Daniel Dennett

Edited by

ANDREW BROOK
Carleton University

DON ROSS
University of Cape Town

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United States of America

Typefaces Janson Text 10/13 pt. and ITC Officina Sans *System* LaTeX 2ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data
Daniel Dennett / edited by Andrew Brook, Don Ross.
p. cm. – (Contemporary philosophy in focus)
Includes bibliographical references and index.
ISBN 0-521-80394-2 – ISBN 0-521-00864-6 (pbk.)
1. Dennett, Daniel Clement. I. Brook, Andrew. II. Ross, Don.
III. Series.
B945.D394 D36 2001
191 – dc21 2001037489

ISBN 0 521 80394 2 hardback
ISBN 0 521 00864 6 paperback

Contents

Contributors

KATHLEEN AKINS is associate professor of philosophy at Simon Fraser University, Canada. Her work focuses on the theoretical problems common to both philosophy of mind/metaphysics and the neurosciences. She is coordinator of the McDonnell Project in Philosophy and the Neurosciences, funded by a James S. McDonnell Centennial Fellowship. She is currently working on a book on the physiology and philosophy of colour vision with Martin Hahn.

SIMON BARON-COHEN is Co-Director of the Autism Research Centre, Departments of Experimental Psychology and Psychiatry at Cambridge University. He wrote *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press/A Bradford Book, 1995).

ANDREW BROOK is professor of philosophy, director of the Institute for Interdisciplinary Studies, and chair of the Ph.D. Programme in Cognitive Science at Carleton University, Canada. He is the author of *Kant and the Mind* (Cambridge University Press, 1994), coauthor with Robert Stainton of *Knowledge and Mind* (MIT Press/A Bradford Book, 2000) and coeditor with Don Ross and David Thompson of *Dennett's Philosophy: A Comprehensive Assessment* (MIT Press/A Bradford Book, 2000).

DOROTHY L. CHENEY is professor of biology at the University of Pennsylvania. She is coauthor with Robert Seyfarth of *How Monkeys See the World: Inside the Mind of Another Species* (University of Chicago Press, 1990).

PAUL M. CHURCHLAND is professor of philosophy at the University of San Diego. He is author of *A Neurocomputational Perspective* (1989), *The Engine of Reason, the Seat of the Soul* (1995), and coauthor with Patricia Churchland of *On the Contrary: Critical Essays 1987–1997* (1998) (all from MIT Press/Bradford Books).

ANDY CLARK is professor of philosophy and cognitive science at the University of Sussex, U.K. He has written extensively on artificial neural

networks, robotics, and artificial life. His latest books are *Being There: Putting Brain, Body and World Together Again* (MIT Press/A Bradford Book, 1997) and *Mindware: An Introduction to the Philosophy of Cognitive Science* (Oxford University Press, 2001).

RICHARD GRIFFIN is a research scientist at the Autism Research Centre's Infant-Toddler Lab, Departments of Experimental Psychology and Psychiatry, University of Cambridge. He studied philosophy and psychology at Harvard University, and was a frequent participant in Dennett's philosophy seminars at Tufts University.

DON ROSS is professor of economics and convenor of the programme in philosophy, politics, and economics at the University of Cape Town. He is the author of *Metaphor, Meaning and Cognition* (Peter Lang, 1993) and *What People Want: The Concept of Utility from Bentham to Game Theory* (UCT Press, 2001), and coeditor with Andrew Brook and David Thompson of *Dennett's Philosophy: A Comprehensive Assessment* (MIT Press/A Bradford Book, 2000). He is currently at work on a book on the relationships between economics and cognitive science.

ROBERT M. SEYFARTH is professor of psychology at the University of Pennsylvania and coauthor with Dorothy Cheney of *How Monkeys See the World: Inside the Mind of Another Species* (University of Chicago Press, 1990).

YORICK WILKS is professor of computer science and director of the Institute of Language, Speech, and Hearing at the University of Sheffield, U.K. He has published numerous articles and five books in the area of artificial intelligence and computer language processing, of which the most recent are *Artificial Believers*, coauthored with Afzal Ballim (Lawrence Erlbaum, 1991), and *Electric Words: Dictionaries, Computers and Meanings*, coauthored with Brian Slator and Louise Guthrie (MIT Press/A Bradford Book, 1996). He is editor of *Machine Conversations* (Kluwer, 1999).

1 Dennett's Position in the Intellectual World

ANDREW BROOK AND DON ROSS

1. DENNETT'S LIFE AND WORK

Over the past thirty years, Daniel Clement Dennett has had a major influence on our understanding of human intentionality and agency, consciousness (and thereby phenomenology and the architecture and neuroscience of consciousness), developmental psychology, cognitive ethology, artificial intelligence, and evolutionary theory. In this introductory essay, we will first give a chronological survey of these contributions and then, starting with Dennett's place in the intellectual history of the last half of the twentieth century, construct an overview of his philosophy. Dennett has played a central role in one of the most significant theoretical revolutions of the past fifty years, the cognitivist revolution. This revolution demolished simple empiricism and put in its stead a view of human action as requiring interpretation in terms of a rich reservoir of cognitive resources and, many argue, evolutionary history. Dennett has played a role in this revolution for thirty-five years now.

Recently, a number of collections have appeared on Dennett's work. This one is different. We are interested in the influence Dennett has had beyond the bounds of academic philosophy. To assess this influence, we have assembled a team of experts who either specialize in one of the areas in which Dennett has had such an influence (Simon Baron-Cohen in developmental psychology, Robert Seyfarth and Dorothy Cheney in cognitive ethology, Yorick Wilks in artificial intelligence) or who, although trained in philosophy, have attained expertise in a discipline beyond philosophy (in alphabetical order: Akins in neuroscience, Brook in cognitive science, Clark in cognitive science, Churchland in neuroscience, Ross in economics and evolutionary theory). We wrote the introduction as much for these contributors as for our readers. Rather than each contributor trying to summarize the aspect of Dennett's work most relevant to his or her contribution, we put together a single common overview for all to share. The various contributors then took this overview as read when they prepared

their essays. Thus, the essays should be read together with this introduction.

Although Dennett has deep roots in American pragmatism, he was actually born in Lebanon (during World War II; his father was representing the U.S. government there) and he did his doctorate at Oxford. He is a seasoned world traveller and has spent time in most of the major universities of the Western world (including, he once said with some pride, every university with a philosophy graduate programme in Canada). He did his undergraduate work at Harvard, where, he says, he vigorously resisted his teacher and the most influential American philosopher of the twentieth century, Willard van Orman Quine (doctrines for which he later developed more sympathy) (Dennett 1998, p. 357). His doctoral studies at Oxford were done under the tutelage of the most influential Oxford philosopher of his time, Gilbert Ryle. Dennett received a D.Phil. in the remarkably speedy time of two years in 1965.

Dennett's life has been as stable and unsensational as his work has been brilliant and influential. After a brief stint at the University of California Riverside, he has been at one university ever since, Tufts University in Medford, Massachusetts, a suburb of Boston. He married in university and has been married to the same woman ever since. They have two grown children. He runs a working farm in Maine. (Dennett is said to get some of his best ideas while working the land. Douglas Hofstadter calls this 'tillosophy.') He is an expert high-seas sailor and navigator and an accomplished drummer and choral singer. He is one of a very few philosophers who commands bestseller-sized advances for his books. Dennett once described himself as an ACLU (American Civil Liberties Union) liberal.

Dennett is famous for his generosity to students. A small army of increasingly important young philosophers, psychologists, and even M.D.s have passed through his Centre for Cognitive Studies at Tufts as postdoctoral fellows or visitors. Although Dennett has had a deep influence on thousands of researchers, there are few 'Dennettians.' Unlike many world-class intellectuals, Dennett has never sought to create disciples. His former students are remarkably independent-minded.

Dennett is an accomplished philosophical humourist. He was the main force behind *The Philosopher's Lexicon* (1987d), which consists of 'definitions' of philosopher's names that reflect (or at least spoof) some aspect of their work. The entry for 'Dennett' (not written by Dennett) reads in part: "Dennett, v. To while away the hours defining surnames; hence, *dennettation*, n. The meaning of a surname. 'Every surname has both a meinong and a dennettation.'"

There are many, many sides to Dennett's contributions but one of his most important contributions has been to challenge unexamined orthodoxies. One of his characteristic ways of doing so is to go after comfortable assumptions with what he calls intuition pumps. Here is an example. Prereflectively, most of us would think that there is a clear difference between how something tastes to us and how we react to that taste (with pleasure, indifference, disgust, etc.). But consider the case of Mr. Chase and Mr. Sanborn:

> Mr. Chase and Mr. Sanborn both used to like a certain coffee. More recently it has lost its appeal. The reasons they give seem to differ markedly. Chase: "The flavour of the coffee hasn't changed but I just don't like that flavour very much now." Sanborn: "No, no, you are quite wrong. I would still like *that* flavour as much as ever. The problem is that the coffee *doesn't* taste that way any more." (taken from Dennett 1988a)

We are meant to say to ourselves, "Hmmm, maybe this distinction I want to draw between how something tastes and how much one likes the taste does not correspond to a real difference." Then we are meant to generalize the doubt: "Well, if there isn't a clear boundary to be drawn here, what about with other mental states?" – And we are well on our way to shaking up our traditional philosophical conception of the mind as a place populated by a bunch of clearly demarcated mental states.

Dennett is not just a philosophical gadfly. The unsettling intuition pumps, awkward rhetorical questions, and so on, are used not just in the service of iconoclasm (although they are used in the service of iconoclasm). Dennett has a deep philosophical mission, one articulated in his very first book and carried through with verve, ingenuity, and great continuity ever since.

His first book grew out of the work he did with Ryle. It was called *Content and Consciousness* (1969). These two words, 'content' and 'consciousness,' encapsulate much of Dennett's mission. 'Content' refers to the contents of the mind: all the beliefs and desires and values and emotions and hopes and expectations and memories and . . . and . . . and . . . that make up the mental life of all cognitively intact human beings. And 'consciousness' refers, of course, to our consciousness of our world, our mental contents, and ourselves.

In Dennett's view, the correct order in which to examine these topics is content first, then consciousness. (The essays in the current volume might seem to be in the reverse order but they are not. We will explain why later.) To oversimplify mightily, he devoted the 1970s and 1980s to content

(with some important forays into consciousness) and the 1990s to consciousness (with some important additional work on content). Mentioning just these two topics may make Dennett's contribution look narrow. It is not. To the contrary, his work on consciousness has led him to study how consciousness evolved, pathologies of consciousness such as Dissociative Identity Disorder (what used to be called Multiple Personality Disorder), whether there is any real difference between how a mental state functions in us and how it feels to us (what the philosophers call, somewhat quaintly, its *qualia* or felt quality), what "selves" might be, methods for studying consciousness, how to model consciousness as a cognitive system, the nature of introspection (the consciousness we have of ourselves and our own mental states), the neural implementation of consciousness, and so on – just about every issue in connection with consciousness that one could think of. And his influence has reached to just about everybody interested in any of these issues.

Similarly with content. His work on mental content has led him to questions about artificial content (AI), the evolution of content, the relationship of content to the environment and brain (neuroscience), content in nonhumans (cognitive ethology), the nature of explanation in psychology and science generally, how content is represented and the different styles of mental representation, the relationship of representations to the brain, how we ascribe mental content to ourselves and others, and so on – all the issues alive in current work on the mind, how it evolved, and its place in the world.

His first major work, as we said, was *Content and Consciousness* (1969). It will figure prominently in the second part of this introduction. His next book was a collection of essays written during the 1970s, *Brainstorms.* This work not only brought together an extraordinarily interesting group of papers on mental content (and four on aspects of consciousness), it helped launch a unique publishing enterprise, Bradford Books. Founded by Harry and Betty Stanton and subsequently absorbed by MIT Press, the Bradford Books insignia has become one of the most important collections of books in philosophy of mind and cognitive science in the English language. Except for some trade books of the 1990s, all Dennett's books since *Brainstorms* have been published under the Bradford insignia (*Brainstorms* has been recently rereleased as a Penguin softcover).

Brainstorms begins with the first full articulation of Dennett's distinctive approach to mental content. The approach is called the intentional stance and the paper is called "Intentional Systems." Says Dennett, we can approach something in order to explain it from three stances, the physical

stance, the design stance, and the intentional stance. Each has its own advantages and costs, as we will see when we examine some of the details of the approach in Section 2.2. This mode of explanation yields impressive results when we seek to understand people's reasons for what they say and do and in other areas. The best current theory of autism, for example, is that autistic people lack the implicit notion that others are intentional systems, so their ability to adopt this stance to themselves and others is impaired or absent (Griffin and Baron-Cohen). The approach also yields impressive results in cognitive ethology (Seyfarth and Cheney). Although few economists are aware of the connection, the game-theoretic logic that underlies the intentional stance has also become one of the main approaches in economics (Ross 1).

As well as these system-anchoring reflections on content, *Brainstorms* also contained a number of papers on behaviourism and its early cognitivist replacements, a fascinating paper on AI as what you might call philosophy under the discipline of reality (you have to implement your ideas in AI) (Wilks), no less than four papers on aspects of consciousness that presaged a book still fifteen years away, and a number of papers on decision making and responsibility.

It was decision making and responsibility that Dennett turned to next, in an idiosyncratic little book called *Elbow Room: The Varieties of Free Will Worth Wanting* (1984). Beginning life as John Locke Lectures in Oxford, the book advocates a brisk compatibilism between decisions being causally determined and decisions being free in any way that is "worth wanting." For better or for worse, this book has had less influence than Dennett's other books. Interestingly, he is working on a new book on free will as we write (in the middle of 2001). People are awaiting his new thoughts on the matter with great anticipation. (Strictly speaking, *Elbow Room* was not his next book after *Brainstorms. The Mind's I*, edited with Douglas Hofstadter (1981), came in between. It is mainly a collection of works by others, however, so we won't comment on it here, quirky and entertaining though it is.)

The year 1987 saw his second major collection of papers on content, *The Intentional Stance.* Dennett tells us that the first paper in the collection, "True Believers," replaces "Intentional Systems" as his flagship paper. The main difference between the two is that in determining what beliefs, desires, and so on, to attribute to a system, that is, what beliefs, desires, and so on, the system should have, evolutionary considerations now play a much bigger role than they did in the earlier paper (Ross 1). The concern to get clear about the "ontological status" of mental states that was finally put to rest in "Real Patterns" is very much in evidence in Dennett's reflections

in the early part of the collection. Two later papers in the collection are on evolutionary theory, including one of the most famous pieces he has ever written, "Evolution, Error, and Intentionality," a paper in which he explicitly and very firmly sets his face against all forms of the idea that mental content can ever be intrinsic to brain states. The volume contains the first full expression of the method for studying consciousness that was to become the centre of a large book on consciousness four years later, in a paper called (enigmatically) "Beyond Belief." There is nothing else in the collection on consciousness. This is firmly a volume about content.

At this point, Dennett left content behind for a while and turned to consciousness. His next book was a huge, sprawling work called *Consciousness Explained (CE)* (1991a). (For Dennett, modesty is a virtue to be kept for special occasions.) *CE* was aimed at a wide audience. For the first time (with the exception of *The Mind's I*), Dennett chose a trade publisher. It would not be the last.

In *CE*, Dennett has two main targets. One is the picture of conscious states that the tradition received from Descartes. This is the idea that there is something to a conscious state, some felt quality, that is unmistakably clear and clearly different from all other properties of mental states (Brook). The other is the picture of the conscious system that the tradition received from Descartes. This is the picture of the conscious system as a kind of screen on which conscious states play before a little homunculus sitting in the middle of the theatre (Dennett calls it a Cartesian theatre), where the conscious states themselves are conceived of as discrete, separately identifiable states, states with, for example, clear stop and start points. Dennett wants to replace the Cartesian picture with what he calls a Multiple Drafts Model (MDM) of consciousness. The MDM treats consciousness as a kind of mental content, almost a matter of programming, a highly controversial point of view (Churchland). The book concludes with a chapter pulling together a picture of the self and a final attempt to beat back two of the more esoteric attempts to make consciousness mysterious, one each by Thomas Nagel and John Searle.

Both Dennett's theory of content and his theory of consciousness require that the brain have certain capacities and structures. It must have the capacity to produce that incredible array of behaviour expressing mental content that we find in ourselves and others. And it must house an MD-type structure, which, Dennett suggests, probably consists in a Pandemonium architecture of some kind (Akins).

Having settled accounts with consciousness, Dennett next took up a task that many had been expecting him to turn to for a long time, evolutionary

theory. *Darwin's Dangerous Idea (DDI)* (1995) was also published as a trade book and enjoyed the same phenomenal success as *CE*. Two bestsellers on abstract issues in the philosophy of mind and philosophy of biology in a row is not bad!

In *DDI*, Dennett argues for two main claims:

1. Darwin's theory of evolution is a powerful 'universal acid' for dissolving all manner of intellectual 'skyhooks' and other pseudoscientific props that philosophers (and not just philosophers) have wheeled onto the stage to try to patch up hopeless theories; and yet,
2. Darwin's theory of evolution may deflate the pretensions of many accounts of morality but the ones it deflates are highly problematic in any case. Contrary to those who see Darwin as the destroyer of all morality, however, the theory of evolution leaves one perfectly satisfactory approach to morality and political philosophy untouched, namely, traditional Western liberalism.

In the course of developing his picture of evolution, Dennett tackles the central debates in contemporary evolutionary theory: adaptationism versus the idea that much of evolution has consisted in good tricks developed for one function being coopted for other functions; smooth evolution versus Gould's punctuated equilibrium; the role of genetic drift, climate changes, and other such accidental elements in evolutionary change; and so on. As readers of the *New York Review of Books* will know, some of Dennett's claims on these issues generated a firestorm of controversy. Although it's a little hard to understand why, Dennett found himself accused of being an 'ultra-Darwinian' (whatever that is) and Dawkins's lapdog. Needless to say, he responded with equal vigour and a hot argument ensued (Ross 2).

Among the most important claims introduced in *DDI* is a claim that it is language that makes it possible for us to have our kind of mind, a kind of mind that, by being able to cooperate with other minds and record the results of cooperation for others to build on, can figure out the physics of the universe, find cures for most serious diseases, build Hubble telescopes and Channel tunnels, create artificial speech–interpreters and problem solvers, and so on and so forth. By endowing us with language, evolution has utterly separated us from all cognitive systems that do not have language, where by "language" we mean something that is syntactically articulated and functions by building sentences that are structured compositions of lexical units. This became the basis of his next book, *Kinds of Minds* (1996), an attempt to

pull some of his most important ideas about minds in general and our kind of mind in particular together in one place and to say more than was said in *CE* (or anywhere else) on the evolution of our kind of intentionality. How far Dennett wants to go here, whether for example he wants to restrict even consciousness to animals with our kind of language, is not entirely clear (Clark).

Dennett's most recent work (as of the time of writing) is *Brainchildren* (1998), a collection that pulls together a remarkably diverse array of pieces written over the previous decade or so and appearing in various sometimes quite obscure places. The pieces range widely. They include rich philosophical essays such as "Real Patterns" (already mentioned); responses to criticism, especially of *CE*; a strikingly wide group of papers on artificial intelligence and artificial life; and some new papers on animal cognition and consciousness. The collection closes with two occasional pieces, a self-portrait and one of Dennett's few forays into morality, a paper on what the dangers of information technology are and are not. This collection is fun to read – much of it is written in an even more relaxed, polemical style than is usual with Dennett. One of the papers on animal cognition is on animal pain. For interpretationists about mental life such as Dennett, pain poses a special challenge because if anything in mental life just is, just hits us no matter how we interpret it, pain would seem to be it. Dennett does an impressive job of 'downsizing' the range of pains that animals could plausibly be said to feel.

In addition to the books we have discussed, Dennett has written hundreds of papers, critical studies, reviews, proposals, and so on and so forth. He has averaged about ten publications a year for over thirty years!

What does the future hold for Dennett? Well, he is not even at normal retirement age yet, so he is far from the end of his productive life. As we said, a new book on free will is being prepared even as we write. Beyond that, who knows? Dennett has written deep, groundbreaking books on all the topics that he set himself over thirty years ago and has lived in a sea of cut-and-thrust for his whole career as a result. (His public confrontations with John Searle and Jerry Fodor and especially Stephen Jay Gould are the stuff of legends.) But over and over he has shown that he still has new things to say about the topics that matter to him.

This ends the chronological summary of Dennett's corpus. We turn now to the interesting project of fitting the pieces together into a coherent whole. We will start with how the man fits into his time. It is not by accident that Dennett has been so influential. As well as being a brilliant polemicist, he started work at a pivotal point in the study of cognition, the beginning of

the cognitive revolution. His own work at the time was an important part of that pivot.

2. DENNETT'S PHILOSOPHY

There are a number of ways in which a scholar or scientist can be influential outside their home discipline. One is to have a single large idea. Another is to develop an important new technique. A third is to be so penetrating that just about anything one says on any topic is of interest. It might be supposed that Daniel Dennett's influence beyond philosophy has been of the third kind. His main analytic tool is probably the ingenious rhetorical question, the revealing intuition pump. Dennett's two best-known works outside of professional philosophy, *Consciousness Explained* (1991a) and *Darwin's Dangerous Idea* (1995), seem consistent with this diagnosis. Nonetheless, there is a distinctively Dennettian point of view and its influence has been more important than Dennett's work as a critical gadfly, brilliant and entertaining as the latter often is.

This point of view is clear already in *Content and Consciousness (C&C)* (1969), a blueprint for Dennett's entire corpus. One of the decisive moves in the cognitive revolution was the rejection of empiricism as it existed at mid-century. Dennett was one of those who charted a course beyond this form of empiricism and we will take this issue as our starting point.

2.1. Empiricism in Dennett's Time

As we said, Dennett was a student of the iconoclastic Oxford philosopher of mind, Gilbert Ryle, and, before that, of the Grand Admiral of logic and epistemology, W. V. O. Quine. Ryle wrote during the heyday of 'ordinary language philosophy,' the methodological movement inspired by J. L. Austin and given a tremendous boost by a certain (mis)reading of Wittgenstein's *Philosophical Investigations.* Ordinary language philosophers supposed, roughly, that some philosophical problems appear to be insoluble because philosophers misinterpret the language used to state them. Take a notion dear to Dennett's heart, belief. We speak of someone's *believing* that snow is white, and of someone's *believing* in the Loch Ness monster, and of their cat's *belief* that it's suppertime. It is easy to overinterpret these remarks as implying that 'belief' must denote a kind of state or object, as 'gas' does. We then go searching for properties that could gather all and only beliefs into a well-behaved set of such states or objects. Terrible difficulties

at once arise. No one (philosophers excepted) ever entertains to themselves the phrase 'snow is white, by gum,' yet most believe that proposition if they believe anything. It is hard to see, however, how a typical person could be said to believe in the reality of the Loch Ness monster unless they *had* sometime entertained the phrase 'Loch Ness monster.' As for the cat, its inability to entertain *any* phrases leaves its status as a believer-of-specific-contents puzzling: Can something have beliefs about suppertimes if, as is likely, it is incapable of believing that supper comes after lunch? Questions like these led ordinary language philosophers to think that something had gone deeply wrong. Typical English speakers seem to use words such as 'belief' without feeling compelled to suppose that cats secretly talk to themselves, or that tiny signs bearing the words 'snow is white' flash inside people's brains from time to time, or other absurdities that their words *seem* to imply.

This example of dissolving a philosophical puzzle is far too crude to represent the work of any actual ordinary language philosopher but it will suffice as an illustration of the method. The crucial move in it is to take the way people express themselves in language as *data*, that is, as a basis for philosophical reflection. However, data can be used in more than one way. In *C&C*, at least initially Dennett started from the same data as ordinary language philosophers but found that they led him to very different conclusions. Dennett never lost his respect for what he learned from his mentors about dissolving philosophical pseudo-puzzles. However, he found in ordinary uses of mentalistic language support, not for bland behaviourism, but for a kind of evolutionary cognitivism, as it is now called. (The kind of evolutionary cognitivism in question is the view that to understand behaviour, we must invoke a rich repertoire of mental states and processes, and that we can best identify the states and processes concerned by viewing the mental as having the shape it has due to natural selection.) Dennett's turn against his philosophical upbringing here was momentous; it marks one of the clearest turning points that we know of in the movement away from simple-minded behaviourism to the current (near)-consensus cognitivism.

Ryle's chief conclusion in *The Concept of Mind* is that philosophical hyper-literalness about the mentalistic idiom had, over several centuries, given rise to an absurd metaphysics of mind. Mentality is not a part of physical objects; physical objects do not hope for things or fear things or believe things. But minds are not parts of objects of any other kind, either. Mental objects no more hope or fear or believe than do physical ones. *People* hope or fear or believe – but, Ryle gives us reason to think, it is not clear that the word 'person' is being used to refer to an object here, certainly

not to an object by itself. Ryle demonstrates that ascription of mentalistic predicates to objects is a 'category mistake,' a logical error akin to ascribing a colour to an abstraction ("What colour is the square root of 3?") or contents to a vacuum ("It's chock-full of Nothingness"). Dennett, in *C&C*, summarizes the logical problem involved in applying mentalistic predicates to objects and processes by reference to the concept of *intentionality*. (The concept of intentionality was originally identified by medieval philosophers and reintroduced into philosophy by the nineteenth-century philosopher-psychologist Brentano.) The mental is *about* other states of affairs, whereas physical objects or states of physical objects are not. (Thus, even a good candidate case for mindless about-ness, such as "That tree's rings are about its age," is not one; it means, "That tree's rings *show* its age.") It is a category mistake to take anything having aboutness to be straightforwardly a property of an object of any kind.

It will be objected, "We *do* speak about animate bodies as engaging in activities that are about something; we speak of creatures' activities as directed to achieving goals and goals are about whatever they seek. For example, we say that she is walking *to get to* the store or he is dancing *because* it is fun. And so on. If Ryle is right, how is this possible?" It is possible because we see the motions involved in cases like these as *actions*, that is to say, as expressing reasons for doing things. Reasons for action are not causes, not causes as we usually think of them in the Humean tradition at any rate. Now, how does this distinction between reasons and (usual) causes account for our ascribing goals and other things having intentionality to bodies (i.e., actions) without introducing metaphysical extravagances (mentalistic 'ghosts in the machine,' to use Ryle's phrase)? Answer: Only if we can see how our reasons for doing the actions we do and the undeniably present causes of the bodily motions in which those actions consist relate to one another. In ordinary language, they peacefully coexist. We know that if we wish our arm to go up, then, under normal circumstances, it will go up, but we need not have any explanation of how either the reasons for this action or the causes of the related motion work.

This peaceful coexistence breaks down when we start doing scientific psychology. If all events must have physical causes, then, since beliefs, and so on, are not normal physical causes, the psychological theorist cannot invoke things like beliefs to explain behaviour. At this point, theorists are tempted to go in one of two directions. Either one may *identify* beliefs with states of the body (usually the brain), or one may deny that there are any such things as beliefs. The first path requires abandoning the everyday concept of belief and is otiose, since now the notion of belief is not doing any

explanatory work that notions of causes in brain circuitry can't do. Radical behaviourists who denied that minds exist spoke as if they favoured the second option, but they were just adding redundant metaphysical noises to their scientific talk, according to Ryle. That is, radical behaviourists seemed to be saying, "We're not content that it's methodologically unnecessary for us to talk about these things; we insist that they don't exist. We don't need this claim; we just . . . " – what? Hate it when people use alternative language about things like beliefs? It is not among a scientist's responsibilities to be a dogmatic metaphysician. (The *worst* move is to imagine that beliefs *cause* brain-states, for this multiplies entities gratuitously while *still* abusing the everyday concept.)

There is another alternative: logical behaviourism (the term usually applied to Ryle's position). On this view, talk about the causes of behaviour and talk about the reasons for behaviour can coexist because talk about reasons is a way of finding a particular *pattern in* behaviour, not a way of finding the *causes of* behaviour. On this view, scientific psychology and the mentalistic conceptual scheme are taken to have no direct connection to one another; it is only the philosopher's unwarranted assumption that they *must* serve the same functions that forces the postulation of 'the ghost in the machine': an object-like mind that both furnishes reasons for actions *and* is the cause of behaviour.

This happy ecumenicism ceases to be any comfort as soon as someone adds the opinion that science is in the business of describing the world *as it actually is*. In that case, what matters is not whether mentalistic talk is logically compatible with talk about causes of behaviour; all that matters is that science finds no place for such talk. Dennett's other teacher, Quine, explicitly said what Ryle only implied, namely, that mentalistic language describes no facts and is merely a 'dramatic idiom.' Quine marshalled an important point of logical analysis in support of this view. Scientific language, it is generally supposed, ascribes properties to objects (e.g., 'Spacetime is curved'). If so, one ought to be able to interchange coreferring terms within scientific sentences without changing their truth-values. Thus, if "The morning star orbits the sun" is true, then "The evening star orbits the sun" must also be true, since 'the morning star' and 'the evening star' refer to the same object. However, this logical property, called '*substitutability salva veritate*,' does not apply inside many sentences in mentalistic discourse. "Dweezil believes that the morning star orbits the sun" is perfectly consistent with the falsehood of "Dweezil believes that the evening star orbits the sun." In uttering such clauses, one is committed only to someone's intentional state being about something believed in, desired, hoped for, dreaded or whatnot,

not something that exists. By contrast, if one ascribes a property to chairs, one commits oneself to the existence of chairs.

So 'the unicorn in which John believes' refers (if it refers to anything) to what some philosophers call an 'intentional object,' not to an actual animal. Unless, it seems, brain-states can be directly taken to somehow be about intentional objects, sentences describing mental states won't have truth-values that derive in any straightforward way from those of nonmentalistic sentences. For Quine, there is a ready explanation for this semantic peculiarity: Sentences describing mental states do not literally *describe* anything. They are merely what Quine calls a dramatic idiom for something else. In both Quine's eyes and those of his friend Skinner, this analysis buttressed the scepticism of the radical behaviourist about the existence of the mental.

This was the high point of twentieth-century empiricism. Nothing exists, it was thought, except what we can sense using the sensitive surfaces of the body (eyes, ears, nose, taste, touch) and the correlations of these inputs with behavioural outputs. There may in addition be some apparatus connecting inputs to outputs or there may not but it does not matter – no such apparatus would be mental, as the mental has always been understood. Such empiricism pervaded the social sciences and some of the natural ones as well. Anthropologists 'discovered' that humans consist of arbitrary, infinitely variable belief-systems incommensurable with one another. For sociologists, the pressures of class membership caused people to invent value-systems that are merely rationalizations of power. Economists turned their venerable axioms of rationality into summaries of observed choice-behaviour lacking any deeper justification. Political scientists focussed on patterns in voting and other political behaviour. Many biologists emphasized the extent to which species are chains of descent shaped by geological and other environmental accidents. In the case of the biologists, this might have been more spin than substance but the social sciences were dominated by the idea that contingent environmental forces drive all behavioural processes. This was the environment into which Dennett came as a student.

The above description suggests a degree of doctrinal uniformity and simplicity that of course never really existed; but notice that now, thirty-five years later, the picture has no aptness at all. Articles and monographs now pour forth in which social scientists urge their colleagues to concentrate on the universal features of human behavioural patterns shaped by natural selection and fixed in the structures of cognition. Among biologists, the remaining defenders of pervasive contingency as the motor of evolution snarl defensively at 'ultra-Darwinian' adaptationists who depict

Mother Nature as an engineer carving a rationalizable trajectory through design space. Barkow, Cosmides, and Tooby (1992), in their manifesto for evolutionary psychology, aim explicitly at the old empiricism, which they call 'the standard social science model' (SSSM; the phrase is increasingly used as a rallying banner across the disciplines). The foundation of the empiricist SSSM was its concept of mind. In the cognitive revolution against it, Chomsky is the Copernicus. Jerry Fodor and Dan Dennett perhaps fight for the role of Galileo.

Unlike Chomsky and Fodor, Dennett did not overthrow the empiricist concept of mind wholesale. Ryle's analysis is not an inaccurate review of what people *do* in fact say about minds, and Quine's logical semantics remains the single most influential body of work in the field. Dennett showed that we can preserve Ryle's and some of Quine's leading insights while pushing our study of behaviour away from the periphery and deep into the organism.

2.2. The Foundations of Dennett's System

So how did Dennett respond to the radical empiricism of his time? To answer this question, we need to learn more about how he thinks about the mind and explanations of behaviour in terms of mental states. In many of his works, he starts by distinguishing three explanatory stances that one can take toward a complex organism or system. Consider a simple chess-playing computer (a favourite example of Dennett's).

One stance is to explain its current behaviour and predict its future behaviour by understanding how it is built and what the causal processes in it are like. This would give us an extraordinarily detailed, secure explanation of the system but at a severe price – extreme complexity. Not even the programmers who wrote the chess-playing programme could give this kind of explanation of the system. This stance Dennett calls *the physical stance*.

Another stance is to predict and explain the system's behaviour by understanding the design built into it, in this case the programme controlling its operations. Dennett calls this *the design stance*. This stance will produce an explanation much simpler than the first one – all we need to understand is how the system is designed to behave, not all the details of how and how well it implements this design – but the simplicity comes at a price. For our explanations to be any good, the system has to function as it was designed to function – and we have to assume that it is doing so or the design stance is useless to us. Because of this assumption built into them, explanations