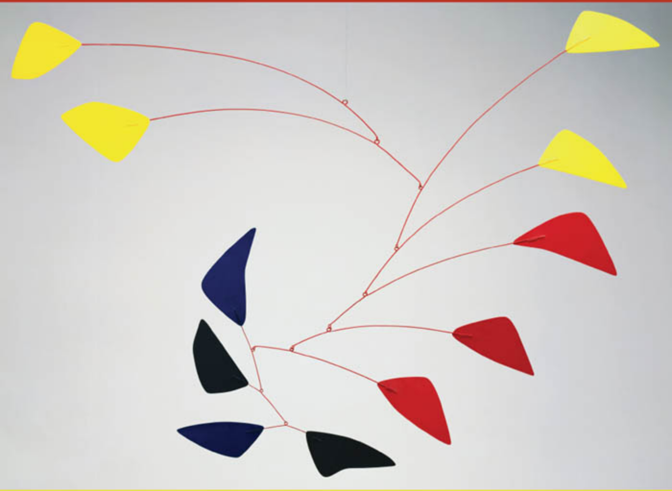
OXFORD

the Architecture of the Mind



PETER CARRUTHERS

The Architecture of the Mind

This book is a comprehensive development and defense of one of the guiding assumptions of evolutionary psychology: that the human mind is composed of a large number of semi-independent modules. *The Architecture of the Mind* has three main goals. One is to argue for massive mental modularity. Another is to answer a 'How possibly?' challenge to any such approach. The first part of the book lays out the positive case supporting massive modularity. It also outlines how the thesis should best be developed, and articulates the notion of 'module' that is in question. Then the second part of the book takes up the challenge of explaining how the sorts of flexibility and creativity that are distinctive of the human mind could possibly be grounded in the operations of a massive number of modules.

Peter Carruthers's third aim is to show how the various components of the mind are likely to be linked and interact with one another—indeed, this is crucial to demonstrating how the human mind, together with its familiar capacities, can be underpinned by a massively modular set of mechanisms. He outlines and defends the basic framework of a perception / belief / desire / planning / motor-control architecture, as well as detailing the likely components and their modes of connectivity. Many specific claims about the place within this architecture of natural language, of a mind-reading system, and others are explained and motivated. A number of novel proposals are made in the course of these discussions, one of which is that creative human thought depends upon a prior kind of creativity of action.

Written with unusual clarity and directness, and surveying an extensive range of research in cognitive science, this book will be essential reading for anyone with an interest in the nature and organization of the mind.

Peter Carruthers is Professor of Philosophy at the University of Maryland, College Park.

This page intentionally left blank

The Architecture of the Mind

Massive Modularity and the Flexibility of Thought

Peter Carruthers

CLARENDON PRESS · OXFORD

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 trademark 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

Carruthers, Peter, 1952–
The architecture of the mind : massive modularity and the flexibility of thought /
Peter Carruthers.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-19-920707-7 (alk. paper)
ISBN-10: 0-19-920707-0 (alk. paper)
ISBN-13: 978-0-19-920708-4 (alk. paper)
ISBN-10: 0-19-920708-9 (alk. paper)
1. Psychology—Philosophy. 2. Modularity (Psychology) 3. Brain. I. Title.
BF38.C375 2006

Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 0–19–920708–9 978–0–19–920708–4
ISBN 0–19–920707–0 (Pbk.) 978–0–19–920707–7 (Pbk.)

10 9 8 7 6 5 4 3 2 1

For Susan
the architecture of my life

This page intentionally left blank

Contents

List of Figures

Preface

This book has three main aims. One is to motivate and argue for a massively modular account of the architecture of the human mind. Another is to answer a 'How possibly?' challenge to any such approach. In the first part of the book (Chapters 1–3) the positive case for massive modularity is laid out. I also outline how the thesis of massive mental modularity should best be developed, and articulate the notion of 'module' that is appropriate to serve within such an account. And then in the second part of the book (Chapters 4–7) I take up the challenge of explaining how a massively modular mind could possibly display the sorts of flexibility and creativity that are distinctive of the human mind. Here the account that I provide finds a central place for representations of natural language sentences, among other things.

The third aim of this book is to give at least a sketch of the ways in which the various components of the mind are likely to be linked up to one another, and to interact with one another—indeed, this will be crucial to demonstrating how it is possible for the human mind, together with its familiar capacities, to be underpinned by a massively modular set of structures and components. Chapter 2 outlines and defends the basic framework of a perception / belief / desire / planning / motor-control architecture, as well as making proposals about many of the likely components and their modes of connectivity. And then in the chapters thereafter many specific claims about the place within this architecture of natural language, of a mind-reading system, and others are explained and motivated.

Although these three main strands in the book are mutually supporting, they are also to some degree independent of one another. Someone might find the arguments for massive modularity convincing, for example, while being unconvinced of my account of human flexibility, and while disagreeing with the overall architecture of components that I lay out. Or someone might think that the case for massive modularity is weak, while agreeing that my account of human flexibility and the component-architecture underpinning it are along the right lines—only requiring far fewer elements than a massive modularist would allege. And so on. But I believe that, taken together, the claims that I make under each of these three headings should add up to be (or rather multiply to be) an attractive overall package that is greater than the mere sum of its individual parts.

Our topic is massive modularity. But unfortunately there exists a wide variety of notions of modularity, put to work in a diverse range of literatures, extending from biology (Schlosser and Wagner, 2004), through computer science and artificial intelligence (Bryson, 2000, McDermott, 2001), to psychology (Karmiloff-Smith, 1992), to philosophy (Fodor, 1983, 2000; Samuels, 1998). Of these, Fodor's (1983) conception of a module has been especially influential, and many of the uses of the notion of modularity within cognitive science are to some degree variations upon it. In addition, a number of different researchers in cognitive science have argued for a form of *massive* mental modularity, and have done so in a variety of distinct ways (Tooby and Cosmides, 1992; Sperber, 1996; Pinker, 1997). But they, too, are by no means in complete agreement with one another about what modules, as such, *are*.

The way out of this morass is to line up the *arguments* for massive modularity with the *notion* of modularity that those arguments support. This is what I do in Chapter 1. I present and defend the cogency of three main arguments for massive modularity, while carefully teasing out the notion of 'module' that would be supported if those arguments are, indeed, cogent. The result is a notion of modularity that is some distance from Fodor's (in particular, modules needn't be informationally *encapsulated*). It is much closer to the use of the term 'module' in biology, and it is even closer to the notion used by researchers in artificial intelligence. On this account, a module is a functionally distinct processing system of the mind, whose operations are at least partly independent of those of others, and whose existence and properties are partly dissociable from the others. Moreover, modular systems must be *frugal* in their use of information and other cognitive resources, and they will have internal operations that are widely *inaccessible* to other systems. The thesis of massive mental modularity is then the claim that the mind is composed of *many* functionally isolable processing systems which possess such properties, and which have multiple input and output connections with others.

Chapter 2 then defends the major premise of one of the main arguments for massive modularity, claiming that the minds of non-human animals—from insects to chimpanzees—are massively modular in their organization. The chapter also locates those modules within a basic perception / belief / desire / practical reason / motor-control architecture, which will serve as the framework for the account of the structure of the human mind provided in later chapters. It also puts in place many specific ideas that will be needed in later chapters, including the claim that there is a limited capacity for mental rehearsal of action present in the minds of some of our primate cousins.

In Chapter 3 I discuss the modules that are likely to have been added, or enhanced, in the transition from the minds of the great-ape common-ancestors to our own. I defend the view that these are *multiple*, and argue at some length against the competing 'one major new adaptation' hypothesis. They include a mind-reading system, a natural language system, and systems for normative belief, reasoning, and motivation. In each case I discuss the probable internal organization of the module in question, and the ways in which it is likely to be embedded into the overall architecture of the mind.

Chapter 4 starts to take up the challenge of explaining the distinctive *flexibility* of the human mind in massively modular terms. It distinguishes a number of different kinds of flexibility, arguing that some are relatively easy to address while others are harder to explain. It then discusses how the language faculty may be responsible for flexibility of *content*, combining the outputs of other conceptual modules into a single representation that can then be mentally rehearsed, 'globally broadcast' (in the sense of Baars, 1988), and received as input by a whole suite of conceptual modules once again. Increasingly flexible *cycles* of modular processing thereby become possible, as do new kinds of reasoning.

Chapter 5 then tackles the problem of creativity. It advances the thesis that all forms of creative cognition reduce, ultimately, to creative *action*. In contrast with traditional views that see creative thought as prior to creative activity, I here argue the reverse. (Think of creative 'on-line' improvisation in jazz, to get a feel for the kind of thing that I have in mind.) The root of all creativity, I claim, lies in the creative activation and rehearsal of action schemata. The first manifestations of this ability are to be found in the problem-solving abilities of chimpanzees, and are then found—greatly enriched—in the pretend play of young children. Indeed, I claim that the *function* of childhood pretend play is to practice and further enhance that ability.

Chapter 6 turns to our capacity for science, and for abductive reasoning more generally (sometimes called 'inference to the best explanation'). Some people have claimed that our capacity for science is one of the remaining deep mysteries (comparable to the problem of consciousness, or the problem of the origin of the universe), and that it presents a formidable challenge for *any* cognitive theory to explain, let alone a massively modular one (Pinker, 1997; Fodor, 2000). Chapter 6 claims to solve this problem, in outline at least. Once again language and mental rehearsal play an important role in the account, as do principles employed in the interpretation of speech and the evaluation of testimony.

Chapter 7 discusses how the thesis of massive modularity can accommodate the distinctive features of human *practical* reasoning. This chapter is relatively brief, since most of the materials needed for the account have been put into place earlier in the book. What is new in the chapter is the suggestion that human practical reason can exploit the resources of our distinctively human *theoretical* reason. And I defend the belief / desire framework that I have adopted throughout against those who claim that, in the case of human beings, it is perceptions of *reasons*, rather than desires, that motivate our actions (Dancy, 1993, 2000; Scanlon, 1998). I also argue that conscious will is an illusion, developing one of the arguments of Wegner (2002). Chapter 8 then briefly summarizes the book's arguments and conclusions.

It hardly needs saying that this book is an ambitious one—indeed, it is almost absurdly so. For it takes as its goal nothing less than the elaboration and defense of a massively modular architecture for the human mind, to which many, many, different bodies of research are relevant. In consequence, there are numerous places where I have had to touch on topics on which I am by no means an expert. And there is, no doubt, a wealth of further evidence and theorizing out there in a variety of literatures that would be thoroughly germane to my project, if only I had the good fortune to know of it. Moreover, there are a wide range of kinds of expertise that are surely relevant to the evaluation of my various claims and proposals, some of which I simply don't possess. I can only console myself with the thought that *someone* has to step back from the details and paint the big picture, albeit taking big risks in doing so. And I hope that even if I have made many mistakes, and even if the architecture that I sketch in outline proves incorrect in many of its details, still what I have done might nevertheless be *roughly* along the right lines. At the very least, I hope that it will provide a useful foil and stimulus for the research of others.

It is worth remarking that many academic philosophers might fail to recognize what occurs within these pages as a form of philosophy. For the book contains very little that can be considered to be conceptual analysis, and most of the arguments are empirically grounded inferences to the best explanation, rather than deductive in form. If this is so, then so be it, and so much the worse for the philosophers in question. For by the same token much of the work of Aristotle, and of Hume, wouldn't count as philosophy, either. I believe that I have good role-models. And I believe that naturalistic philosophy, of the sort exemplified here, is the way (or at least, *a* way) that philosophy should be.

Many cognitive scientists, likewise, might fail to recognize what occurs within these pages as a form of science. For I report no new data or experiments. And although I do review a great deal of scientific data, I also make proposals and outline theories that go well beyond anything that the current evidence might strictly warrant, and many of the ideas that I defend are avowedly speculative. I can only plead that science always contains what might be called 'framework assumptions', as well as detailed theories closely grounded in the empirical data. And the examination and defense of those assumptions can be the work of naturalistically minded philosophers.

Again Hume (1739) provides us with a model. Although he, too, conducted no experiments, he saw himself as attempting to ground an empirical science of psychology, and the framework that he laid out has proven immensely influential amongst working psychologists ever since. (Even those of us who reject his associationism and empiricism can continue to find much that is of value in his work—see Fodor, 2003.) I should stress that the present book is a good deal less ambitious, of course. I am not attempting to *found* a massively modular framework for psychology, since much excellent work in that tradition already exists. Instead, I aim to lay out the best case for it, to defend it, and to show how it might be able to overcome its greatest difficulties. I thus see myself in the more modest role of 'under-laborer for science', championed by Locke (1690).

Acknowledgments

This book is the indirect product of twenty years of interdisciplinary theorizing and research, and the direct product of ten years of thinking about modularity issues. I have been heavily involved in interdisciplinary projects linking both philosophy and psychology since I began attending workshops and participating in reading groups while at the University of Essex through the latter half of the 1980s. (That was a time of transition for me. I was finishing up my work on Wittgenstein's *Tractatus*—which resulted in a pair of monographs at the close of the decade—while at the same time 'retraining' myself as a philosopher of psychology.) And in the year 1992 I was fortunate enough to secure funding from the Hang Seng Bank of Hong Kong to launch the Hang Seng Centre for Cognitive Studies at the University of Sheffield, where I then worked. In that same year I instituted a series of interdisciplinary colloquia, and began running (in collaboration with various others) a series of interdisciplinary workshops and for-publication conferences that lasted until the summer of 2004. The result has been a total of seven edited volumes of interdisciplinary essays (Carruthers and Smith, 1996; Carruthers and Boucher, 1998; Carruthers and Chamberlain, 2000; Carruthers, Stich, and Siegal, 2002; Carruthers, Laurence, and Stich, 2005, 2006, and one planned for 2007). It hardly needs saying that I have benefited immeasurably from conversations with the participants in these projects, from hearing their presentations, and from reading their work. They are, unfortunately, too numerous to list here. (I estimate that there are more than 250 of them.) But I should like to record my gratitude for all that I have learned from these colleagues over the years.

Some portions of the present book draw on previous publications of mine, to a greater or lesser extent. (I have been working out my ideas gradually as I have gone along, with frequent additions and changes of mind.) I am grateful to the editors and publishers of the pieces listed here for permission to reproduce material. And I am grateful, too, to all those who gave me advice and critical commentary on those items, at various stages of their production. (Again, they are too numerous to list; their names can be found in the acknowledgments sections of the original publications.) The publications are:

'The roots of scientific reasoning: infancy, modularity, and the art of tracking.' In P. Carruthers, S. Stich and M. Siegal (eds.), *The Cognitive Basis of Science*, Cambridge University Press, 2002, 73–95. 'Human creativity: its evolution, its cognitive basis, and its connections with childhood pretence.' *British Journal for the Philosophy of Science*, 53 (2002), 1–25; published by Oxford University Press. 'The cognitive functions of language.' *Behavioral and Brain Sciences*, 25 (2002), 657–719; published by Cambridge University Press. 'On Fodor's Problem.' *Mind and Language*, 18 (2003), 502–23; published by Blackwell Press. 'On being simple minded.' *American Philosophical Quarterly*, 41 (2004), 205–20; published by University of Illinois Press. 'Practical reasoning in a modular mind.' *Mind and Language*, 19 (2004), 259–78; published by Blackwell Press. 'Distinctively human thinking: modular precursors and components.' In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: structure and contents*, Oxford University Press, 2005, 69–88. 'Why the question of animal consciousness might not matter very much.' *Philosophical Psychology*, 18 (2005), 83–102; published by Routledge. 'The case for massively modular models of mind.' In R. Stainton (ed.), *Contemporary Debates in Cognitive Science*, Blackwell, 2006, 3–21. 'Simple heuristics meet massive modularity.' In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: culture and cognition*, Oxford University Press, 2006.

I should stress that the present book is far more than any sort of compilation of these previous publications, however. There is much in it that is new. And the last item listed, in particular, involved a very substantial change of mind concerning the nature of modularity. This required me to rethink much that I had previously written on the topic. I am especially grateful to Stephen Stich for forcing that rethinking on me over the course of a series of conversations. (Reading and commenting on drafts of Samuels, 2005, probably also had some significant effect in this regard.) And I am grateful to Randy Gallistel for a dinner-time tutorial that set me right on the architecture of practical reasoning in non-human animals, which again occasioned some important rethinking on my part.

Thanks to the following friends and colleagues for their comments on some or all of the first draft of this book: Greg Currie, Andrew Coward, Dustin Stokes, Jonathan Weisberg, and especially Mike Anderson, Keith Frankish, Paul Pietroski, and Richard Samuels. In addition, I am grateful for recent discussions with Isaac Carruthers, Erich Diese, Jerry Levinson, Louis Liebenberg, Joe Mikhael, Petter Sannum, and Mike Tetzlaff (the latter of whom also participated in writing one of the sections in Chapter 1, on languages of thought). Thanks also go to Shaun Nichols and Stephen Stich for permission to reproduce two figures from their 2003, printed here as Figures 3.1 and 5.1; to Daniel Felleman and David Van Essen for permission to reproduce their map of the visual areas in the macaque visual cortex from their 1991, printed here as Figure 1.3; and to Randy Gallistel for permission to reproduce his ant navigation figure from his 2000, printed here as Figure 2.6.

Some of the ideas in this book have been presented as talks at the following venues over the last half-dozen years. I am grateful to all those who participated in the discussions that followed these presentations:

Austin Texas (Human Behavior and Evolution Society), Georgetown (Philosophy), London (Royal Institute of Philosophy), Michigan (Ann Arbor, Cognitive Science), Paris (CNRS), Penn (Cognitive Science), Providence RI (American Society for Aesthetics, symposium on creativity), Rutgers (Cognitive Science), Sheffield (three conferences as part of a three-year project on the innate mind), Toronto (Cognitive Science), Tucson AZ (Towards a Science of Consciousness 2004, symposium), Washington DC (American Philosophical Association, Eastern Division symposium), Washington University in St Louis (Philosophy), and Western Ontario (Philosophy graduate conference).

In addition, an early version of some of the material in this book was presented to a graduate seminar at the University of Maryland in the Spring of 2003; and then a polished penultimate version of the book was presented to another group of students in the Fall of 2005. I am grateful to all the students involved for their discussion and critical comment, mentioning especially:

Mike Anderson, Bryan Baltzly, Matt Barker, Susan Bilek, Benedict Chan, Melissa Ebbers, Kent Erickson, Joel Gibson, William Kallfelz, Andrew Kania, Dimiter Kirilov, Ryan Millsap, Harjeet Parmar, Jeremy Pober, Chris Pyne, Natalia Romanova, Leyland Saunders, Elizabeth Schecter, Elizabeth Stoll, and Benedicte Veillet.

Finally, I am grateful to the General Research Board at the University of Maryland for an award that enabled me to take a year without teaching to complete the first full draft of this book, and to my Dean, Jim Harris, for his support.

1

The Case for Massively Modular Models of Mind

My goal in this chapter is to set out the positive case supporting massively modular models of the human mind. Unfortunately, there is no generally accepted understanding of what a massively modular model of the mind *is*. So at least some of our discussion will have to be terminological. I shall begin by laying out the range of things that can be meant by 'modularity'. I shall then adopt a pair of strategies. One will be to distinguish some things that 'modularity' definitely *can't* mean, if the thesis of massive modularity is to be even remotely plausible. The other will be to look at the main arguments that have been offered in support of massive modularity, discussing what notion of 'module' they might warrant. It will turn out that there is, indeed, a strong case in support of massively modular models of the mind on *one* reasonably natural understanding of 'module'. But what really matters in the end, of course, is the substantive question what sorts of structure are adequate to account for the organization and operations of the human mind, not whether or not the components appealed to in that account get described as 'modules'. So the more interesting question before us is what the arguments that have been offered in support of massive modularity can succeed in showing us about those structures, whatever they get called. This substantive issue will occupy the bulk of the chapter.

1 Introduction: On Modularity

We begin our discussion with a consideration of the range of things that can (and have) been meant by 'modularity'. I shall pay special attention to the work of Fodor (1983), which has been particularly influential in some areas of cognitive science.

1.1 A Spectrum of Options

In the weakest sense, a module can just be something like: a dissociable functional component. This is pretty much the everyday sense in which one can speak of buying a hi-fi system on a modular basis, for example. The hi-fi is modular if one can purchase the speakers independently of the tape-deck, say, or substitute one set of speakers for another for use with the same tape-deck. Moreover, it counts towards the modularity of the system if one doesn't have to buy a tape-deck at all—just purchasing a CD player along with the rest—or if the tape-deck can be broken while the remainder of the system continues to operate normally. It is important to stress, however, that independence amongst modules is by no means total. The different parts need to be connected up with one another in the right way, and coupled to a source of electrical power, in order for the whole hi-fi system to work; and the amplifier is an indispensable—but still distinct—component.

Understood in this weak way, the thesis of massive mental modularity would claim that the mind consists entirely of distinct components, each of which has some specific job to do in the functioning of the whole. It would predict that the properties of many of these components could vary independently of the properties of the others. (This would be consistent with the hypothesis of 'special intelligences'—see Gardner, 1983.) And it would predict that the components should be separately modifiable, being differentially affected by at least some other factors.[1] Moreover, the theory would predict that it is possible for some of these components to be damaged or absent altogether, while leaving the functioning of the remainder at least partially intact.

Would a thesis of *massive* mental modularity of this sort be either interesting or controversial? That would depend upon whether the thesis in question were just that the mind consists of some modular components, on the one hand; or whether it is that the mind consists of *a great many* modular components, on the other. Read in the first way, then nearly everyone is a massive modularist, given the weak sense of 'module' that is in play. For everyone will allow that the mind does consist of distinct components; and everyone will allow that at least some of these components can be damaged without destroying the functionality of the whole. The simple facts of cortical blindness and deafness are enough to establish these weak claims.

Read in the second way, however, the thesis of massive modularity would be by no means anodyne—although obviously it would admit of a range

[1] Sternberg (2001) develops this aspect of (weak) modularity into an elaborate and well worked-out methodology for the discovery of distinct modules in many different areas of cognitive science, relying on their separate modifiability by other factors.

of strengths, depending upon *how many* components the mind is thought to consist of. Certainly it isn't the case that everyone believes that the mind is composed of a great many distinct functional components. For example, those who (like Fodor, 1983) picture the mind as a big general-purpose computer with a limited number of distinct input and output links to the world (vision, audition, etc.) don't believe this, even though they may allow that the input systems themselves are composed of multiple parts.

In reply it might be said that almost everyone accepts that the mind contains lots and lots of *representations* (e.g. sentences in a 'language of thought'). And shouldn't each one of these count as a distinct component? If so, then the thesis of massive modularity (in the weak sense of 'module') will be accepted by almost all cognitive scientists—certainly by all who believe in local representations. The thesis of massive modularity is generally understood to apply only to *processing systems*, however, not to the representations that might be produced by such systems. And this is how I myself propose to understand it. So for these purposes, perceptual systems and inferential systems are candidate modules, but the individual percepts and beliefs produced by such systems are not.

It is clear, then, that a thesis of massive (in the sense of 'multiple') modularity is a controversial one, even when the term 'module' is taken in its weakest sense. So those evolutionary psychologists who have defended the claim that the mind consists of a great many different modular processing systems (Tooby and Cosmides, 1992; Sperber, 1996; Pinker, 1997) are defending a thesis of considerable interest, even if 'module' just *means* 'component'.

At the other end of the spectrum of notions of modularity, and in the strongest sense, a module would have all of the properties of what is sometimes called a 'Fodor-module' (Fodor, 1983). That is, it would be a domain-specific innately specified processing system, with its own proprietary transducers, and delivering 'shallow' (non-conceptual) outputs (e.g., in the case of the visual system, delivering a 2½-D sketch; Marr, 1983). In addition, a module in this sense would be mandatory in its operations, swift in its processing, isolated from and inaccessible to the rest of cognition, associated with particular neural structures, liable to specific and characteristic patterns of breakdown, and would develop according to a paced and distinctively arranged sequence of growth. I shall need to comment briefly on the various elements of this account.

1.2 On Fodor-Modularity

According to Fodor (1983) modules are domain-specific processing systems of the mind. Like most others who have written about modularity since, he understands this to mean that a module will be restricted in the kinds of content that it takes as input.[2] It is restricted to those contents that constitute its *domain*, indeed. So the visual system is restricted to visual inputs; the auditory system is restricted to auditory inputs; and so on. Furthermore, Fodor claims that each module should have its own transducers: the rods and cones of the retina for the visual system; the eardrum for the auditory system; and so forth.

According to Fodor (1983), moreover, the outputs of a module are *shallow* in the sense of being non-conceptual. So modules generate *information* of various sorts, but they don't issue in *thoughts* or *beliefs*. On the contrary, belief-fixation is argued by Fodor to be the very archetype of a *non*-modular (or holistic) process. Hence the visual module might deliver a representation of surfaces and edges in the perceived scene, say, but it wouldn't as such issue in *recognition* of the object as a chair, nor in the *belief* that a chair is present. This would require the cooperation of some other (non-modular) system or systems.

Fodor-modules are supposed to be innate, in some sense of that term,[3] and to be localized to specific structures in the brain (although these structures might not, themselves, be local ones, but could rather be distributed across a set of dispersed neural systems). Their growth and development would be under significant genetic control, therefore, and might be liable to distinctive patterns of breakdown, either genetic or developmental. And one would expect their growth to unfold according to a genetically guided developmental timetable, buffered against the vagaries of the environment and the individual's learning opportunities.

Fodor-modules are also supposed to be mandatory and swift in their processing. So their operations aren't under voluntary control (one can't turn them off), and they generate their outputs extremely quickly by comparison with other (non-modular) systems. When we have our eyes open we can't help but see what is in front of us. And nor can our better judgment (e.g. about the equal lengths of the two lines in a Müller-Lyer illusion) over-ride

[2] Many evolutionary psychologists understand domain-specificity somewhat differently. They tend to regard the domain of a module to be its *function*. The domain of a module is what it is *supposed to do*, on this account, rather than the class of contents that it can receive as input. I shall follow the more common *content* reading of 'domain' in the present chapter. But the two notions turn out to be intimately connected with one another when the notion of 'input' is elucidated properly, as we shall see shortly.

[3] Samuels (2002) argues convincingly that 'innate', in the context of cognitive science, should mean something like 'cognitively primitive'. Innate properties of the mind are ones that emerge in the course of development that is normal for that genotype, but that admit of no cognitivist explanation. (For example, they aren't explicable as resulting from some sort of *learning* process.) So they are cognitively basic—they can be appealed to in the explanation of other mental properties, but don't themselves admit of a cognitive explanation (as opposed, e.g., to a biological one).

the operations of the visual system. Moreover, compare the speed with which vision is processed with the (much slower) speed of conscious decision-making.

Finally, modules are supposed by Fodor to be both isolated from the remainder of cognition (i.e. encapsulated) and to have internal operations that are inaccessible elsewhere. These properties are often run together with each other (and also with domain specificity), but they are really quite distinct. To say that a processing system is *encapsulated* is to say that its internal operations can't draw on any information held outside of that system in addition to its input. (This isn't to say that the system can't access any stored information at all, of course, for it might have its own dedicated database that it consults during its operations.) In contrast, to say that a system is *inaccessible* is to say that other systems can have no access to its internal processing, but only to its outputs, or to the results of that processing.

Note that neither of these notions should be confused with that of *domain specificity*. The latter is about restrictions on the input to a system. To say that a system is domain specific is to say that it only receives inputs of a particular sort, concerning a certain kind of subject matter. Whereas to say that the processing of a system is encapsulated, on the one hand, or inaccessible, on the other, is to say something about the access-relations that obtain between the internal operations of that system and others. Hence one can easily envisage systems that might *lack* domain specificity, for example (being capable of receiving any sort of content as input), but whose internal operations are nevertheless encapsulated and inaccessible (Carruthers, 2002a; Sperber, 2002; and see the discussion of the supposed logic module in Section 2 below).

The explanations just given depend crucially on the notion of the *input* to a system, however. And this, too, needs some elucidation. One option would see the notion of *input* as contrasting with that of *database*, or stored information of any kind (as proposed by Carruthers, 2003). In which case any sort of activated information generated by and received from other processing systems would count as input, provided that the receiving system could do something with that information. But understanding the notion of input in this way would deliver a highly counter-intuitive account of the notion of *encapsulation*. For suppose that a processing system were so set up that it could query a wide range of other systems for information in the course of its normal operations; but suppose that the system in question couldn't, itself, access any *stored* information. Then if 'encapsulation' meant 'processing that can't draw on any information besides the input', and 'input' just meant 'information received from another system (rather than accessed from a database)', then the system envisaged would count as an entirely encapsulated one! For while it can draw on information from many other systems in the mind, all of this would be categorized as 'input' to the system, and hence wouldn't count against the system's encapsulation. This is a conclusion that surely needs to be avoided.

Another way to see the point is to notice that the distinction between the database that a system can search, on the one hand, and the information that is made available to that system by other processing systems, on the other, can't bear the weight required of it. Imagine two systems A and B. System A conducts searches across all stored beliefs in the course of its own operations, and hence isn't encapsulated by the above account. System B doesn't do this, but queries a wide range of other systems, which collectively search all stored beliefs on its behalf. It is absurd to say that while System A is unencapsulated, System B is an encapsulated one (on the grounds that all of the information made available to it counts as *input*, and hence doesn't count against its encapsulation). For neither is in any meaningful sense isolated from information held or generated elsewhere in the mind.

It is better to understand the input to a system to be the set of items of information that can *turn the system on*. For example, the face-recognition system is turned on by representations of eye-like and mouth-like shapes related to one another in such a way as to form a rough triangle. And the mind-reading system is turned on by representations of behavior. As we shall see in Section 2, the mind-reading system might need to send queries elsewhere in order to do its work, seeking information from the folk-physics system, perhaps. But the answers to those queries don't count as input to the mind-reading system, because the latter can't be *activated by* mere representations of physical movement.

This account gives us plausible readings of both 'domain specificity' and 'encapsulation'. The domain of the mind-reading system includes intentional behavior, but not mere physical movements like the sight of a ball rolling down a hill, because only the former will cue the system into action. But if the mind-reading system operates by querying a wide range of other systems for information (including the folk-physics system, say), then it won't count as an encapsulated one. The terminology that has been introduced here is summarized in Figure 1.1, for ease of reference later on.[4]

[4] Note that this account also ties the *content* reading of 'domain' more closely to the *function* reading favored by some evolutionary psychologists. For the representations of behavior that turn the mind-reading system on—its inputs, and hence its content-domain—are (for the most part) the ones that it was *designed* to process, and hence also constitute its functional domain. (The qualification 'for the most part' is needed because of the distinction that Sperber, 1996, draws between the *proper* domain of a module—which is the set of inputs that it was designed to process—and the *actual* domain, which could be much wider. Geometric shapes moving around on a computer screen in the right sort of contingent way will cue me into interpreting their movements in intentional terms; but the mind-reading system wasn't designed to process the movements of geometric shapes.)

Input: The input to a system to be the set of items of information that can *turn the system on*. (The notion of *input* contrasts with that of *information accessed in the course of processing*, whether activated or stored.)

Domain specific: To say that a system is domain specific is to say that it only receives inputs of a particular sort, concerning a certain kind of subject matter.

Encapsulation: To say that a processing system is encapsulated is to say that its internal operations can't draw on any information held outside that system in addition to its input.

Inaccessible: To say that a system is inaccessible is to say that other systems can have no access to its internal processing, but only to its outputs, or to the results of that processing.

Figure 1.1. Some key terminology

2 What Massive Modularity could not be

There is nothing incoherent in the idea that the mind might consist of a great many Fodor-modules (or at least in systems that closely resemble Fodor-modules). Indeed, such an idea is consistent with Brooks's (1986) *subsumption architecture* for the mind. On such an account the mind consists of a whole suite of input-to-output modular processing systems, with the overall behavior of the mind, and of the organism that it governs, being determined by competition amongst such modules. Each module will receive its input from a set of sensory transducers (whether shared or proprietary), and will serve to control some specific type of behavior of the organism. The operations of each module will be mandatory, encapsulated, and inaccessible. And modules might be innate, each with its own specific neural realization, distinctive developmental trajectory, and characteristic patterns of breakdown.

I shall argue in Chapter 2, however, that a subsumption architecture isn't even plausible as an account of the minds of insects and arachnids, let alone of the human mind. Indeed, I shall argue that perception / belief / desire / planning architectures are well nigh ubiquitous in the animal kingdom. On such an account, perception gives rise to beliefs; a combination of perception and the organism's bodily states gives rise to desires; and then beliefs and desires are combined with one another within some sort of practical-reasoning system to select an appropriate behavior. I shall assume the correctness of this kind of view in the present chapter, returning to defend it in the chapter that follows.

If belief / desire architectures are presupposed, then it is obvious that by 'module' we can't possibly mean 'Fodor-module', if a thesis of massive mental modularity is to be even remotely plausible. In particular, some of the items in Fodor's list will need to get struck out as soon as we move to endorse any sort of central-systems modularity, let alone entertain the idea of *massive* modularity. (This is no accident, since Fodor's analysis was explicitly designed to apply to modular input and output systems like color perception or face recognition. Fodor has consistently maintained that there is nothing modular about central cognitive processes of believing and reasoning. See Fodor, 1983, 2000.) If there are to be conceptual modules—modules dealing with common-sense physics, say, or common-sense biology, or with cheater-detection, to name but a few examples that have been proposed by cognitive scientists in recent decades—then it is obvious that modules can't have their own proprietary transducers. Nor can they have shallow outputs. On the contrary, their outputs will be fully conceptual thoughts or beliefs.

Is this way of proceeding question-begging? Can one insist, on the contrary, that since modules *are* systems with shallow outputs we can see at a glance that the mind can't be massively modular in its organization? This would be fine if there were already a pre-existing agreed understanding of what modules are supposed to be. But there isn't. As stressed in Section 1, there are a *range* of meanings of 'module' available. And we surely shouldn't allow ourselves to become fixated on Fodor-modularity (as seems to have happened to most philosophers who write on these topics). On the contrary, principles of charity of interpretation dictate that we should select the meaning that makes the best sense of the claims of massive modularists. That is what I aim to do in this chapter.

What of domain specificity, in the context of a thesis of massive modularity? I once argued that this also needs to be dropped (or to be re-conceptualized in terms of functional rather than content domains), on the grounds that the practical-reasoning system should be considered as a distinct module, but one that would be capable of receiving any belief and any desire as input (Carruthers, 2004a). I now think, however, that practical reasoning is underpinned by a whole host of different systems, each of which is turned on by a specific sort of motivation, and each of which then searches for information in the service of that motivation within specific locations in the mind. (See Chapter 2.7 and 2.8.) So each such system can probably count as domain specific in character.

Although it may well be the case that *most* modules are domain specific, we could surely accept that some aren't, without thereby compromising the thesis of massive mental modularity. Sperber (1996), for example, hypothesizes the existence of a formal logic module, whose task is to deduce some of the simpler logical consequences of any set of beliefs that it receives as input. For example, when it receives as input any pair of propositions of the form, 'P' and 'P ⊃ Q', it immediately deduces the conclusion 'Q'. The operations of such a module might be entirely encapsulated (as well as sharing many other elements of Fodor-modularity). But it plainly couldn't be domain specific, since in order to do its job it would have to be capable of getting turned on by *any* set of beliefs of the right form.

While we should accept that most conceptual modules are likely to be domain specific, then, we shouldn't absolutely require it. Swiftness of processing, in contrast, surely needs to go, in the context of massive modularity (except perhaps in comparison with the speed of *conscious* thought processes, if the latter are realized in cycles of modular activity, as I shall argue in Chapter 4 that they are). For if the mind is *massively* modular, then we will lack any significant comparison-class. Fodor-modules were characterized as swift in relation to *central* processes; but a massive modularist will maintain that the latter are modular too.

It looks plausible that the claim of mandatory operation should be retained, however. Each component system of the mind can be such that it automatically processes any input that it receives. (Indeed, such a claim is almost analytic, if the input to a system is just the information that is capable of turning on its operations, as we suggested above.) And certainly it seems that *some* of the alleged central modules, at least, have such a property. As Segal (1998) points out, we can't help but see the actions of an actor on the stage as displaying anger, or jealousy, or whatever; despite our knowledge that he is thinking and feeling none of the things that he appears to be. So the operations of our mind-reading faculty seem to be mandatory. Another way to put the point is that the operations of a system are mandatory if they can't be *turned off at will*. And it seems very likely that most (if not all) of the component systems that make up the human mind are mandatory in this sense, and that conscious decisions shouldn't be capable of determining whether or not a given system continues operating. This is what explains Segal's point: just by reminding ourselves that this is only an actor, we can't stop the mind-reading system from processing the behavioral input that it receives.

Now what of claims of innateness, and of neural specificity? Certainly one *could* maintain that the mind consists almost exclusively of innately channeled processing systems, realized in specific neural structures. This would be a highly controversial claim, but it wouldn't be immediately absurd. Whether this is the *best* way to develop and defend a thesis of massive modularity is moot. Certainly, innateness has been emphasized by evolutionary psychologists, who have argued that natural selection has led to the development of multiple innately channeled cognitive systems (Tooby and Cosmides, 1992). But others have argued that modularity is the product of learning and development (Karmiloff-Smith, 1992; Paterson et al., 1999). Both sides in this debate agree, however, that modules will be realized in specific neural structures (not necessarily the same from individual to individual). And both sides are agreed, at least, that development begins with a set of innate attention biases and a variety of innately structured learning mechanisms.

My own sympathies in this debate are towards the nativist end of the spectrum. I suspect that much of the structure, and many of the contents, of the human mind are innate or innately channeled.[5] But in the context of developing a thesis of *massive* modularity, it seems wisest to drop the innateness-constraint from our definition of what modules are. For one might want to allow that some aspects of the mature language faculty are modular, for example, even though it is saturated with acquired information about the lexicon of a specific natural language like English. And one might want to allow that modules concerned with particular behavioral skills can be constructed by over-learning, say, in such a way that it might be appropriate to describe someone's reading competence as modular.[6]

Finally, we come to the properties of encapsulated and inaccessible processing. These are thought by many (including Fodor, 2000) to be the core properties of modular systems. And there seems to be no a priori reason why the mind shouldn't be composed exclusively out of such systems, and cycles of operation of such systems. At any rate, such claims have been defended by a number of those who describe themselves as massive modularists (Sperber, 1996, 2002, 2005; Carruthers, 2002a, 2003, 2004a). I shall leave the claim of inaccessibility untouched for the moment, pending closer examination of the arguments in support of massive modularity. But I do want briefly to argue that massive modularists shouldn't claim that the mind must consist exclusively of systems that are encapsulated. (I shall then return to this point at greater length in Section 6, in the context of an examination of Fodor's arguments.)

As we noted above, even where a system has been designed to focus on and process a particular domain of inputs, one might expect that in the course of its normal processing it might need to query a range of other systems for information of other sorts. Consider the mind-reading system, for example, which virtually every massive modularist would consider to be realized in

[5] What does it *mean* to say that some property of the mind is innate? Certainly not that it is built from a fully-specified genetic blueprint. For that isn't the way in which the development of any aspect of an organism occurs: it is always an interaction of genes and gene-environments (Carroll, 2005). Nor does it mean that the property is present at birth, nor universal in all members of the species. Rather, I endorse the analysis given by Samuels (2002). For the purposes of cognitive science, a trait is innate if (a) it emerges during the course of development that is normal for the genotype, and (b) it is cognitively *basic*, not admitting of a cognitive (e.g. learning-based) explanation.

[6] Indeed, recent theories suggest that there are *many* fine-grained motor-control modules constructed via learning, with different modules being constructed each time we acquire a new skill, or learn to manipulate a new tool (Ghahramani and Wolpert, 1997; Haruno et al., 2001).

a module (or collection of modules). This is designed to focus on behavior together with attributions of mental states, and to generate predictions of further behavior and/or attributions of yet other mental states. Yet in the course of its normal operations it may need to query a whole range of other systems for information relevant to solving the task in hand. In which case the system isn't an encapsulated one.

Consider an example used by Currie and Sterelny (2000) in their criticism of the view that the mind-reading faculty might be modular (which they take to require both domain specificity and encapsulation). They write:

> If Max's confederate says he drew money from their London bank today there are all sorts of reasons Max might suspect him: because it is a public holiday there; because there was a total blackout of the city; because the confederate was spotted in New York at lunch time. Just where are the bits of information to which we are systematically blind in making our social judgments? The whole genre of the detective story depends on our interest and skill in bringing improbable bits of far-away information to bear on the question of someone's credibility. To suggest that we don't have that skill defies belief.

While the example perhaps shows that mind-reading isn't encapsulated, it shows nothing about lack of domain specificity; nor does it show that mind-reading isn't underpinned by a specialized function-specific processing system. (That is, it does nothing to show that mind-reading is just an aspect of some sort of holistic general-purpose cognition, as Currie and Sterelny believe.) This is because the skill in question arguably isn't (or isn't largely) a mind-reading one. Let me elaborate.

Roughly speaking, to lie is to assert that P while believing that not-P. So evidence of lying is evidence that the person is speaking contrary to their belief. In the case of Max's confederate it is evidence that, although he *says* that he drew money from the London account today, he actually believes that he didn't. Now the folk-psychological principle, 'If someone didn't do something, then they believe that they didn't do that thing', is surely pretty robust, at least for actions that are salient and recent (like traveling to, and drawing money from, a bank on the day in question). So almost all of the onus in demonstrating that the confederate is lying will fall onto showing that he didn't in fact do what he said he did. And this isn't anything to do with mind-reading per se. Evidence that he was somewhere else at the time, or evidence that physical constraints of one sort or another would have prevented the action (e.g. the bank was closed), will (in the circumstances) provide sufficient evidence of duplicity. Granted, many different kinds of information can be relevant to the question what actually happened, and what the confederate actually did or didn't do. But this doesn't in itself count against the domain-specificity and function-specificity of a separately effectible mind-reading system (although it *does* count decisively against its encapsulation).

All that the example really shows is that the mind-reading faculty may need to work in conjunction with other elements of cognition in providing us with a solution to a problem, querying other systems for information.[7] In fact most of the burden in detective-work is placed on physical enquiries of one sort or another—investigating foot-prints, finger-prints, closed banks, whether the suspect was somewhere else at the time, and so forth—rather than on mind-reading. The contribution of the latter to the example in question is limited to (a) assisting in the interpretation of the original utterance (Does the confederate mean what he says? Is he joking or teasing?); (b) providing us with the concept of a lie, and perhaps a disposition to be on the lookout for lies; and (c) providing us with the principle that people generally know whether they have or haven't performed a salient action in the recent past.

What we have so far, then, is that if a thesis of massive mental modularity is to be remotely plausible, then by 'module' we cannot mean 'Fodor-module'. In particular, the properties of having proprietary transducers, shallow outputs, fast processing, significant innateness or innate channeling, and encapsulation will very likely have to be struck out. That leaves us with the idea that modules might be isolable function-specific processing systems, all or almost all of which are domain specific (in the content sense), whose operations aren't subject to the will, which are associated with specific neural structures (albeit sometimes spatially dispersed ones), and whose internal operations may be inaccessible to the remainder of cognition. Whether all of these properties should be retained in the most defensible version of a thesis of massive mental modularity will be the subject of the remainder of this chapter.

In the sections that follow I shall be considering the main arguments that can be offered in support of a thesis of massively modular mental organization. I shall be simultaneously examining not only the strength of those arguments, but also the notion of 'module' that they might warrant.

3 The Argument from Design

The first argument derives ultimately from Simon (1962), and concerns the design of complex functional systems quite generally, and in biology in

[7] See Nichols and Stich (2003) who develop an admirably detailed account of our mind-reading capacities, which involves a complex array of both domain-specific and domain-general mechanisms and processes, including the operations of a domain-general planning system and a domain-general suppositional system, or 'possible worlds box'. I shall discuss their model in greater detail in Chapter 3.

particular. According to this line of thought, we should expect such systems to be constructed hierarchically out of dissociable sub-systems (each of which is made up of yet further sub-systems), in such a way that the whole assembly can be built up gradually, adding sub-system to sub-system; where the properties of sub-systems can be varied independently of one another; and in such a way that the functionality of the whole is buffered, to some extent, from changes or damage occurring to the parts.

Simon (1962) uses the famous analogy of the two watchmakers to illustrate the point. One watchmaker assembles one watch at a time, attempting to construct the whole finished product at once from a given set of micro-components. This makes it easy for him to forget the proper ordering of parts, and if he is interrupted he may have to start again from the beginning. The second watchmaker first builds sets of sub-components out of the given micro-component parts, and then combines those into larger sub-component assemblies, until eventually the watches are complete. This helps organize and sequence the whole process, and makes it much less vulnerable to interruption.

3.1 Modules in Biology

Consistent with such an account, there is a very great deal of evidence from across many levels in biology to the effect that complex functional systems are built up out of assemblies of sub-components (West-Eberhard, 2003; Schlosser and Wagner, 2004; Callebaut and Rasskin-Gutman, 2005). Each of these components is constructed out of further sub-components and has a distinctive role to play in the functioning of the whole, and many of them can be damaged or lost while leaving the functionality of the remainder at least partially intact. This is true for the operations of genes, of cells, of cellular assemblies, of whole organs, of whole organisms, and of multi-organism units like a bee colony (Seeley, 1995). And by extension, we should expect it to be true of cognition also, provided that it is appropriate to think of cognitive systems as biological ones, which have been subject to natural selection. (I shall return to examine this question in Section 3.2.)

West-Eberhard (2003) argues that a belief in massive biological modularity—in the sense of discreteness and dissociability amongst parts combined with integration within parts—is well nigh ubiquitous across the biological sciences. But not everyone uses the term 'module' to designate this same concept, however. Many other words are used to describe the same thing, including 'atomization' (Wagner, 1996), 'autonomy' (Nijhout, 1991), 'compartmentalization' (Maynard-Smith and Szathmáry, 1995; Kirschner and Gerhart, 1998), 'individualization' (Larson and Losos, 1996), and 'sub-unit organization' (West-Eberhard, 1996). But the phenomenon in question—and the belief in its omnipresence—is the same in each case.

West-Eberhard (2003) herself argues that in the context of evolutionary developmental biology, the most fruitful way of individuating modules is in terms of the developmental / genetic switches that lead to their development. For each such switch leads to a distinct or partly distinct compartment in the individual's phenotype, in which a distinctive set of genes is expressed, and which can hence become a target of natural selection. And developmental determination by switches must occur prior to the resulting modular system becoming an adaptation, as well as prior to any further shaping of the functional integration of the module through the evolutionary process.

The important point for our purposes, however, is that modular organization is a prerequisite of—or is at least an *extremely* common solution to—evolvability (Wagner and Altenberg, 1996; Raff and Raff, 2000; Wimsatt and Schank, 2004). Since the properties of modules are to some significant degree independent of one another, both they and the developmental pathways that lead to them can have distinctive effects on the overall fitness of the organism. But by the same token, since modules are separately modifiable, natural selection can act on one without having to make alterations in all (which would have potentially disastrous effects). So evolution can tinker with the separate components of the overall organism, at many levels of organization, responding to particular evolutionary pressures by factoring the overall fitness of the organism into the distinctive fitness-effects of the component modules. Since only a modular organization can enable this to happen, the question for us is whether or not it is appropriate to think of the mind as a *biological* system, subject to the same evolvability requirements as any other such system. We will turn to that question shortly.

Before we do so, I want to stress that biological modularity is always a matter of degree (Rasskin-Gutman, 2005). Hence the notion doesn't just apply to so-called 'mosaic' traits like eye-color that can vary quite independently of all others, as Woodward and Cowie (2004) allege. Biological systems like hearts and lungs are closely interconnected with many others, of course—each is tightly tied into the bilateral organization of the body, and presupposes the existence of the other, for example. Nevertheless, each follows a developmental trajectory that is significantly independent of the other; events like cancer can affect the one without affecting the other; and there can be genetic variations in each one that don't lead to alterations in the other. There is therefore an important sense in which hearts and lungs can be regarded as distinct bodily modules.

What we should expect, then, is that cognitive systems can be more or less deeply embedded in the developmental / genetic hierarchy, and more or less closely dependent upon other such systems. Some might be more like lungs—homologous across many species, and crucial to the functioning of the whole. (The cognitive modules that process basic spatial properties and relationships might be a good example, here.) And some might be more like eye-color—varying across individual members of the species, and comparatively peripheral in function. (Some genetic variations in personality type might be an example of this.) But all should be to some important degree discrete and dissociable, while displaying significant internal integration.

3.2 Did the Mind Evolve?

What sorts of properties of organisms are apt to have fitness-effects? These are many and various, ranging from gross anatomical features such as size, shape, and color of fur or skin, through the detailed functional organization of specific physical systems such as the eye or the liver, to behavioral tendencies such as the disposition that cuckoo chicks have to push other baby birds out of the nest. And for anyone who is neither an epiphenomenalist nor an eliminativist about the mind, it is manifest that the human mind is amongst the properties of the human organism that has fitness-effects. For it will be by virtue of the mind that almost all fitness-enhancing behaviors—such as running from a predator, taking resources from a competitor, or wooing a mate—are caused.

On any broadly realist construal of the mind and its states, then, the mind is at least a prime *candidate* to have been shaped by natural selection. How could such a possibility fail to have been realized? How could the mind be a major cause of fitness-enhancing behaviors without being a product of natural selection? One alternative would be a truly radical empiricist one. It might be said that not only most of the contents of the mind, but also its structure and organization, are acquired from the environment. Perhaps the only direct product of natural selection is some sort of extremely powerful learning algorithm, which could operate almost equally well in a wide range of environments, both actual and non-actual. The fitness-enhancing properties that we observe in adult minds, then, aren't (except very indirectly) a product of natural selection, but are rather a result of learning from the environment within which fitness-enhancing behaviors will need to be manifested.

Such a proposal is an obvious non-starter, however. It is one thing to claim that all the *contents* of the mind are acquired from the environment using general learning principles, as empiricists have traditionally claimed. (This is implausible enough by itself; see Section 4 below, briefly, and Chapter 2, at length.) And it is quite another thing to claim that the structure and organization of the mind is similarly learned. How could the differences between, and characteristic causal roles of, beliefs, desires, emotions, and intentions be learned from experience?[8] For there is nothing corresponding to them in the world from which they could be learned; and in any case, any process of learning must surely presuppose that a basic mental architecture is already in place. Moreover, how could the differences between personal (or 'episodic') memory, factual (or 'semantic') memory, and short-term (or 'working') memory be acquired from the environment? The idea seems barely coherent. And indeed, no empiricist has ever been foolish enough to suggest such things.

We have no other option, then, but to see the structure and organization of the mind as a product of the human genotype, in exactly the same sense as, and to the same extent that, the structure and organization of the human body is a product of our genotype. But someone could still try to maintain that the mind isn't the result of any process of natural selection. Rather, it might be said, the structure of the mind might be the product of a single macro-mutation, which became general in the population through sheer chance, and which has remained thereafter through mere inertia. Or it might be the case that the organization in question was arrived at through random genetic drift—that is to say, a random walk through a whole series of minor genetic mutations, each of which just happened to become general in the population, and the sequence of which just happened to produce the structure of our minds as its end-point.

These possibilities are so immensely unlikely that they can effectively be dismissed out of hand. Evolution by natural selection remains the only explanation of organized functional complexity that we have (Dawkins, 1986). Any complex phenotypic structure, such as the human eye or the human mind, will require the cooperation of many thousands of genes to build it. And the possibility that all of these thousands of tiny genetic mutations might have occurred all at once by chance, or might have become established in sequence (again by chance), is unlikely in the extreme. The odds in favor of either thing happening are vanishingly small. (Throwing a '6' with a fair die many thousands of times in a row would be much more likely.) We can be confident that each of the required small changes, initially occurring through chance mutation, conferred at least some minor fitness-benefit on its possessor, sufficient to stabilize it in the population, and thus providing a platform on which the next small change could occur.

The strength of this argument, in respect of any given biological system, is directly proportional to the degree of its organized functional complexity—the more complex the organization of the system, the more implausible it is that

[8] Note that we aren't asking how one could learn from experience *of* beliefs, desires, and the other mental states. Rather, we are asking how the differences between these states themselves could be learned. The point concerns our acquisition of the mind itself, not the acquisition of a *theory* of mind.

it might have arisen by chance macro-mutation or random genetic walk. Now, even from the perspective of common-sense psychology the mind is an immensely complex system, which seems to be organized in ways that are largely adaptive. (As evidence of the latter point, witness the success of our species as a whole, which has burgeoned in numbers and spread across the whole planet in the course of a mere 100,000 years or so.) And the more we learn about the mind from a scientific perspective, the more it seems that it is even more complex than we might initially have been inclined to think. Systems such as vision, for example—that are treated as 'simples' from the perspective of common-sense psychology—turn out to have a hugely complex internal structure.

Before leaving this topic I should stress that it remains possible that some properties of the mind might be 'spandrels' (in the sense of Gould and Lewontin, 1979). From the claim that the mind as a whole is an adaptation resulting from natural selection, it of course doesn't follow that every property of the mind is an adaptation likewise. For some might be by-products of those that *are* adaptations. And I should also stress that when they happen against the right background, small changes can sometimes have large adaptive effects without any need for a history of selection. Thus consider the hypothesis put forward by Hauser et al. (2002), concerning the evolution of the language faculty. They suggest that many of the systems that enable language in humans are shared with other animal species, such as the capacity to carve a speech stream into phonemes, and the capacity for vocal imitation (Hauser, 1996). Against a sufficiently rich background, it might then have needed but a small and relatively simple change—perhaps to enable a particular sort of recursion in the generation of mental representations—to make fully human language possible. And this change itself might either have resulted from a single random mutation, or be a spandrel of some other selected-for change. None of this is ruled out by the claim that the mind as a whole has been shaped by natural selection.

3.3 How many Modules?

The prediction of this line of reasoning, then, is that cognition will be structured out of systems that are to some significant degree dissociable, and each of which has a distinctive function, or set of functions, to perform.[9] This gives us a notion of a cognitive 'module' that is pretty close to the everyday

[9] We should expect many cognitive systems to have a *set* of functions, rather than a unique function, since multi-functionality is rife in the biological world. Once a component has been selected, it can be co-opted, and partly maintained and shaped, in the service of other tasks. By the same token, we should expect many sub-modules to be *shared* amongst more than one superordinate system. I return to this point in Section 3.4.

sense in which one can talk about a hi-fi system as 'modular' provided that the tape-deck can be purchased, can function, and can vary its properties independently of the CD player, and so forth. Roughly, a module is just a dissociable *component*.

Consistent with the above prediction, there is now a great deal of evidence of a neuro-psychological sort that something like massive modularity (in the everyday sense of 'module') is indeed true of the human mind. People can have their language system damaged while leaving much of the remainder of cognition intact (aphasia); people can lack the ability to reason about mental states while still being capable of much else (autism); people can lose their capacity to recognize just human faces; someone can lose the capacity to reason about cheating in a social exchange while retaining otherwise parallel capacities to reason about risks and dangers; someone can lose the capacity to name living things while retaining the capacity to name non-living things, or vice versa; someone can lose the capacity to name fruits and vegetables while retaining their ability to name animals; and so on and so forth (Sachs, 1985; Shallice, 1988; Hart and Gordon, 1992; Sacchett and Humphreys, 1992; Baron-Cohen, 1995; Farah et al., 1996; Tager-Flusberg, 1999; Stone et al., 2002; Varley, 2002).[10]

But just *how many* components does this argument suggest that the mind consists of, however? Simon's (1962) argument makes the case for hierarchical organization, but Samuels (2006) claims that the argument fails to establish modularity of mind in any interesting sense. At the top of the hierarchy will be the target system in question (a cell, a bodily organ, the human mind). And at the base will be the smallest micro-components of the system, bottoming out (in the case of the mind) in the detailed neural processes that realize cognitive ones. But it might seem that it is left entirely open how high or how low the pyramid is (i.e. how many 'levels' the hierarchy consists of), and whether the 'pyramid' has concave or convex edges. If the pyramid is quite low with concave sides, then the mind might decompose at the first level of analysis into just a few constituents such as *perception, belief, desire*, and *the will*, much as traditional 'faculty psychologies' have always assumed; and these might then

[10] In fact very few of these disorders are 'clean', with just the target capacity damaged and all else left intact. In most cases where one capacity is damaged, others will be damaged also. Where the damage results from an acquired brain-injury, this is hardly very surprising. For few such injuries are likely to affect just a single brain system. But even where the damage is genetic, we should not be surprised. For as Marcus (2004) points out, a very high proportion of the genes involved in building any one bodily system will also be involved in building others; so the vast majority of genetically based disorders should be expected to have broad effects. In addition, where modules share parts, damage to one of those parts will have an impact on the functioning of more than one superordinate system. And such sharing of parts is likely to be rife in cognitive systems, just as it is in biological ones.

get implemented quite rapidly in neural processes. In contrast, only if the pyramid is high with convex sides should we expect the mind to decompose into *many* components, each of which in turn consists of many components, and so on. (See Figure 1.2.)

There is more mileage to be derived from Simon's argument yet, however. For the range and complexity of the functions that the overall system needs to execute will surely give us a direct measure of the height of the pyramid and manner in which it will slope. (The greater the complexity, the greater the number of sub-systems into which the system will decompose at each level of organization, and the greater the number of levels.) This is because the hierarchical organization is there in the first place to ensure evolvability and robustness of function. Evolution needs to be able to tinker with one function in response to selection pressures without necessarily impacting any of the others.[11]

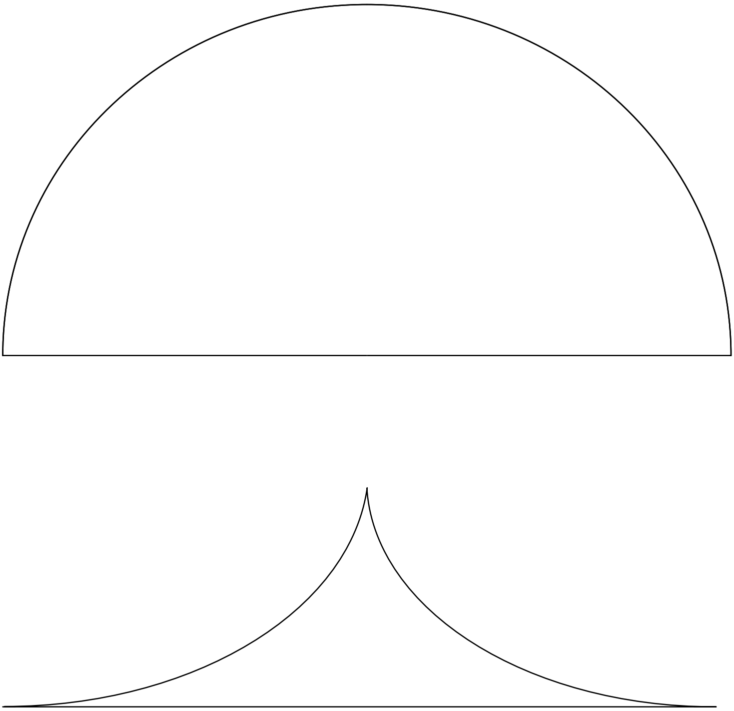


Figure 1.2. A convex deep 'pyramid' versus a concave shallow one

[11] So does learning, since once you have learned one fact, you need to be able to hold it unchanged while you learn others. Likewise, once you have learned one skill, you need to be able to isolate and preserve it while you acquire other skills; see Ghahramani and Wolpert, 1997; Manoel et al., 2002. I shall return to this point in Section 3.5.

Roughly speaking, then, we should expect there to be one distinct sub-system for each reliably recurring function that human minds are called upon to perform. (And whenever the function performed is complex, the sub-system in question should itself decompose into an array of sub-sub-systems, and so on.) But as evolutionary psychologists have often emphasized, the functions of the human mind are *myriad* (Tooby and Cosmides, 1992; Pinker, 1997). Focusing just on the social domain, for example, humans need to: identify degrees of relatedness of kin, care for and assist kin, avoid incest, select and woo a mate, identify and care for offspring, make friends and build coalitions, enter into contracts, identify and punish those who are cheating on a contract, identify and acquire the norms of their surrounding culture, identify the beliefs and goals of other agents, predict the behavior of other agents, and so on and so forth—plainly this is just the tip of a huge iceberg, even in this one domain. In which case the argument from biology enables us to conclude that the mind will consist in a *very great many* functionally distinct components, which is a (weak) form of massive modularity thesis.

If the argument for massive modularity depends upon counting (large numbers of) distinct cognitive functions, then it is important for us to know how *functions* are to be individuated for these purposes. How are we to tell, for example, whether identifying kin and identifying cheaters are two distinct functions (in which case we should expect two distinct processing mechanisms to underlie them), or whether they are really just one function (some generalized form of learning)? There are broadly two approaches that one can take to the individuation of cognitive functions. One is to do task analysis, showing that the abilities in question make structurally different demands on processing, requiring different sorts of algorithms in order to extract the knowledge or ability in question from the initial data or starting-point. This is the approach that will loom large in Chapter 2, where I shall argue on just these sorts of grounds that the learning challenges that animals face require *multiple* differently structured learning mechanisms, and not just one or a few.

There is, however, a second class of ways in which we can individuate cognitive functions, which enables us to keep the argument from general biological design separate from the argument from animal minds. This is by reflecting on the extremely powerful constraint that speed of processing places on the design of cognitive systems, as we shall see more fully in Sections 3.4 and 3.5 below. Wherever different materials need to be processed at the same time, it will be much more efficient to employ two distinct processing systems which can handle those materials in parallel, rather than relying upon a single general-purpose system which would have to tackle the tasks sequentially. For in the animal world generally, speed of learning, speed of decision-making, and speed of reaction time are crucial determinants of survival and reproduction. And just as we would therefore predict, and as everyone now acknowledges, the human brain is massively parallel in its organization and in its operations.

Before concluding this part of our discussion, let me note that it is important not to be misled by talk of 'mechanisms' and 'components' in the context of discussions of the massively modular mind. For these are supposed to be distinct functionally specialized cognitive systems, that is all. In particular, it is important not to think of modules as *objects*, by analogy with hammers and screwdrivers. (In this respect, comparing the massively modular mind to an 'adaptive toolbox' or to a 'Swiss army knife', in the way that evolutionary psychologists often do, is actually highly misleading.) For as Barrett (2006) argues, thinking in this way can lead critics of massive modularity—such as Woodward and Cowie (2004), and Buller (2005)—to attribute to modules properties that they don't (or at least needn't) have. In consequence, the critics end up attacking a straw man. Let me briefly elaborate.

If one conceives of modules by analogy with artifacts like a screwdriver, then one will naturally think that modules must be physically discrete from one another, that they aren't readily modifiable, and that they have been produced in accordance with some kind of design blueprint (in this case written in the genes). But modules are biological systems, and like most such systems they are likely to be built by co-opting and connecting in novel ways resources that were antecedently available in the service of other functions. In consequence, modules are likely to exhibit massive sharing of parts. This is still consistent with them being functionally specialized, as well as being independently effectible and independently disruptable. And modules (again like biological systems generally) are likely to show significant plasticity in the course of development and in response to environmental change. Moreover, they will develop through extensive and complicated gene–environment interactions, in which much of the 'information' that is used to build each system comes from the environment. Certainly it would be a mistake to think of modules as somehow 'pre-formed' in the genes (even in cases where they are significantly innate, given the sense of 'innate' sketched in Footnotes 3 and 5 above).

I have set out Simon's (1962) argument thus far as if it were an argument specifically about biological systems. But it is actually much broader, applying to complex functional systems quite generally, and to complex computer programs in particular.

3.4 Computer Science and AI

The basic reason why biological systems are organized hierarchically in modular fashion is a constraint of evolvability. Evolution needs to be able to add new functions without disrupting those that already exist; and it needs to be able to tinker with the operations of a given functional sub-system—either debugging it, or altering its processing in response to changes in external circumstances—without affecting the functionality of the remainder. Human software engineers have hit upon the same problem, and the same solution. (Although the language of modularity isn't so often used by computer scientists, the same concept arguably gets deployed under the heading of 'object-oriented programs'; see below.) In order that new functions can be added to a program, or in order that one part of it can be debugged, improved, or updated, but without any danger of introducing errors elsewhere, software engineers routinely modularize their programs. And for just these reasons, the automatic electronic control systems that manage complex telephone networks are always organized hierarchically out of modular sub-components, for example (Kamel, 1987; Coward, 2001).

Sometimes modularity is actually enforced by the computer language employed, although sometimes it isn't (Aaron Sloman, personal communication). Much low-level programming is still done using the language C, for example, which doesn't mandate modular organization. (Likewise for languages like Fortran.) But a good programmer will still try to write modular code, with well-defined interfaces between the different parts of the system. However, two of the most widely used languages nowadays are C++ and Java. Both support the use of well-defined interfaces *enforced* by the language, as do many other languages. Languages in this class are often described as 'object oriented'.

Thus many programming languages now require a total processing system to treat some of its parts as 'objects' which can be queried or informed, but where the processing that takes place within those objects isn't accessible elsewhere. This enables the code within the 'objects' to be altered without having to make alterations in code elsewhere, with all the attendant risks that this would bring; and it likewise allows new 'objects' to be added to the system without necessitating wholesale re-writings of code elsewhere. And the resulting architecture is regarded as well nigh inevitable (irrespective of the programming language used) once a certain threshold in the overall degree of complexity of the system gets passed.[12]

[12] Interestingly, since the need for modular organization increases with increasing complexity, we can predict that the human mind will be the *most* modular amongst animal minds, whereas the minds of insects (say) might hardly be modular at all. This is the reverse of the intuition shared by many philosophers and social scientists, who would be prepared to allow that animal minds might be organized along modular lines, while believing that with the appearance of the human mind most of that organization was somehow superseded and swept away.

AI researchers charged with trying to build intelligent systems, likewise, have increasingly converged on architectures in which the processing within the total system is divided up amongst a much wider set of task-specific processing mechanisms, which can query one another, make their outputs available to others, and many of which can access shared databases (personal communication: Mike Anderson, John Horty, Aaron Sloman). But many of these systems will deploy processing algorithms that aren't operated by any of the others. And most of them won't know or care about what is going on within the others. The fact that human designers of intelligent systems have converged on modular organization is evidence that the human mind, similarly, will be modular in its design.

In some respects the constraints on the design of computer-based systems are different from the constraints on the design of biological systems (especially brains), however. And this has implications for the sorts of modular organization that are likely to result. In particular, resource constraints are much more important in brains than within modern computers, and this differential has been increasing rapidly with developments in computing technology. Brains are very expensive to build and maintain, in comparison with other components of the body. Thus Aiello and Wheeler (1995) point out that the brain consumes energy at about eight times the rate that would be predicted from its mass alone (accounting for about 20% of the total). So adding extra processing power doesn't come cheap. Moreover, increases in brain size carry other sorts of cost as well, resulting from the consequent increases in head size. This is especially evident in the hominid line, where increased head sizes have resulted in much elevated dangers of maternal death during labor. They have also necessitated extended periods of maternal dependency, since hominid infants have to be born less mature than would be predicted from other biological measures (Barrett et al., 2002).

To some extent this resource constraint pulls in the opposite direction from the constraint of separate modifiability, and we should therefore predict that the actual design of the brain will involve some sort of trade-off between them (Coward, 2005). If minimizing energetic costs were the major design criterion, then one would expect that the fewer brain systems that there are, the better. But on the other hand the evolution of multiple functionality requires that those functions should be underlain by separately modifiable systems, as we have seen. As a result, what we should predict is that while there will be many modules, those modules should *share parts* wherever this can be achieved without losing too much processing efficiency (and subject to other constrains: see below). And indeed, there is now a great deal of evidence supporting what Anderson (forthcoming) calls 'the massive redeployment hypothesis'. This is the view that the components of brain systems are frequently deployed in the service of multiple functions.

A second sort of resource constraint, however, is speed. This often militates *against* sharing of parts, as we shall see. Brains are extremely slow in comparison with turn-of-the-century desktop computers. Information is propagated down the axons of the nerves in the human brain at speeds of only a few meters per second—below the 55 mph speed limit! So a signal passing down an axon of length ten centimeters, say (quite a common long-distance connection within the brain), will take around a tenth of a second on its own, even before time is allowed for electronic spread within the dendrites, and for synaptic transmission. The signal propagation rate within the microchip that forms the central processing unit of a standard desktop computer is about one *million* times greater than this (Christopher Cherniak, personal communication). Moreover, many real life processing tasks on which the organism's survival may depend will need to be *completed* within fractions of a second. (Think of reacting to an attack by a predator, for example.) The result is that we should predict massive *parallelism* in the functional organization of the brain. A great deal of evidence supports this prediction, too.

There are two sorts of circumstance in which speed constraints militate against modules sharing parts. One is where the processing sub-system in question would have to operate on different inputs, and generate distinct outputs, in the service of the two modules in question. For this means that the sub-system would then have to operate *sequentially*, significantly slowing the processing time for one of the two modules. Parts will therefore be shared, in general, only where the two containing systems need the *same* information to be generated from the *same* input at the same time. Otherwise we would expect two distinct sub-systems to evolve, which can operate in parallel. The other sort of circumstance in which speed counts against sharing is whenever the two down-stream consumer systems for a given sub-system are significantly spatially separated. In such cases it may be more efficient to build two distinct systems to do the job, even if they precisely replicate each other's processing. Hence parts are much more likely to be shared within two adjacent modules than within two distant ones.

How powerful is the pressure exerted on the design and organization of brains by energetic and temporal resource constraints? Notice that the length of any given neural connection is both positively correlated with mass (and hence with energy consumption) and negatively correlated with speed (since the distance traveled is greater). So if the pressure exerted by these resource constraints were powerful, then we would expect that the wiring diagram for the brain would minimize signaling distance. And indeed, it turns out that the wiring diagram for different areas of the cortex is almost as efficient in its layout as it is theoretically possible to be (Cherniak et al., 2004).

The resource constraints that distinctively constrain the design of brains (energetic and temporal), don't imply that the brain should be any *less* modular in its organization than other complex biological systems, then. But they do suggest that parts should be shared whenever this can be done without increasing signaling distances or processing time, and without too much loss of reliability. And they also suggest that there should be massive parallelism and duplication of structure whenever signaling distances get too great, or whenever different sorts of information need to be processed within the same brief time-frame. We should see these resource constraints as modulating and adding complexity to the 'one-function / one-module' principle that can be derived from the constraints of separate modifiability and separate evolvability, therefore, without altering that prediction in any fundamental way.

The biological argument for massive modularity, as discussed in Sections 3.1 through 3.3 above, might naturally be read as having the following form: (1) Biological systems are, when complex, massively modularly organized. (2) The human mind is a biological system, and is complex. (3) So the human mind will be massively modularly organized. But reflection on the *reasons why* biological systems are modular shows that the argument really has the following form: (1) Biological systems are designed systems, constructed incrementally. (2) Such systems, when complex, need to have massively modular organization. (3) The human mind is a biological system, and is complex. (4) So the human mind will be massively modular in its organization. And it now emerges that the argument from computer science and AI has exactly the same form, only with 'computational system' substituted for 'biological system' throughout.

So Simon's (1962) argument is really an argument from *design*, then, whether the designer is natural selection (in the case of biological systems) or human engineers (in the case of computer programs). It predicts that, in general, each element added incrementally to the design should be realized in a functionally distinct sub-system, whose properties can be varied independently of the others (to a significant degree, modulated by the extent to which component parts are shared between them). It should be possible for these elements to be added to the design without necessitating changes within the other systems, and their functionality might be lost altogether without destroying the functioning of the whole arrangement. And since there are *many* ancient and evolutionarily significant capacities of the human mind (as well as many capacities constructed by learning of various sorts—see Section 3.5), we should expect the human mind to be *massively* modular in its organization (using the weak sense of 'module').

3.5 A Design for Learning

The human mind (together with most animal minds and some AI systems, of course) is distinctive amongst designed systems in being designed for learning. Amongst the evolved modules that make up the human mind are many whose function is to generate and store new information of various sorts (both episodic memories and factual information), as well as building and retaining new skills. Indeed, many of the modules that make up human and animal minds are best characterized as learning systems of various sorts. (This will loom large in Section 4, when we come to consider a second line of argument in support of massive mental modularity, and again in Section 5, when we come to defend evolutionary psychology against its philosophical critics.)

So the argument from design suggests that the mind should contain multiple learning modules. But it might be objected again that a lot will turn, in evaluating that argument, on how the capacities of the mind are individuated. For example, if we treat 'learning who has cheated on a contract' as a distinct capacity from 'learning a social norm', then the design argument predicts that they will be realized in distinct modules. But if we just characterize both as 'learning', *simpliciter*, then why shouldn't there be just one system involved? Indeed, it might be urged that the argument from design is consistent with there being just *one* general learning mechanism in the human mind, which can be directed in the service of all the many different learning tasks.

In fact, however, the structure of the many different learning tasks that humans and other animals face is highly varied, and it is very doubtful indeed whether there could be just a single leaning mechanism capable of undertaking every one of them. (Stronger still, it is highly doubtful whether even the collective activity of a *few* leaning mechanisms would be adequate to the task.) These claims will be sketched in Section 4, and then developed at length in Chapter 2, in the course of presenting and defending what I call 'the argument from animals' in support of massive modularity. So at this point it should be acknowledged that the design argument gains strength from being integrated with the argument from animals. But even if we set this aside, it remains very implausible that there should be just one (or a few) mechanisms of learning.

Even if learning were everywhere and always the same—in such a way that the mechanism of learning that is used in one task could always in principle be used to undertake any other—we should *still* expect that the mind would contain many different versions, or instances, of that type of mechanism. One reason has to do with robustness of function: if there were just a single learning mechanism, then damage to it would cripple the organism; whereas if there are multiple learning mechanisms, each dedicated to a particular learning task, then damage to one can leave all the others intact. But the more important reason has to do with *speed* and *reliability* of learning, as I shall now argue.

If there were just a single learning mechanism, then it would presumably have to be deployed serially to undertake the various learning tasks. This would take time, and many opportunities for learning might consequently be lost. If there are multiple mechanisms, in contrast, then they can operate in parallel, greatly reducing the overall acquisition time. (And given the slow speeds characteristic of neuronal transmission—already mentioned in Section 3.4—this is a very significant constraint.) Nor would it be at all surprising that evolution should operate in this way. As Marcus (2004) points out, natural selection frequently operates through a process of copying sets of genes (and hence the structures that they build), before the new structure is turned to the service of a novel function.

For example, consider learning about the movements and interactions of physical objects (common-sense physics) and learning about the goals, thoughts, and intentions of human subjects (common-sense psychology). Even if we thought that the very same learning algorithms could serve both tasks equally well (which is actually *extremely* implausible), we should still predict the existence of two distinct learning mechanisms for these two domains. This is because events in those domains will vary independently of each other while often occurring simultaneously. If there were just a single learning mechanism, then it would need to operate on the physical and psychological aspects of a given event sequentially, or risk confounding them. This would retard learning, and might lead to many opportunities for learning being missed altogether when complex chains of events unfold in real time. The obvious design solution is to have two distinct copies of the learning system in question, each focused on a proprietary domain.

In addition to predicting that there will be multiple learning mechanisms, the argument from design predicts that, where the *products* of learning are multiple and complex, displaying significant internal organization, then those products themselves should have a modular character. Hence each learned element should have a realization that is distinct from the others, in such a way that new elements can be added through learning without interfering with those that already exist, and in such a way that the properties of any one element can be altered without altering the others. And the clearest examples of forms of learning that display such properties are learned skills of one sort or another, from reading and writing, through piano playing, cooking, to the various stages of kayak making.[13]

Consistent with these predictions, Ghahramani and Wolpert (1997) provide evidence of the modular decomposition of distinct visuomotor skills. They had subjects learn to reach to a perceived location that was displaced from actual (as if seen through a prism) from two distinct starting locations of their right hands, where the displacement was different for the two starting locations. The model they were testing assumed that a distinct module would be built for each of the two learned behaviors, and that subsequently a gating mechanism would take input from each of these modules (in the form of a weighted average) for starting positions of the hand that were intermediate between those that had been used in the initial training. The data that they obtained matched the predictions of this model very nicely, whereas those data proved inconsistent with a number of other models that assumed just a single learned system of some sort.

Likewise, Kharraz-Tavakol et al. (2000) set out to test the modular decomposition of acquired skills, on the assumption that if the different movements that make up a skill are stored as separate modules, then there should be transfers of learning to novel skills that recombine those movements in different ways. And this is just what they found, using a letter-writing task in which subjects first had to learn to write a letter from the Nashki alphabet (a precursor of the modern Arabic alphabet), and then had to learn to write that same letter rotated through 180 degrees. Manoel et al. (2002) produced further evidence in the same vein. They, too, had subjects learn to write a novel letter (this time a Chinese character), but then they embedded that task within a more complex character-writing assignment. They found, as predicted, that the sequencing and relative timing of movements from the original task remained invariant within the context of the new one, suggesting that it had been stored as a module that was left unaffected when embedded into a new skill.

The picture that emerges from these and other similar data, then, is that the components of acquired skills—like the modules of the mind more generally—are organized hierarchically out of motor-control systems that are constructed via learning. (See Wolpert et al., 2003, for a review.) So not only are the various learning systems of the mind realized in distinct modules, as the argument from design implies, but the products of those systems add yet further modules to the architecture of the mind, at least in the case of skill-learning.

[13] But memory, too, appears to display a modular organization, even when attention is confined just to long-term memory. Sites associated with memory are found in many different locations in the brain, with different regions associated with different forms of memory (Marcus, 2004). I shall return to develop this point in Chapter 2.6.

4 The Argument from Animals

Another line of reasoning supporting massive modularity starts initially from reflection on the differing task demands of the very different learning challenges that people and other animals must face, as well as the demands of generating appropriate fitness-enhancing intrinsic desires (Gallistel, 1990, 2000; Tooby and Cosmides, 1992, 2005). I shall provide just a sketch of this line of argument here. It will then be one of the tasks of Chapter 2 to establish its main premise: the massively modular organization of animal minds.

4.1 Extracting New Information

It is one sort of task to learn the sun's azimuth (its height in the sky at any given time of day and year) so as to provide a source of direction. It is quite another sort of task to perform the calculations required for dead reckoning, integrating distance traveled with the angle of each turn, so as to provide the direction and distance to home from one's current position. It is yet another sort of task to navigate via landmarks, recognizing each landmark and locating it on a mental map. And it is quite another task again to learn the center of rotation of the night sky from observation of the stars, extracting from it the polar north. These are all learning problems that animals can solve. But they require quite different learning mechanisms to succeed (Gallistel, 2000).

When we widen our focus from navigation to other sorts of learning problem, the argument is further reinforced. Many such problems pose computational challenges—to extract the information required from the data provided—that are distinct from any others. From vision, to speech recognition, to mind-reading, to cheater detection, to complex skill acquisition, the challenges posed are plainly quite distinct. So for each such problem, we should postulate the existence of a distinct learning mechanism, whose internal processes are computationally specialized in the way required to solve the task. It is very hard to believe that there could be any sort of *general* learning mechanism that could perform all of these different roles.

One might think that conditioning experiments fly in the face of these claims. But as we shall see more fully in Chapter 2, general-purpose conditioning is rare at best. Indeed, Gallistel (2000; Gallistel and Gibbon, 2001; Gallistel et al., 2001) has argued forcefully that *there is no such thing as* a general learning mechanism. Specifically, he argues that the results from conditioning experiments are best explained in terms of the computational operations of a specialized rate-estimation module, rather than some sort of generalized associative process.

4.2 Acquiring New Desires

Desires aren't learned in any normal sense of the term 'learning', of course. Yet much of evolutionary psychology is concerned with the genesis of human motivational states. This is an area where we need to construct a new concept, in fact—the desiderative equivalent of learning. Learning is a process that issues in true beliefs, or beliefs that are close enough to the truth to support (or at least not to hinder) inclusive fitness.[14] But desires, too, need to be formed in ways that will support (or not hinder) the inclusive fitness of the individual. Some desires are instrumental ones, of course, being derived from ultimate goals together with beliefs about the means that would be sufficient for realizing those goals. But it is hardly very plausible that all acquired desires are formed in this way.

Anti-modular theorists such as Dupré (2001) are apt to talk vaguely about the influence of surrounding culture, at this point. Somehow goals such as a woman's desire to purchase a wrinkle-removing skin-cream, or an older man's desire to be seen in the company of a beautiful young girl, are supposed to be caused by cultural influences of one sort or another—prevailing attitudes to women, perceived power structures, media images, and so forth. But it is left entirely unclear what the mechanism of such influences is supposed to be. How do facts about culture generate new desires? We are not told, beyond vague (and obviously inadequate) appeals to imitation (Campbell, 2002).

In contrast, evolutionary psychology postulates a rich network of systems for generating new desires in the light of input from the environment and background beliefs. Many of these desires will be 'ultimate', in the sense that they haven't been produced by reasoning backwards from the means sufficient to fulfill some other desire. But they will still have been produced by inferences taking place in systems dedicated to creating desires of that sort. A desire to have sex with a specific person in a particular context, for example, won't (of course) have been produced by reasoning that such an act is likely to fulfill some sort of evolutionary goal of producing many healthy descendants. Rather, it will have been generated by some system (a module) that has evolved for the purpose, which takes as input a variety of kinds of perceptual and non-perceptual information, and then generates, when appropriate, a desire of some

[14] This isn't meant to be a definition, of course. If there are innate beliefs, then evolution might also be a process that issues in true beliefs, but evolving isn't learning. What is distinctive of learning is that it should involve some method (not necessarily a *general* one, let alone one that we already have a name for, such as 'enumerative induction') for extracting novel information from the environment within at least the lifetime of the individual organism. And what distinguishes learning from mere triggering, is that it is a process that admits of a correct cognitive description—learning is a cognitive as opposed to a brute-biological process.