RICHARD DAWKINS

HOW A SCIENTIST CHANGED THE WAY WE THINK

Reflections by scientists, writers, and philosophers

Edited by

ALAN GRAFEN

AND

MARK RIDLEY

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2006

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by RefineCatch Ltd., Bungay, Suffolk
Printed in Great Britain by
Clays Ltd., St Ives plc

ISBN 0–19–929116–0 978–0–19–284055–8

1 3 5 7 9 10 8 6 4 2

DEEP COMMONALITIES BETWEEN LIFE AND MIND

Steven Pinker

US television talk-show host Jay Leno, interviewing a passerby: How do you think Mount Rushmore was formed?
Passerby: Erosion?
Leno: Well, how do you think the rain knew to not only pick four presidents—but four of our *greatest* presidents? How did the rain know to put the beard on Lincoln and not on Jefferson?
Passerby: Oh, just luck, I guess.

I AM a cognitive scientist, someone who studies the nature of intelligence and the workings of the mind. Yet one of my most profound scientific influences has been Richard Dawkins, an evolutionary biologist. The influence runs deeper than the fact that the mind is a product of the brain and the brain a product of evolution; such an influence could apply to someone who studies any organ of any organism. The significance of Dawkins' ideas, for me and many others, runs to his characterization of the very nature of life and to a theme that runs throughout his writings: the possibility of deep commonalities between life and mind.

Scientists, unlike literary scholars, are ordinarily not a fitting subject of exegesis and thematic analysis. A scientist's writings should be transparent, revealing facts and explanations directly. Yet I find that Dawkins' ideas repay close reflection and reexamination, not because he is a guru issuing enigmatic pronouncements for others to ponder, but because he continually

engages the deepest problems in biology, problems that continue to challenge our understanding.

When I first read Dawkins I was immediately gripped by concerns in his writings on life that were richer versions of ones that guided my thinking on the mind. The parallels concerned both the content and the practice of the relevant sciences.

The first critical theme is an attention to adaptive complexity as the paramount phenomenon in need of an explanation, most forcibly expressed in *The Blind Watchmaker* and *Climbing Mount Improbable*. In the case of life, we have the remarkable adaptations of living things: echolocation, camouflage, the vertebrate eye, and countless other 'organs of extreme perfection and complication', in Darwin's words, which represent solutions to formidable engineering problems. In the case of mind, we have the remarkable powers of human cognition: the ability to recognize objects and materials, plan and execute movement, reason and remember, speak and understand.

I shared, moreover, Dawkins' impatience with fellow scientists who provide passable accounts of relatively peripheral aspects of their subject matter, but who, when it came to mechanistic explanations for adaptive complexity, were too easily satisfied with verbal formulae and vague hand-waving. Dawkins did not disguise his exasperation with Stephen Jay Gould's claims that he had revolutionized the theory of evolution with addenda such as punctuated equilibrium, species-selection, and exaptation. But these addenda, Dawkins pointed out, did not address the main problem of adaptive complexity in life and so left the core of the theory of natural selection (which does solve the problem) untouched. Many cognitive scientists, I often grumble, also seem to content themselves with verbal substitutes for explanatory mechanisms, such as 'strategies', 'general intelligence', 'plasticity', or 'extracting regularities'.

The discomfort with inadequate explanations of key phenomena underlies another area of resonance—the conviction that in some areas of science there is an indispensable role for the exploration of ideas, their logical adequacy, and their explanatory

power, rather than equating science with the obsessive gathering of data. Biology today, especially molecular biology, is massively weighted toward laboratory work, and any hint of theory is considered scholastic or archaic. In the case of molecular biology this attitude is particularly amnesic, because at the dawn of the field in the 1940s there was an obsession with the theoretical preconditions constraining any putative candidate for the machinery of life (as expressed, for example, in the influential treatise 'What is Life?' by Erwin Schrödinger, a theoretical physicist).

Dawkins has been unapologetic in insisting that a complete biology must lay out the implications of its theories, perhaps most forcibly in his essay 'Universal Darwinism', which audaciously argued that natural selection is not only the best theory of the evolution of life on earth, but almost certainly the best theory of the evolution of life anywhere in the universe. I believe that in cognitive science, too, the demands on adequate theories are so stringent as to carve out an essential place for theoretical analysis. In Dawkins' case, this encourages a blurring of his writing for his fellow scientists and for informed nonspecialists: his more popular books certainly cannot be considered 'popularization', nor is his most technical book, *The Extended Phenotype*, restricted to specialists. This is an example I try to emulate.

A second major theme in Dawkins' writings on life that has important parallels in the understanding of the mind is a focus on *information*. In *The Blind Watchmaker* Dawkins wrote, 'If you want to understand life, don't think about vibrant, throbbing gels and oozes, think about information technology.' Dawkins has tirelessly emphasized the centrality of information in biology—the storage of genetic information in DNA, the computations embodied in transcription and translation, and the cybernetic feedback loop that constitutes the central mechanism of natural selection itself, in which seemingly goal-oriented behavior results from the directed adjustment of some process by its recent consequences. The centrality of information was captured in the metaphor in Dawkins' book title *River Out of Eden*, the river being a flow of information in the generation-to-generation copying of

genetic material since the origin of complex life. It figured into his *Blind Watchmaker* simulations of the evolutionary process, an early example of the burgeoning field of artificial life. It also lies behind his influential theory of memes, which illustrates that the logic of natural selection applies to any replicator which carries information with a certain degree of fidelity. Dawkins' emphasis on the ethereal commodity called 'information' in an age of biology dominated by the concrete molecular mechanisms is another courageous stance. There is no contradiction, of course, between a system being understood in terms of its information content and it being understood in terms of its material substrate. But when it comes down to the deepest understanding of what life is, how it works, and what forms it is likely to take elsewhere in the universe, Dawkins implies that it is abstract conceptions of information, computation, and feedback, and not nucleic acids, sugars, lipids, and proteins, that will lie at the root of the explanation.

All this has clear parallels in the understanding of the mind. The 'cognitive revolution' of the 1950s, which connected psychology with the nascent fields of information theory, computer science, generative linguistics, and artificial intelligence, had as its central premise the idea that knowledge is a form of information, thinking a form of computation, and organized behavior a product of feedback and other control processes. This gave birth to a new science of cognition that continues to dominate psychology today, embracing computer simulations of cognition as a fundamental theoretical tool, and the framing of hypotheses about computational architecture (serial versus parallel processing, analogue versus digital computation, graphical versus list-like representations, etc.) as a fundamental source of experimental predictions. As with biology, an emphasis on information allows one to discuss cognition in a broader framework from the particular species found on earth, extending to the nature of processes we would wish to consider intelligent anywhere in the universe. And, as in biology, an emphasis on information unfortunately must withstand a strong current toward experimental studies of physical mechanisms (in this case the physiology of the brain)

accompanied by a mistrust of theory and analysis. Again there is no contradiction between studying information processing systems and studying their physical implementation, but there has been a recent tendency to downplay the former, at a cost of explanatory adequacy.

The parallel use of information-theoretic concepts in biology and cognitive science (particularly linguistics) is no secret, of course, and is evident in the reliance of genetics on a vocabulary borrowed from linguistics. DNA sequences are said to contain letters and punctuation, may be palindromic, meaningless, or synonymous, are transcribed and translated, and are even stored in libraries. Biologists occasionally describe development and physiology as following rules, most notably in the immunologist Niels Jerne's concept of the 'generative grammar of the immune system'.

A final shared theme in life and mind made prominent in Dawkins' writings is the use of mentalistic concepts in biology, most boldly in his title *The Selfish Gene*. The expression evoked a certain amount of abuse, most notoriously in the philosopher Mary Midgley's pronouncement that 'Genes cannot be selfish or unselfish, any more than atoms can be jealous, elephants abstract or biscuits teleological' (a throwback to the era in which philosophers thought that their contribution to science was to educate scientists on elementary errors of logic encouraged by their sloppy use of language). Dawkins' main point was that one can understand the logic of natural selection by imagining that the genes are agents executing strategies to make more copies of themselves. This is very different from imaging natural selection as a process that works toward the survival of the group or species or the harmony of the ecosystem or planet. Indeed, as Dawkins argued in *The Extended Phenotype*, the selfish-gene stance in many ways offers a more perspicuous and less distorting lens with which to view natural selection than the logically equivalent alternative in which natural selection is seen as maximizing the inclusive fitness of individuals. Dawkins' use of intentional, mentalistic expression was extended in later

writings in which he alluded to animals' knowing or remembering the past environments of their lineage, as when a camouflaged animal could be said to display a knowledge of its ancestors' environments on its skin.

The proper domain of mentalistic language, one might think, is the human mind, but its application there has not been without controversy either. During the reign of behaviorism in psychology in the middle decades of the twentieth century, it was considered as erroneous to attribute beliefs, desires, and emotions to humans as it would be to genes, atoms, elephants, or biscuits. Mentalistic concepts, being unobservable and subjective, were considered as unscientific as ghosts and fairies and were to be eschewed in favor of explaining behavior directly in terms of an organism's current stimulus situation and its past history of associations among stimuli and rewards. Since the cognitive revolution, this taboo has been lifted, and psychology profitably explains intelligent behavior in terms of beliefs and desires. This allows it to tap into the world of folk psychology (which still has more predictive power when it comes to day-to-day behavior than any body of scientific psychology) while still grounding it in the mechanistic explanation of computational theory.

In defending his use of mentalistic language in biological explanation, Dawkins has been meticulous in explaining that he does not impute conscious intent to genes, nor does he attribute to them the kind of foresight and flexible cleverness we are accustomed to in humans. His definitions of 'selfishness', 'altruism', 'spite', and other traits ordinarily used for humans is entirely behavioristic, he notes, and no harm will come if one remembers that these terms are mnemonics for technical concepts and heuristics for generating predictions rather than direct attributions of the human traits.

I sometimes wonder, though, whether caveats about the use of mentalistic vocabulary in biology are stronger than they need to be—whether there is an abstract sense in which we can *literally* say that genes are selfish, that they try to replicate, that they know about their past environments, and so on. Now of course

we have no reason to believe that genes have conscious experience, but a dirty secret of modern science is that we have no way of explaining the fact that *humans* have conscious experience either (conscious experience in the sense of raw first-person subjective awareness—the distinction between conscious and unconscious processes, and the nature of self-consciousness, are entirely tractable scientific topics). No one has really explained why it *feels like something* to be a hunk of neural tissue processing information in certain complex patterns. So even in the case of humans, our use of mentalistic terms does not depend on a commitment on how to explain the subjective aspects of the relevant states, but only on their functional role within a chain of computations.

Taking this to its logical conclusion, it seems to me that if information-processing gives us a good explanation for the states of knowing and wanting that are embodied in the hunk of matter called a human brain, there is no principled reason to avoid attributing states of knowing and wanting to other hunks of matter. To be specific, nothing prevents us from seeking a generic characterization of 'knowing' (in terms of the storage of usable information) that would embrace both the way in which people know things (in their case, in the patterns of synaptic connectivity in brain tissue) and the ways in which the genes know things (presumably in the sequence of bases in their DNA). Similarly, we could frame an abstract characterization of 'trying' in terms of negative feedback loops, that is, a causal nexus consisting of repeated or continuous operations, a mechanism that is sensitive to the effects of those operations on some state of the environment, and an adjustment process that alters the operation on the next iteration in a particular direction, thereby increasing the chance that that aspect of the environment will be caused to be in a given state. In the case of the human mind, the actions would be muscle movements, the effects would be detected by the senses, and the adjustments would be made by neural circuitry programming the next iteration of the movement. In the case of the evolution of genes, the actions would be extended phenotypes, the effects would be sensed as differential mortality and fecundity,

and the adjustment would be made in terms of the number of descendants resulting in the next generation.

This characterization of beliefs and desires in terms of information rather than physical incarnation may overarch not only life and mind but other intelligent systems such as machines and societies. By the same token it would embrace the various forms of intelligence implicit in the bodies of animals and plants, which we would not want to attribute either to fully human cogitation nor to the monomaniacal agenda of replication characterizing the genes. When the coloration of a viceroy butterfly fools the butterfly's predators by mimicking that of a more noxious monarch butterfly, there is a kind of intelligence being manifest. But its immediate goal is to fool the predator rather than replicate the genes, and its proximate mechanism is the overall developmental plan of the organism rather than the transcription of a single gene.

In other words the attribution of mentalistic states such as knowing and trying can be hierarchical. The genes, in order to effect their goal of making copies of themselves, can help build an organ whose goal is to fool a predator. The human mind is another intelligent mechanism built as part of the intelligent agenda of the genes, and it is the seat of a third (and the most familiar) level of intelligence: the internal simulation of possible behaviors and their anticipated consequences that makes our intelligence more flexible and powerful than the limited forms implicit in the genes or in the bodies of plants and animals. Inside the mind, too, we find a hierarchy of subgoals (to make a cup of coffee, put coffee grounds in the coffeemaker; to get coffee grounds, grind the beans; to get the beans, find the package; if there is no package, go to the store; and so on). Computer scientists often visualize hierarchies of goals as a stack, in which a program designed to achieve some goal often has to accomplish a subgoal as a means to its end, whereupon it 'pushes down' to an appropriate subroutine, and then 'pops' back up when the subroutine has accomplished the subgoal. The subroutine, in turn, can call a subroutine of its own to accomplish an even smaller

and more specialized subgoal. (The stack image comes from a memory structure that keeps track of which subroutine called which other subroutine, and works like a spring-loaded stack of cafeteria trays.) In this image, the best laid plans of mice and men are the bottom layers of the stack, and above them is the intelligence implicit in their bodies and genes, with the topmost goal being the replication of genes that makes up the core of natural selection.

It would take a good philosopher to forge bulletproof characterizations of 'intelligence', 'goal', 'want', 'try', 'know', 'selfish', 'think', and so on, that would embrace minds, robots, living bodies, genes, and other intelligent systems. (It would take an even better one to figure out how to reintroduce subjective experience into this picture when it comes to human and animal minds.) But the promise that such a characterization is possible—that we can sensibly apply mentalistic terms to biology without shudder quotes—is one of Dawkins' legacies. If so, we would have a substantive and deep explanation of our own minds, in which parochial activities like our own thinking and wanting would be seen as manifestations of more general and abstract phenomena.

The idea that life and mind are in some ways manifestations of a common set of principles can enrich the understanding of both. But it also mandates not confusing the two manifestations—not forgetting what it is (a gene? an entire organism? the mind of a person?) that knows something, or tries something, or wants something, or acts selfishly. I suspect that the biggest impediment to accepting the insights of evolutionary biology in understanding the human mind is in people's tendency to confuse the various entities to which a given mentalistic explanation may be applied.

One example is the common tendency to assume that Dawkins' portrayal of 'selfish genes' implies that organisms in general, and people in particular, are ruthlessly egoistic and self-serving. In fact nothing in the selfish-gene view predicts that this should be so. Selfish genes are perfectly compatible with selfless organisms, since the genes' goal of selfishly replicating themselves can be implemented via the sub-goal of building organisms that

are wired to do unselfish things such as being nice to relatives, extending favors in certain circumstances, flaunting their generosity under other circumstances, and so on. (Indeed much of *The Selfish Gene* consists of explanations of how the altruism of organisms is a consequence of the selfishness of genes.) Another example of this kind of confusion is the common claim that sociobiology is refuted by the many things people do that don't help to spread their genes, such as adopting children or using contraception. In this case the confusion is between the motive of genes to replicate themselves (which does exist) and the motive of people to spread their genes (which doesn't). Genes effect their goal of replication via the sub-goal of wiring people with certain goals of their own, but replication per se need not be among those sub-sub-goals: it's sufficient for people to seek sex and to nurture their children. In the environment in which our ancestors were selected, people pursuing those goals automatically helped the relevant genes pursue theirs (since sex tended to lead to babies), but when the environment changed (such as when we invented contraception) the causal chain that used to make sub-goals bring about superordinate goals no longer were in operation.

I suspect that these common fallacies arise from applying a Freudian mindset to evolutionary psychology. People conceive of the genes as the deepest, truest essence of a person, the part that harbors his deepest wishes, and think of conscious experience and overt behavior as a superficial veneer hiding these ulterior motives. This is a fallacy because the motives of the genes are entirely different from the motives of the person—they are a play within a play, not the interior monologue of a single cast of players.

More generally, I think it was the ease of confusing one level of intelligence with another that led to the proscription of mentalistic terms in behaviorism and to the phobia of anthropomorphizing organisms or genes in biology. But as long as we are meticulous about keeping genes, organisms, and brains straight, there is no reason to avoid applying common explanatory mechanisms (such as goals and knowledge) if they promise insight and explanation.

The promise of applying common tools to life and mind, and the danger in failing to distinguish which is the target of any particular explanation, can also, I think, be seen in discussions of the relevance of memes to human mind and culture. Dawkins has suggested that his discussion of memes was largely meant to illustrate the information-theoretic nature of the mechanism of natural selection—that it was not particular to DNA or carbon-based organisms or life on earth but applied to replicators of any sort. Others have treated his suggestions about memes as an actual theory of cultural change, some cleaving to a tight analogy between the selection of genes and the selection of memes, others exploring a larger family of models of cultural evolution, epidemiology, demographics, and gene-culture coevolution, I think that the mind-life parallel inherent in memetics holds out the promise of new ways of understanding cultural and historical change, but that it also poses a danger.

Many theorists, partly on the basis of Dawkins' arguments about the indispensability of natural selection in explaining complex design in living things, write as if natural selection, applied to memes rather than genes, is the only adequate explanation of complex design in human cultural achievements. To bring culture into biology, they reason, one shows how it evolved by its own version of natural selection. But that doesn't follow, because the products of evolution don't have to *look like* the process of evolution. In the case of cultural evolution they certainly don't look alike—human cultural products are not the result of an accumulation of copying errors, but are crafted through bouts of concerted brainwork by intelligent designers. And there is nothing in Dawkins' Universal Darwinism argument that makes this observation suspect. While it remains true that the origin of complex design on earth requires invoking selection (given the absence of any alternative mechanisms adequate to the task), in the case of complex design in culture we do have an alternative, namely the creative powers of the human brain. Ultimately we have to explain the complexity of the brain itself in terms of genetic selection, but then the ladder can be kicked away and

the actual process of cultural creation and transmission studied without prejudice.

A final connection. Religion has become a major theme of Dawkins' recent writings, and here too life and mind figure into the argument in related ways. The appearance of complex design in the living world was, of course, a major argument for belief in God throughout history, and a defensible one before it was undermined by the ability of natural selection to generate the appearance of design without a designer. As Dawkins wrote in *The Blind Watchmaker*, 'Although atheism might have been logically tenable before Charles Darwin, Darwin made it possible to be an intellectually fulfilled atheist.' I believe that a parallel development has taken place with regard to beliefs about the mind. The complexity of human intelligence strikes many people as compelling evidence for the existence of a soul in the same way that the complexity of life was seen as evidence for the existence of a designer. Now that intelligence may be explicable in material terms, as a kind of information processing in neural circuitry (with the circuitry itself explicable by natural selection), this second source of intuitive support for spiritual beings is being undermined. Just as evolutionary biology made it possible for intellectually fulfilled people to do without creationism, computational cognitive science makes it possible for them to do without dualism.