IS THE MIND A SYSTEM OF MODULES SHAPED BY NATURAL SELECTION?

Evolutionary psychology (EP) is a new field that attempts to understand the mind in evolutionary terms. It posits that the mind consists of many "modules," each one performing a very specific kind of task, and operating independently of the others. These modules evolved because they were good solutions to adaptive problems faced by our ancestors. Within this framework, we can gain an understanding of the mind by discovering these modules and the environmental pressures that led to their being naturally selected. Peter Carruthers, in his chapter, presents some of the empirical evidence that bolsters the EP program, and then responds to an important philosophical objection to the modular conception of mental architecture. Fiona Cowie and Jim Woodward concede that the mind has been shaped by natural selection and that the mind consists of a variety of distinct capacities; but surely EP is not merely asserting this banal point. A major problem, they contend, is that evolutionary psychologists are often unclear about what a "module" is supposed to be. They argue that there is no unambiguous sense of the word "module" that is faithful to the usage of evolutionary psychologists and that would also render the central claim of EP true. In particular, the assumption that individual modules can evolve in isolation from one another is at odds with much of what we know about the evolution of brains. This exchange shows how empirical and conceptual issues can merge, blurring the line between philosophy of science and science itself.

CHAPTER FIFTEEN

The Mind is a System of Modules Shaped by Natural Selection

Peter Carruthers

This chapter defends the positive thesis that constitutes its title. It argues, first, that the mind has been shaped by natural selection; and, secondly, that the result of that shaping process is a modular mental architecture. The arguments presented are all broadly empirical in character, drawing on evidence provided by biologists, neuro-scientists, and psychologists (evolutionary, cognitive, and developmental), as well as by researchers in artificial intelligence. Yet the conclusion is at odds with the manifest image of ourselves provided both by introspection and by common-sense psychology. The chapter concludes by sketching how a modular architecture might be developed to account for the patently unconstrained character of human thought, which has served as an assumption in a number of recent philosophical attacks on mental modularity.

15.1 Introduction: Evolutionary Psychology and Modularity

If we are to address the topic picked out by the title of this chapter, then there are two main questions for us to answer:

1 Is the human mind, as well as the human body and brain, a product of natural selection? Is the human mind an adaptation/collection of adaptations?
2 Supposing that the answer to question 1 is positive, has the effect of natural selection on the human mind been to impose on it a modular architecture?

The reasons why a positive answer should be returned to question 1 are quite straightforward, and are not disputed within the present debate. Everyone should agree that the human mind is an adaptation, provided that one thinks (i) that the mind is causally and explanatorily relevant in the production of behavior, and supposing that one

agrees (ii) that the mind is significantly structured. For assumption (i) will make the mind a major determinant of evolutionary fitness, and so a prime *candidate* to be shaped by natural selection. And since natural selection is the only serious contender for explaining the appearance of functional complexity in the biological world (Williams, 1966; Dawkins, 1976, 1986), assumption (ii) means that we shall have good reason to think that the mind has *actually* been shaped by evolution. These points will be taken for granted in what follows.

The main burden of this chapter will be to motivate a positive answer to question 2, then. This actually breaks down into two sub-questions: (a) *Does* the mind have a modular architecture? (b) Supposing that it does, is that architecture a product of natural selection or, rather, does it get created within each individual through general learning and some sort of process of gradual modularization? Most of my attention will be devoted to question (2a), in sections 15.2 and 15.3 below. In section 15.4, I shall turn briefly to question (2b).

The thesis to be defended here – that the mind consists largely of evolved modules – is the claim of "massive modularity," which has been proposed and argued for in recent decades by evolutionary psychologists (Gallistel, 1990; Tooby and Cosmides, 1992; Sperber, 1996; Pinker, 1999). It is important to realize at the outset that evolutionary psychology is a broad church (somewhat like utilitarianism), embracing a variety of different theoretical claims and approaches. So one should be wary of arguments that take the form: "Many evolutionary psychologists claim that *P*. *P* is implausible. So evolutionary psychology is implausible." (No one would ever dream of arguing like this against utilitarianism. One would, of course, have to explore other avenues and options within a utilitarian perspective before concluding with a negative verdict.) In what follows, I shall try to present the thesis of massive modularity in what seems to me its strongest light. But the reader should be aware that there are other ways in which the approach might be developed, and that some of these might eventually turn out to be better. Evolutionary psychology is best seen as a *research program* (in the sense of Lakatos, 1970), not a fixed body of theory.

15.1.1 What are modules?

A natural first thought would be that the answer to our main question might depend crucially upon what one means by "module." To the question, "Does the mind consist of evolved modules?" one might naturally reply, "That largely depends on what you understand by a 'module'." And in a way, this is obviously quite correct. However, it would be a mistake to clog up our inquiry by attempting an analysis of what we should mean by "module" in advance. As generally happens in science, a proper understanding of our terms should be the *outcome* of empirical and theoretical enquiry, not something which is stipulated at the start.

According to Fodor's now-classic (1983) account, for example, modules are stipulated to be domain-specific innately specified processing systems, with their own proprietary transducers, and delivering "shallow" (nonconceptual) outputs; they are held to be mandatory in their operation, swift in their processing, encapsulated from and inaccessible to the rest of cognition, associated with particular neural structures, liable to specific and characteristic patterns of breakdown, and to develop according to

a paced and distinctively arranged sequence of growth. But such an account was designed to apply specifically to *input* systems, such as vision or touch. Those in the cognitive sciences who have wished to extend the notion of "modularity" to apply to the central systems that generate new beliefs from old (such as our "theory of mind," or the faculty of common-sense psychology), or which generate new desires, would obviously have to modify that account a great deal. Yet many in philosophy remain strangely fixated on Fodor's original analysis, using it as a stick with which to beat some of the more ambitious claims for mental modularity (see, e.g., Currie and Sterelny, 1999).

I propose, in contrast, that we should begin our work with "module" understood in its weakest and loosest everyday sense, to mean something like, "isolable functional sub-component." Thus a company organized in modular fashion has separate units that operate independently and perform distinct functions. And a hi-fi system that can be purchased on a modular basis is one in which the separate component parts can be bought independently, in which some at least of those components can operate independently of the others – you can have the tape-deck without the CD player – and in which different versions of the same part can be substituted for others without altering the remainder. Note that even this weak notion of modularity isn't anodyne in relation to the mind. Specifically, functionalism about mental properties might be correct without it being the case that personal memory, say, forms an isolable type that can be damaged or lost while leaving the rest of the mind intact. And when this notion is conjoined with the claim that different modules will have at least partially distinct selective histories, the result is anything but anodyne. This weak everyday notion should be enough to get us started. In the end, a module will be whatever is warranted by the various *arguments for* mental modularity.

These arguments are of various different types, and have somewhat different strengths. And actually they warrant rather different notions of modularity too; so our final answer to the question concerning the meaning of "modularity" might have to be a multiple one. I will not have space to develop this point fully in the present chapter, but let me just assert the following. We should expect modules to be relatively *encapsulated*, in the sense of processing their inputs independently of most of the information stored elsewhere in the mind. We should expect many (but not necessarily all) central modules to be *domain specific*, in the sense of having been designed to operate on particular sorts of conceptual inputs or conceptual structures. We should expect that some (but only some) modules will deploy algorithms that are unique to them, not being redeployed elsewhere in the mind to subserve other functions. And, of course (if we are to believe in *evolved* modules), we should expect that the development of modules is innately channeled to some significant degree, and that many of them, at least, will be liable to local impairment.

Notice that we need *not* be committed to saying that modules are "elegantly engineered" processing systems, which have simple and streamlined internal structures, and which exist independently of other such systems. On the contrary, we should expect that many modules are *kludges* (Clark, 1987), recruiting and cobbling together in quite inelegant ways resources that existed antecedently. (This is routine in biology, where the evolution of any new structure has to begin from what was already present. Thus the penguin's flippers, used for swimming, evolved from the wings that it once

used for flight.) We should also expect many modules to consist of arrangements of sub-modules, which are linked together in complex ways to fulfill the function in question. (A glance at any of the now-familiar wiring diagrams for visual cortex will give the flavor of what I have in mind.) And, for similar reasons, we should expect many modules and sub-modules to have multiple functions, passing on their outputs to a variety of other systems for different purposes. In that case, distinct modules may share parts in common with one another. (This need not mean that they cannot be selectively damaged, since it may be possible to damage some of the links in a system without destroying any of the shared components.)

I shall begin our discussion of the arguments for mental modularity in section 15.2, with a number of broadly programmatic/methodological lines of reasoning for the conclusion that we should *expect* the mind to be modular in structure. In section 15.3 I shall look briefly at some of the evidence – developmental, pathological, and experimental – for thinking that the mind really does have such an architecture. Then in section 15.4 I shall consider whether modular architectures are innately structured adaptations, or result from some sort of process of modularization-through-learning.

Finally, in section 15.5 I shall take up what seems to many people to be the most powerful objection to a modular conception of the human mind, deriving from the nondomain-specific and unconstrained character of human thinking. (We can combine together highly disparate concepts in our thoughts, and we can apparently do so at will.) In that section, I shall sketch a language-involving cognitive architecture according to which a modular natural language faculty serves to integrate the outputs of other modular systems, and where that language faculty can be used to generate new contents in imagination. But there won't be space, here, to explain these ideas fully; a mere sketch will have to suffice (for further development, see Carruthers, 2003b, 2003c, forthcoming).

15.1.2 Belief/desire psychology and learning

There is one final point that I want to emphasize and develop in this introduction, before we get down to the arguments. This is that, contrary to many people's first impressions, evolved modules aren't at all inconsistent with learning. Opponents of evolutionary psychology (e.g., Dupré, 2001) are apt to suppose that what is at issue are a suite of evolved *behavioral dispositions*. Hence they think that such an approach will minimize the role of learning in both human development and in mature behaviors. And they think, in consequence, that evolutionary psychology is incapable of explaining the distinctive *flexibility* of human behavior. But nothing could be further from the truth. There are at least two distinct confusions here, which need to be sorted out. The first is that evolutionary psychology is concerned not with behavioral dispositions *per se* but, rather, with the cognitive systems that operate and interact to produce those dispositions. (In fact, this is the crucial difference between evolutionary psychology and the earlier scientific research program of sociobiology.) And the second is that most of these postulated systems are actually systems of *learning*. I shall establish these points in turn. In this case, rapid assimilation of new information leading to flexibility in behavior is exactly what we should predict.

Evolutionary psychology takes quite seriously a belief/desire (or an information/goal) organization of psychological systems. This is true even in the case of insects, where it turns out that the desert ant has states representing that a food source is 44.64 meters northeast of its nest on a bearing of 16.5 degrees, say, which it can deploy either in the service of the goal of carrying a piece of food in a straight line back to the nest, or in returning directly to the source once again – or, in the case of bees, when the goal is to inform other bees of the location of the food source (Gallistel, 1990, 2000). What has been selected for in the first instance, on this view, are systems for generating beliefs and desires; the behaviors that result from those beliefs and desires can be many and various. (This isn't to deny that insects and other animals will *also* have a suite of innate behaviors and fixed action patterns, of course, in cases in which flexibility isn't required.)

Seen like this, it becomes obvious that the systems in question are learning systems. (At least, this is so in connection with belief-generation; I shall return to the case of desires below.) The system in the ant which uses dead reckoning to figure out the exact distance and direction of a food source in relation to the nest (given the time of day and position of the sun) is a system for *learning* that relationship, or for acquiring a *belief* concerning that relationship. Similarly, the language system in humans is, in early childhood, designed for *learning* the syntax and vocabulary of the surrounding language (albeit using a unique information-rich set of learning algorithms); and in older children and adults it is used to *learn* what someone has just said on a given occasion, extracting this information from patterns in ambient sound.

As already noted, many of these learning systems are hypothesized to deploy learning algorithms that are unique to them. The computations necessary to extract a directional bearing from information about the sun's position in the sky and the time of day and year are obviously very different from those needed to generate the syntax of English from samples of English discourse. But some systems might operate using learning algorithms that are used many times over in the brain. Thus the system that is used in acquiring vocabulary in childhood, and the system in the visual cortex that extracts object shape, may both be designed to do Bayesian inference (Lila Gleitman, personal communication). This is still consistent with the weak sense of "modularity" that we have adopted, since we can claim that the systems in question are highly restricted in their input and output conditions, and are relatively encapsulated in their processing. Each has been designed to take specific kinds of data as input and to generate a certain sort of output. Whether the algorithms used are unique or replicated many times over in other learning systems is much less significant.

15.1.3 Practical reasoning

If it is being taken for granted that belief/desire psychology applies throughout the animal kingdom, however, then a system for practical reasoning is being taken for granted too. This gives rise to a natural objection. This is that practical reasoning can't be modular, because if an organism possesses just a single practical reasoning system (as opposed to distinct systems for distinct domains), then such a system obviously couldn't be domain specific in its input conditions. However, such a system

could, nevertheless, be highly restricted in terms of its processing database, and this is all that is needed to secure its modular status in the sense that matters.

For example, we can imagine the following very simple practical reasoning module. It takes as input whatever is the currently strongest desire, P. It then initiates a search for beliefs of the form "$Q \rightarrow P$," cueing a search of memory for beliefs of this form and/or keying into action a suite of belief-forming modules to attempt to generate beliefs of this form. When it receives one, it checks its database to see if Q is something for which an existing motor program exists. And it initiates a search of the contents of current perception to see if the circumstances required to bring about Q are actual (i.e., to see not only whether Q is something doable, but doable here and now). If so, it goes ahead and does it. If not, it initiates a further search for beliefs of the form "$R \rightarrow Q$," or of the form "Q" (for if Q is something that is already happening or is about to happen, the animal just has to wait in order to get what it wants – it doesn't need to do anything more), and so on. Perhaps the system also has a simple stopping rule: if you have to go more than n number of conditionals deep, stop and move on to the next strongest desire.[1]

Note that the sort of module described above would be input unrestricted. Since, in principle, almost anything can be the antecedent of a conditional whose consequent is something desired (or whose consequent is the antecedent of a further conditional whose consequent . . . , and so on), any belief can in principle be taken as input by the module. But what the module can *do with* such inputs is, I am supposing, extremely limited. All it can do is the practical reasoning equivalent of *modus ponens* (I want P; if Q then P; Q is something I can do here-and-now; so I'll do Q), as well as collapsing conditionals ($R \rightarrow Q$, $Q \rightarrow P$, so $R \rightarrow P$), and initiating searches for information of a certain sort by other systems. It can't even do conjunction of inputs, I am supposing, let alone anything fancier. Would such a system deserve to be called a "module," despite its lack of input-encapsulation? It seems to me plain that it would. For it could be a dissociable system of the mind, selected for in evolution to fulfill a specific function, genetically channeled in development, and with a distinct neural realization. And because of its process-encapsulation, its implementation ought to be computationally tractable.

However, it is plain that human practical reasoning is not at all like this. There seem to be no specific limits on the kinds of reasoning in which we can engage while thinking about what to do. We can reason conjunctively, disjunctively, to and from universal or existential claims, and so forth. And contents from all the various allegedly modular domains can be combined together in the course of such reasoning. This makes the practical reasoning system look like an archetypal holistic, a-modular (and hence computationally *in*tractable) central system, of just the sort that Fodor thinks make the prospects for a worked-out computational psychology exceedingly dim (Fodor, 1983, 2000). I shall return to this problem briefly in section 15.5 below, and again at greater length in future work (Carruthers, forthcoming).

1 Although I have described this practical reasoning module as "very simple" (in relation to the sorts of reasoning of which humans are actually capable), its algorithms would not, by any means, be computationally trivial ones. Each of the component tasks should be computationally *tractable*, however – at least, so far as I can see (although intuitions of computational tractability are notoriously unreliable).

15.1.4 Acquiring 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 in which 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) individual fitness.[2] But desires, too, need to be formed in ways that will support (or not hinder) individual fitness. 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.

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 backward 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 nonperceptual information, and then generates, when appropriate, a desire of some given strength. (Whether that desire is then acted upon will, of course, depend upon the other desires that the subject possesses at the time, and on his or her relevant beliefs.)

The issue is not, then, the extent to which learning is involved in the causation of forms of human behavior. Both evolutionary psychologists and their opponents can agree that learning is ubiquitous. Nor is the issue even whether the algorithms used in learning are domain general (being suitable for extracting many different kinds of information) or, rather, are specific to a particular domain (although the domain speci-

---

2   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 which we already have a name for, such as "enumerative induction") for extracting 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.

ficity of learning algorithms is certainly an interesting question). Rather, it is whether the mechanisms that engage in learning are multiple, and have been specifically and separately designed by evolution to extract information (or to generate fitness-enhancing goals) concerning a given domain.[3]

15.2 Why we should Expect the Mind to be Modular

In this section, I shall review a number of general arguments for the thesis that we should expect the evolved structure of the mind to be modular. Considerations of space mean that our discussion of these arguments will have to be quite brisk. For more detailed elaboration, see Tooby and Cosmides (1992).

15.2.1 The argument from biology

One argument for massive modularity appeals to considerations deriving from evolutionary biology in general. The way in which evolution of new systems or structures characteristically operates is by "bolting on" new special-purpose items to the existing repertoire. First, there will be a specific evolutionary pressure – some task or problem that recurs regularly enough and that, if a system can be developed which can solve it and solve it quickly and reliably, will confer fitness advantages on those who possess that system. Then, secondly, some system that is targeted specifically on that task or problem will emerge and become universal in the population. Often, admittedly, these domain-specific systems may emerge by utilizing, coopting, and linking together resources that were antecedently available; and hence they may appear quite inelegant when seen in engineering terms. But they will still have been designed for a specific purpose.

Another way of putting the point is that in biology generally, distinct functions predict distinct (if often overlapping) mechanisms to fulfill those functions (Gallistel, 2000). No one supposes that there could be a general-purpose sensory organ, which could fulfill all of the functions of sight, hearing, taste, touch and smell. On the contrary, what we expect to find – and what we do find – are distinct organs, specialized for the distinctive structure of each domain, and which have been shaped by natural selection to fulfill the function in question. Similarly, no one expects to find that there is a general-purpose organ fulfilling the functions of both a heart and a liver, or fulfilling the functions of digestion and respiration. Likewise, then, in the case of the mind: one should expect that distinct mental functions – estimating numerosity, predicting the effects of a collision, reasoning about the mental states of another person, and so on – are likely to be realized in distinct cognitive learning mechanisms, which have been selected and honed for that very purpose.

3 Note that once the evolutionary psychologist's thesis is seen to be restricted to the genetically channeled character of a suite of learning mechanisms, rather than the innateness of most of the contents of the mind or anything of that sort, then the force of the argument from the relative paucity of genes in relation the large number of neurons in the brain (Elman et al., 1996) is much reduced. And see Marcus (forthcoming) for a very nice demonstration of how a small number of genes can be used to build banks of distinctively organized neurons.

15.2.2 Could general learning evolve?

A different – though closely related – consideration is negative, arguing that a general-purpose problem-solver *couldn't evolve*, partly because it would always be outcompeted by a suite of special-purpose conceptual modules. A general-purpose learning system would, inevitably, have to be very slow and unwieldy in relation to any set of domain-specific competitors. One point here is that such a system would face the problem of combinatorial explosion as it tried to search through the maze of information and options available to it (see section 15.2.3 below). Another point is that it would either have to process many different learning tasks at once using the same learning apparatus (and how would *that* be organized, except on a modular basis?), or it would have to tackle those tasks sequentially, leading to significant delays and bottlenecks.

Yet another point, however, is that many learning tasks simply cannot be solved without substantial innate assumptions about the domain being learned; which argues for the existence of a number of distinct learning mechanisms within which these assumptions can be embedded. The most famous such domain is language, where so-called "poverty of stimulus" arguments show that there must be an innately structured language acquisition device, in order for language learning to be possible (Chomsky, 1988; Crain and Pietroski, 2001; Laurence and Margolis, 2001). But similar arguments can be constructed for other domains: such as common-sense psychology, where a rich causal story has to be extracted somehow (and by the age of four, by all normal children) from the behavioral and introspective data (Carruthers, 1992; Botterill and Carruthers, 1999); and also for normative reasoning, where children again manifest an early grasp of a highly abstract set of concepts and principles, this time lacking any straightforward empirical basis (Cummins, 1996; Núñez and Harris, 1998; Dwyer, 1999).

Further arguments relate more specifically to the mechanisms charged with generating desires. For many of the factors which promote long-term fitness are too subtle to be noticed or learned within the lifetime of an individual; in which case there couldn't be a general-purpose problem-solver with the overall goal "promote fitness" or anything of the kind (Tooby and Cosmides, forthcoming). On the contrary, a whole suite of fitness-promoting goals will have to be provided for, which will then require a corresponding set of desire-generating computational systems.

Consider, for example, the surprising prediction that in certain social species where the reproductive success of males can vary a great deal as a function of fitness and status (such as deer and humans), females should vary their reproductive strategies (Trivers and Willard, 1973). Low-fitness, low-status, females should invest in female offspring, since these offer their best chance of passing on their genes to future generations; high-fitness, high-status, females, by contrast, should invest in male offspring. In deer, it seems that the mechanism by which this is effected is noncognitive, somehow altering the birth ratios, since does that are in poor physical condition are more likely to *give birth to* female offspring. In humans, on the other hand, it would appear that the mechanism is a cognitive one, operating via the mother's desire (or absence of desire) to have another child quickly, and/or via her degree of investment in the child that she has. (But see Grant (1998) for evidence that suggests that human

mothers may also be capable of controlling the sex of an unborn infant, to some small degree.) Thus, low-status women in the USA (measured by income and by the presence or absence of an investing male partner) whose first child is a daughter are likely to wait longer before giving birth to a second child; they are more likely to breast-feed a female child; and they will also breast-feed for a significantly greater time; with high-status women displaying the converse pattern (Gaulin and Robbins, 1991).

Roughly speaking, the prediction is that low-status women should want daughters, whereas high-status women should want sons. And this prediction does seem to be supported, both by the results mentioned above and by extensive data from other measures of parental investment around the globe, including rates of male and female infanticide (Hrdy, 1999). But, of course, no one thinks that the women are reasoning backwards, from a desire to be as reproductively successful as possible to the means most likely to realize that goal. For it requires sophisticated evolutionary–biological reasoning to figure the thing out. Rather, the suggestion should be that evolution has favored a desire-generating mechanism in human women that is sensitive to a variety of indications of expected fitness, and which has been selected for because of its long-term effects on reproductive success.

15.2.3 The argument from computational psychology

Perhaps the most important argument in support of mental modularity for our purposes, however, simply reverses the direction of Fodor's (1983, 2000) argument for pessimism concerning the prospects for computational psychology. It goes like this:

(1) The mind is computationally realized.
(2) A-modular, or holistic, processes are computationally intractable.
(3) So the mind must consist wholly or largely of modular systems.

Now, in a way Fodor doesn't deny either of the premises in this argument; and nor does he deny that the conclusion follows. Rather, he believes that we have independent reasons to think that the conclusion is false; and he believes that we cannot even *begin* to see how a-modular processes could be computationally realized. So he thinks that we had better give up attempting to do computational psychology (in respect of central cognition) for the foreseeable future. Fortunately, however, his reasons for thinking that central cognition is holistic in character are poor ones. For they depend upon the assumption that scientific inquiry (social and public though it is) forms a useful model for the processes of individual cognition; and this supposition turns out to be incorrect (Carruthers, 2003a; but see also section 15.5 below for brief discussion of a related argument against mental modularity).

Premise (1) is the guiding assumption that lies behind all work in computational psychology, hence gaining a degree of inductive support from the successes of the computationalist research program. Just about the only people who reject premise (1) are those who endorse an extreme form of distributed connectionism, believing that the brain (or significant portions of it, dedicated to central processes) forms one vast

connectionist network, in which there are no local representations. The successes of the distributed connectionist program have been limited, however, mostly being confined to various forms of pattern recognition; and there are principled reasons for thinking that such models cannot explain the kinds of one-shot learning of which humans and other animals are manifestly capable (Horgan and Tienson, 1996; Marcus, 2001). Indeed, even the alleged neurological plausibility of connectionist models is now pretty thoroughly undermined, as more and more discoveries are made concerning localist representation in the brain.

Premise (2), too, is almost universally accepted, and has been since the early days of computational modeling of vision. You only have to *begin* thinking in engineering terms about how to realize cognitive processes in computational ones to see that the tasks will need to be broken down into separate problems that can be processed separately (and preferably in parallel). And this is, indeed, exactly what we find in the organization of the visual system. What this premise then does, is to impose on proposed modular systems quite a tight *encapsulation* constraint. For any processor that had to access the full set of the agent's background beliefs (or even a significant subset thereof) would be faced with an unmanageable combinatorial explosion. We should therefore expect the mind to consist of a set of processing systems which are not only modular in the sense of being distinct isolable components, but which operate in isolation from most of the information that is available elsewhere in the mind. (I should emphasize that this point concerns not the *inputs* to modular systems but, rather, the *processing databases* which are accessed by those systems in executing their algorithms; see Carruthers (2003a) and Sperber (2003).)

Modularism is now routinely assumed by just about everyone working in artificial intelligence, in fact (McDermott, 2001). So anyone wishing to *deny* the thesis of massive modularity is forced to take on a heavy burden. It must be claimed either that minds *aren't* computationally realized, or that we haven't the faintest idea how they can be. And, either way, it becomes quite mysterious how mind can exist in a physical universe. (This isn't to say that modularism doesn't have it's own problems, of course. The main ones will be discussed in section 15.5 below.) Modularism in psychology is now warranted in the same sense and to the same degree as the assumption of *mechanism* in biology was warranted prior to the discovery of the double-helix structure of DNA. Biologists in the middle part of the twentieth century were surely justified in believing that the laws of inheritance must be realized *somehow* in biochemical mechanisms, although they couldn't yet say how. In the same way, we are now warranted in believing that the mind must be realized in the operations of a set of modular computational processes, even though there is much that we cannot yet explain.

15.3 Evidence that the Mind is Modular

There are powerful arguments of a general sort, then, for the conclusion that we should *expect* the mind to have a modular organization. In this section, we will review some of the evidence that this expectation is actually fulfilled. Once again, our exposition will have to be extremely brisk.

15.3.1 Developmental evidence

According to the modularist hypothesis, the human mind is made up of isolable components. And the expectations are that such a modular architecture is innate or innately channeled, and also that many of the modular components will operate in accordance with algorithms that are innate, or will make innate assumptions about the domains that they concern.

A variety of kinds of developmental evidence bears on, and supports, these proposals. One point is that developmental psychologists now mostly agree (in marked contrast with the earlier views of Piaget and his supporters) that cognitive development is a domain-specific process. It proceeds at different speeds in different domains (naïve physics, naïve psychology, naïve biology, mathematical understanding, and so on), and the cognitive structures that extract information concerning these domains would seem to be very different from one another. Instead of advancing on a broad front through some sort of general-learning process, it seems that different aspects of our cognition are acquired according to their own separate timetables and trajectories.

Another point is that some degree of competence in at least some of these domains is demonstrable at a very early stage in infancy – in some cases as young as four months. There is now robust evidence of early competence in a simple form of contact-mechanics, as well as in the rudiments of social understanding (Spelke, 1994; Baillargeon, 1995; Woodward, 1998; Phillips et al., 2002). And in other domains too, children acquire competence remarkably early considering the abstract non-observational character of the concepts involved. Thus, children appear to have a good understanding of normative notions such as "should," "must," "permissible," and so on by the age of three or four (Cummins, 1996; Núñez and Harris, 1998).

Under this general heading also fall the various "poverty of stimulus arguments," which have been run so decisively in the case of linguistic knowledge, but which can also be mounted with a good deal of plausibility in other domains too, such as "theory of mind" and moral belief (Carruthers, 1992; Dwyer, 1999). To the extent that it is very hard to see how children could acquire their competencies by the ages at which they do – using only general-learning systems to do it, and given the sorts of evidence available to them – it will be plausible to postulate an innately channeled domain-specific learning module to do the job.

15.3.2 Pathological evidence

The moral of the evidence from human pathology (whether developmental or resulting from later brain injury) is, roughly speaking, that *everything dissociates from everything else* (Shallice, 1988; Tager-Flusberg, 1999). (The data here must inevitably be messy and complicated, of course. For it is known that the same genes can be involved in a number of distinct functions, and any brain damage can be more or less extensive, also having effects on multiple functions at once.) In development, language can be damaged while everything else remains normal ("specific language impairment," or SLI); theory of mind can be damaged while language and physical/spatial thinking are normal (autism); and both theory of mind and language can

be normal while physical/spatial thought are severely damaged (Williams syndrome). Moreover, so-called "general intelligence" can be very impaired while other systems are relatively spared (Downs syndrome).

Similarly amongst adults, aphasias can involve severe loss of linguistic function while much else remains undamaged. Thus aphasic people can often still run many aspects of their own lives, and interact successfully with other people and with the physical world. One severely a-grammatic aphasic man, for example, has been shown to have intact theory of mind abilities, and also to be quite adept at reasoning about physical causes, such as identifying the locus of breakdown in a complex machine (Varley, 1998, 2002). There is also evidence that brain-damaged individuals can lose their capacity to reason normally about biological kinds (folk-biology) while all else is left intact (Job and Surian, 1998), and that a capacity to reason about social contracts can be lost independently of the ability to reason about risks and dangers (Stone et al., 2002); and so on and so forth.

15.3.3 Experimental evidence

There is not as much experimental evidence bearing on the question of modularity of mind as there could be, since most investigators have not thought to go looking for such evidence. This is because experimental psychology remains just as dominated by the empiricist and general-learning-theory assumptions that continue to exert such a hold over philosophy and much of the social sciences. What evidence there is, however, is highly suggestive.

One piece of evidence concerns the existence of a geometric module in both rats and humans (and presumably in many other mammalian species, at least). Rats shown a food source in a rectangular enclosure, who are then disoriented and replaced in that space, will search equally often in the two geometrically equivalent corners (e.g., those having a long wall on the left and a short wall on the right). And they do this despite the fact that there can be highly salient cues, which rats can perfectly well recognize and use in other circumstances, such as heavy scenting or patterning of one of the walls. It appears that, in conditions of disorientation at least, rats cannot integrate geometric information with information of other kinds (Cheng, 1986).

In these same circumstances, human children before the age of six or seven behave just like rats. When disoriented they rely only on geometric information even if, for example, one wall of the room is brightly colored while the others are neutral (Hermer and Spelke, 1994, 1996). And it turns out that human adults, too, are subject to just these effects when they are required to shadow speech (tying up the resources of the language faculty), but not when they have to shadow a complex rhythm (Hermer-Vazquez et al., 1999). In fact, these results provide one of the main sources of evidence for the thesis to be sketched in section 15.5 below, that it is the language faculty that enables information from a variety of otherwise-isolated modular systems to be integrated into a single (natural language) representation.

Another set of experimental evidence concerns the existence of distinct systems for reasoning about social contracts, on the one hand, and about risks and threats, on the other. (Each of these systems was originally predicted on the basis of evolutionary considerations.) When presented with problems that are structurally exactly

similar, except that one concerns a social contract of some sort (often involving the possibility of cheating) and the other concerns a significant risk to self or other, people will adopt very different reasoning strategies. Indeed, presented with just one social contract problem, or one risk-problem, people will reason differently depending upon which role in the contract or situation they are cued to identify with (Cosmides and Tooby, 1992; Gigerenzer and Hug, 1992; Fiddick et al., 2000). (Of course, these data do not *per se* entail the *modularity* of the systems in question. But standard practice in science is that surprising predictions that are made by a theory and turn out to be correct then serve to support that theory even though they do not entail it.)

Finally, there is evidence showing that even the very heartland of empiricist learning-theory – namely classical or associationist conditioning – is actually subserved by a special-purpose computational system that is designed to predict varying temporal contingencies, and which was selected for because of the role that it plays in successful foraging (Gallistel, 2000). There are a number of well-established findings that are important here. One is that when intervals between training trials are kept proportional to the delay between stimulus and reinforcement, then increased delays in reinforcement have no effect on the rate of acquisition. Another is that the number of reinforcements required for acquisition of a new form of behavior is left unaffected by inserting significant numbers of unreinforced trials for every reinforced one. These and other facts are extremely puzzling from a perspective which sees learning as matter of building associative strengths; whereas they can readily be explained within a computational model which assumes that what the animals are really doing is estimating likelihoods and calculating rates of return (Gallistel and Gibson, 2001).

15.4 Adaptation versus Learned Modularization

There is good reason to think, then, that the human mind has a modular organization, to some significant degree. But what of the suggestion that modularization may actually be the outcome of some sort of general-learning process, a product of *over-learning*? On this model, all that would be given at the outset of development are some general-learning abilities and a suite of domain-specific attention biases, which together serve to build a set of organized and automatically operating bodies of knowledge and skill. Then theory of mind, for example, would be a module in the same sense and to the same extent that chess-playing ability in an experienced Grand Master is a module (Karmiloff-Smith, 1992).

Such accounts face a great many difficulties, however. One is that they fail to address the full range of arguments sketched in section 15.2 above for the conclusion that we should expect the mind to have a given (innate) mental architecture. Another is that, in postulating that development is driven by general-learning abilities, the approach has severe difficulty in accounting for the fact that developmental profiles are similar across individuals and across cultures, despite wide variations in the richness and quantity of stimuli that subjects have experienced, and despite equally wide variations in general intelligence (which one would think should correlate well with general-learning ability).

In fact, the thesis of gradual modularization is a product of confusion concerning the relevant alternatives. The contrasting evolutionary psychology view is *not* that modules are there, fully formed, at birth and are realized in specific lumps of neural tissue, in the way that Elman et al. (1996) suppose. So the evidence of neural plasticity in the developing cortex, for example, is simply irrelevant (Samuels, 1998). Rather, the view is that evolved modules develop according to genetically channeled time-scales and profiles, emerging as learning-mechanisms that will serve to build increasingly elaborate bodies of knowledge throughout the lifetime of the individual. When sent up against this, its real opponent, the thesis of modularization through general learning, is not, in my view, a serious contender.

15.5 A Problem for Modularity: The Unconstrained Character of Thought

The arguments that the mind will contain at least *a great many* evolved modules are compelling, then. But what is the full extent of the mind's modularity? It seems obvious to many people that our minds must at least contain a very large and significant central arena, in which thoughts are formed and inferences drawn, which is *non*modular in character. Now this isn't the argument for the holistic character of belief-formation (Fodor, 1983, 2000), which we considered briefly and dismissed in section 15.2.3 above. Rather, it is the fact that we are, manifestly, unconstrained in the way that we can combine together concepts in our thoughts, crossing any putatively modular boundaries. I can be thinking about beliefs one moment and horses the next, and then wonder why I am thinking about both horses and beliefs, in effect then combining concepts from these disparate domains into a single thought.

If, in contrast, the human mind were wholly constructed out of modules, then one might expect that there would be severe limits on the structure and complexity of the kinds of thought that we can think. For some, at least, of these modular systems would be domain specific in character, only handling a given range of proprietary concepts. And for sure there should be limits on information-flow through a modular archi-tecture, since one would expect that while some modules would provide their outputs as input to some others, not every module would be linked with every other one. In that case, there should be some combinations of content that we would find it diffi-cult or impossible to entertain.

Now, one response to this difficulty would be to concede that there is, indeed, a nonmodular arena in which thoughts can be formed; but to continue to insist that this is embedded within an otherwise modular architecture (Cosmides and Tooby, 2001). A better-motivated response, however, is to claim that integration of thought-contents across modular domains is actually subserved by an existing module, namely the natural-language faculty (Carruthers, 1998, 2003b; Hermer-Vazquez et al., 1999). For almost everyone agrees that the language faculty is a distinct module of the mind; yet it is manifest that it would need to take inputs from any other conceptual modules, so that the outputs of those modules should be reportable in speech. The language

system will be ideally *positioned*, then, to facilitate the integration of modular contents; and the abstract and recursive nature of natural language syntax would serve to make such an integration possible in reality.

Three major difficulties remain. One is that not only can we *form* cross-modular thoughts, but we can do new things *with* them – we can use them as premises in reasoning, derive new information from them, and so on. Would this require a suite of nonmodular consumer systems, positioned downstream of the natural language representations (a nonmodular central arena once again)? Arguably not. A case can be made that the further use of cross-modular linguistic representations can be explained in terms of the deployment of a variety of existing modular processes, with perhaps some minor additions and adjustments (Carruthers, 2003c).

The second problem doesn't so much concern the *use* of cross-modular thoughts but, rather, their *creation* or mode of generation. We seem to be able to frame thoughts with arbitrarily novel contents in fantasy and imagination, for example. (So I can easily suppose – and perhaps for the first time in all of human history – that there is a red dragon on the roof, dreaming of diamonds.) How is this to be explained? Would *this* require some sort of radically a-modular thought-forming arena? Again, arguably not. Admittedly, we do have the capacity to *suppose*, and we can put together concepts in novel and creative ways in our suppositions. But this capacity might be a relatively simple addition, built onto the back of the language faculty; and it may be that the function of pretend play in childhood is precisely to construct and develop it (Carruthers, 2002).

Finally, we need to explain how distinctively human forms of practical reasoning are possible, built on the back of the simple processing-encapsulated practical reasoning module which we inherited from our ancestors, together with interactions involving the language system as envisaged above. Here too, it may be possible to construct an account that remains faithful to the spirit of massive modularity (Carruthers, forthcoming).

It should be acknowledged that these are hard problems. But then *everyone* here faces hard problems. As noted earlier, if we gave up on massive modularity then we might lose one set of problems, but we would, instead, face the task of explaining how holistic, a-modular, processes can be computationally realized. Since no one currently believes that *this* problem can be solved, it seems better to continue exploring the resources available to a massive modularist.

15.6 Conclusion

I have argued that there is a powerful case to be made in support of the thesis that forms the title of this chapter, although there has not been the space here to do more than sketch the outlines of that case. There is good reason to think that natural selection has imposed on the human mind a modular architecture. Moreover, there are no overwhelming reasons for thinking otherwise. I should emphasize, however, that the case I have been making is a broadly empirical one, and is therefore subject to empirical falsification. And in the end, of course, our question won't be settled by philosophers, but by scientists.

Acknowledgments

I am grateful to the participants at a Georgetown University philosophy colloquium for discussion of this material.

Bibliography

Baillargeon, R. 1995: Physical reasoning in infancy. In Gazzaniga, op. cit., pp. 181–204.

Barkow, J., Cosmides, L., and Tooby, J. (eds.) 1992: *The Adapted Mind*. New York: Oxford University Press.

Botterill, G. and Carruthers, P. 1999: *The Philosophy of Psychology*. Cambridge: Cambridge University Press.

Carruthers, P. 1992: *Human Knowledge and Human Nature*. Oxford: Oxford University Press.

—— 1998: Thinking in language?. In P. Carruthers and J. Boucher (eds.), *Language and Thought*. Cambridge: Cambridge University Press.

—— 2002: Human creativity. *The British Journal for the Philosophy of Science*, 53, 225–49.

—— 2003a: Moderately massive modularity. In A. O'Hear (ed.), *Mind and Persons*. Cambridge: Cambridge University Press.

—— 2003b: The cognitive functions of language. *Behavioral and Brain Sciences*, 25 (in press).

—— 2003c: On Fodor's problem. *Mind and Language*, 18, 502–23.

—— forthcoming: Practical reasoning in a modular mind.

Cheng, K. 1986: A purely geometric module in the rat's spatial representation. *Cognition*, 23, 149–78.

Chomsky, N. 1988: *Language and Problems of Knowledge*. Cambridge, MA: The MIT Press.

Clark, A. 1987: The kludge in the machine. *Mind and Language*, 2, 277–300.

Cosmides, L. and Tooby, J. 1992: Cognitive adaptations for social exchange. In Barkow et al., op. cit., pp. 163–228.

—— and—— 2001: Unraveling the enigma of human intelligence. In R. Sternberg and J. Kaufman (eds.), *The Evolution of Intelligence*. Mahwah, NJ: Lawrence Erlbaum Associates.

Crain, S. and Pietroski, P. 2001: Nature, nurture and universal grammar. *Linguistics and Philosophy*, 24, 139–86.

Cummins, D. 1996: Evidence for the innateness of deontic reasoning. *Mind and Language*, 11, 160–90.

Currie, G. and Sterelny, K. 1999: How to think about the modularity of mind-reading. *Philosophical Quarterly*, 50, 145–60.

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

—— 1986: *The Blind Watchmaker*. New York: Norton.

Dupré, J. 2001: *Human Nature and the Limits of Science*. Oxford: Oxford University Press.

Dwyer, S. 1999: Moral competence. In K. Murasugi and R. Stainton (eds.), *Philosophy and Linguistics*. Boulder, CO: Westview Press.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996: *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: The MIT Press.

Fiddick, L., Cosmides, L., and Tooby, J. 2000: No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition*, 77, 1–79.

Fodor, J. 1983: *The Modularity of Mind*. Cambridge, MA: The MIT Press.

—— 2000: *The Mind Doesn't Work That Way*. Cambridge, MA: The MIT Press.

Gallistel, R. 1990: *The Organization of Learning*. Cambridge, MA: The MIT Press.

—— 2000: The replacement of general-purpose learning models with adaptively specialized learning modules. In Gazzaniga, op. cit., pp. 1,179–91.

—— and Gibson, J. 2001: Time, rate and conditioning. *Psychological Review*, 108, 289–344.

Gaulin, S. and Robbins, C. 1991: Trivers–Willard effect in contemporary North American society. *American Journal of Physical Anthropology*, 85, 61–9.

Gazzaniga, M. (ed.) 2000: *The New Cognitive Neurosciences*, 2nd edn. Cambridge, MA: The MIT Press.

Gigerenzer, G. and Hug, K. 1992: Domain-specific reasoning: social contracts, cheating and perspective change. *Cognition*, 43, 127–71.

Grant, V. 1998: *Maternal Personality, Evolution and the Sex Ratio*. London: Routledge.

Hermer, L. and Spelke, E. 1994: A geometric process for spatial reorientation in young children. *Nature*, 370, 57–9.

—— and —— 1996: Modularity and development: the case of spatial reorientation. *Cognition*, 61, 195–232.

Hermer-Vazquez, L., Spelke, E., and Katsnelson, A. 1999: Sources of flexibility in human cognition. *Cognitive Psychology*, 39, 3–36.

Horgan, T. and Tienson, J. 1996: *Connectionism and Philosophy of Psychology*. Cambridge, MA: The MIT Press.

Hrdy, S. 1999: *Mother Nature: A History of Mothers, Infants and Natural Selection*. New York: Pantheon.

Job, R. and Surian, L. 1998: A neurocognitive mechanism for folk biology? *Behavioral and Brain Sciences*, 21, 577.

Karmiloff-Smith, A. 1992: *Beyond Modularity*. Cambridge, MA: The MIT Press.

Lakatos, I. 1970: The methodology of scientific research programmes. In I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 91–196.

Laurence, S. and Margolis, E. 2001: The poverty of the stimulus argument. *The British Journal for the Philosophy of Science*, 52, 217–76.

Marcus, G. 2001: *The Algebraic Mind*. Cambridge, MA: The MIT Press.

—— forthcoming: What developmental biology can tell us about innateness. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: Structure and Contents*.

McDermott, D. 2001: *Mind and Mechanism*. Cambridge, MA: The MIT Press.

Núñez, M. and Harris, P. 1998: Psychological and deontic concepts. *Mind and Language*, 13, 153–70.

Phillips, A., Wellman, H., and Spelke, E. 2002: Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, 85, 53–78.

Pinker, S. 1999: *How the Mind Works*. New York: Norton.

Samuels, R. 1998: What brains won't tell us about the mind: a critique of the neurobiological argument against representational nativism. *Mind and Language*, 13, 548–70.

Shallice, T. 1988: *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Spelke, E. 1994: Initial knowledge: six suggestions. *Cognition*, 50, 433–47.

Sperber, D. 1996: *Explaining Culture*. Oxford: Blackwell.

—— 2003: In defense of massive modularity. In I. Dupoux (ed.), *Language, Brain and Cognitive Development. Festschrift for Jacques Mehler*. Cambridge, MA: The MIT Press.

Stone, V., Cosmides, L., Tooby, J., Kroll, N., and Wright, R. 2002: Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Science*, 99, 11,531–6.

Tager-Flusberg, H. (ed.) 1999: *Neurodevelopmental Disorders*. Cambridge, MA: The MIT Press.

Tooby, J. and Cosmides, L. 1992: The psychological foundations of culture. In Barkow et al., op. cit., pp. 19–136.

——and——forthcoming: Motivation and the debate on representational and non-representational innateness. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: Structure and Contents*.

Trivers, R. and Willard, D. 1973: Natural selection of parental ability to vary the sex ratio in offspring. *Science*, 179, 90–2.

Varley, R. 1998: Aphasic language, aphasic thought. In P. Carruthers and J. Boucher (eds.), *Language and Thought*. Cambridge: Cambridge University Press, 128–45.

——2002: Science without grammar: scientific reasoning in severe agrammatic aphasia. In P. Carruthers, S. Stich, and M. Siegal (eds.), *The Cognitive Basis of Science*. Cambridge: Cambridge University Press, 99–116.

Williams, G. 1966: *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Woodward, A. 1998: Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1–34.

CONTEMPORARY DEBATES IN PHILOSOPHY OF SCIENCE

Edited by

Christopher Hitchcock

Blackwell
Publishing

The right of Christopher Hitchcock to be identified as the Author of the Editorial Material in this Work has been asserted in accordance with the UK Copyright, Designs, and Patents Act 1988.