewhere—I'm sorry to say I have source amnesia for this. It is an improvement on "Murphy's Law," that whatever can go wrong will go wrong. It is an improvement because there are real circumstances where a version of it is true, not merely an amusing expression of pessimism. It is supposedly the maxim of every prison warden. Its rationale is that prisons are full of prisoners with time on their hands and a measure of patience and an inquisitive and competitive nature. They will exhaustively search every aspect of the system in which they are incarcerated looking for a way out, or a way of improving their condition, however arduous the path. How clever are the prisoners? Some are brilliant, others just dogged; if they share their discoveries it doesn't much matter who gets credit for the innovations. Instead of testing them for intelligence, test their environments

for *opportunities*. (In *Elbow Room* [1984], I described the existence of a bag of jewels in a trashcan a few feet away as a *bare opportunity*—if only I had known about it, I could have been rich, but I had no information at all that would lead me to check that trashcan for treasure.) According to the Warden's Rule, even bare opportunities must be counted, since prisoners have nothing better to do than hunt for them. No trashcans go untested in a prison.

Bare opportunities abound in the world of evolution, and it doesn't take intelligence to stumble across them; it just takes lots of time for variations to arise and get tried. If there is *information* in an evolutionary environment, a difference that *could* make a difference to something that could detect or just respond to that difference, it *will* make a difference (generally—this is a rule of thumb), a difference that in turn can be amplified by reproduction, making a bigger difference and, in a recursive process, making the detection of that difference more likely in the future, and so forth. Yes, this is just another way of explaining how natural selection works, but it works this way *all the way down*, so that even simple molecular structures can be *usefully seen* to be like those persistent prisoners, ever delving, ever seeking an advantage that they can exploit to further their own "interests." The prisoners have a single overarching goal: *Escape!* These agents have a different goal: *Reproduce!*

When you put these two perspectives together, you get the methodology that bears such delicious fruit in the work of David Haig. Braitenberg's Law invites us to work bottom up, starting with very simple agents, material genes, or even their elements, and treating them as dead stupid *homunculi*—as stupid as we can muster—with just one goal: reproduce! As François Jacob

famously put it, "The dream of every cell is to become two cells," but we can apply his vision one level (or more) lower than he did. When we do this, we recognize that individual tokens of a gene type are rather like a gang of brothers in prison. They are kin-altruistic *or* we can consider the siblings as a unit, a *selfish* unit. For many Warden-like purposes, it is easier to consider the opportunities available to a selfish gang. Their competitive nature is not so much an urge driving them from the inside (they are just parts of macromolecules, after all) as a systematic talent for taking advantage of opportunities if and when they arise. The reason we have to look at the *strategic* gene is that these gangs do have to work together. The copies in the germ line—the eggs and sperm—get to escape (and multiply, if they are lucky) thanks to the "efforts" of their brothers in the soma or body.

So these families of little robots are like the simplest Braitenberg vehicles, and thanks to the law of downhill synthesis we can plot the opportunities that might arise, and the effect of a bunch of such vehicles "discovering" them. Where Haig speaks about what a genetic element can "expect" he is talking about the epistemic predicament facing these robots: is there *any* information in their immediate vicinity that they could ("in principle") exploit to their advantage? Are there bare opportunities that they could discover by chance and then act on? Could they "recognize" an opportunity? They could in the only way such a robot can recognize anything, by trying it as a random stab and being "rewarded" by a benefit. If these opportunities are systematically available, then they are part of what the gang can expect.

But even if we grant, for the sake of argument, or as a crutch for our imaginations, that we can interpret such mindless things

as agents for some purposes, where does *real* agency kick in? How do we get from molecules to masterminds, from selfish genes to selfish (and altruistic) people? Haig tells us that "conscious intentions are special cases of a pervasive intentionality of living things," and he has a lot to say about how these cases are special and how they might arise. Finding illumination in Adam Smith's often-ignored masterpiece, *The Theory of Moral Sentiments* (1759), Haig constructs a version of morality that is distinctly Darwinian but a far cry from some of the simplistic versions of evolution-based ethical thinking currently in vogue. (In other words, any who have been unimpressed by applications of Darwinian thinking to ethics and political theory in relatively crude recent examples should give Haig a chance to salvage the perspective.) "Do I *feel* you feel I am somebody who can be trusted? Do I *feel* you feel I am somebody who can be exploited?" Haig builds a subtly interactive confection of genes, culture, and reason to provide an explanation of how "Integrity is born out of prudence." Not by a miracle, but step by step, in a gradual series of moves to greater and greater complexity, more degrees of freedom to be controlled.

Evolutionary biology has been blessed with many fine expositors, starting of course with Darwin himself. My admiration and gratitude to those fine authors has been often expressed and will not be repeated here, to save the reader from an honor roll that would either run on for pages or neglect somebody whose work has been particularly illuminating to me. David Haig stands out, even in this illustrious company, as a bountiful source of novel insights, a clarifier—and often, in my opinion, resolver—of controversies, a scholar both scrupulous and playful, who reminds me, as well as anybody alive can remind me, how glorious it is to *think* and *understand*.

References

Marx, Karl. 1861/1942. Letter to Lasalle, London, January 16, 1861. In *Gesamtausgabe*. New York: International Publishers.

Braitenberg, Valentino. 1984. *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.

Dawkins, R. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

Dennett, D. C. 1984. *Elbow Room*. Cambridge, MA: MIT Press.

Dennett, D. C. 2003. *Freedom Evolves*. London: Penguin Books.

Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Philosophical Transactions of the Royal Society B* 205:581–598.

Smith, A. 1976. *The Theory of Moral Sentiments*. Oxford: Oxford University Press.