

# The case for massively modular models of mind

**Peter Carruthers**

My charge in this chapter is to set out the positive case supporting massively modular models of the human mind.<sup>1</sup> 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 some of the 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 of 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.

## **1 Introduction: on modularity**

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

---

<sup>1</sup> For the negative case, defending such models against the attacks of opponents, see Carruthers, 2002a, 2002b, 2003, 2004a.

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.

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 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 entirely (or almost entirely) of 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 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 of different strengths, depending upon *how many* components the mind is thought to contain. 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.

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 modular components (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.

Let me comment briefly on the various different elements of this account. 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 can take as input.<sup>2</sup> 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, 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

---

<sup>2</sup> Evolutionary psychologists may well understand domain specificity differently. They tend to understand 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. See Carruthers (forthcoming) for further discussion.

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 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. (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 data-base 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 can only process inputs of a particular sort, concerning a certain kind 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).

## 2 What massive modularity could not be

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.<sup>3</sup> 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 cannot have their own proprietary transducers. Nor can they have shallow outputs. On the contrary, their outputs will be fully-conceptual thoughts or beliefs.

Domain specificity also needs to go, or to be reconceptualized in terms of functional rather than content domains, in the context of a thesis of massive modularity. Although it may well be the case that *many* modules are domain specific, it can’t be the case that *all* are, if the thesis that the mind is built exclusively or almost exclusively out of modules is to be at all believable.<sup>4</sup> Consider practical reasoning, for example. This is plausibly a distinct system of the mind, with a significant innate component, whose internal operations might be encapsulated from and inaccessible to the remainder of cognition (Carruthers, 2004a). And it is a system whose basic architecture is probably very ancient indeed, being common even to insects as well as to ourselves and other mammals (Carruthers, 2004b). But it plainly can’t be domain specific, since in order to do its job it will have to be capable of receiving any belief, and any desire, as input.

Swiftness of processing also needs to go, in the context of massive modularity,

---

<sup>3</sup> 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.

<sup>4</sup> Is this way of proceeding question-begging? Can one insist, on the contrary, that since modules *are* domain specific systems, we can therefore 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 above, there are a *range* of different meanings of ‘module’ available. So principles of charity of interpretation dictate that we should select the meaning that makes the best sense of the claims of massive modularists.

except perhaps in comparison with the speed of *conscious* thought processes, if the latter are realized in cycles of modular activity, as Carruthers (2002a) has maintained. 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. However, it looks like the claim of mandatory operation can be retained. Each component system of the mind can be such that it automatically processes any input that it receives. And certainly it seems that some of the alleged central modules, at least, have such a property. As Segal (1998) points out, we cannot 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 would appear to be mandatory.

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). 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 different 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. 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 can be

constructed by over-learning, say, in such a way that it might be appropriate to describe someone's reading competence as modular.

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). Accordingly, they will be left untouched for the moment, pending closer examination of the arguments in support of massive modularity.

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, domain specificity, comparatively fast processing, and significant innateness or innate channeling will have to be struck out. That leaves us with the idea that modules might be isolable function-specific processing systems, whose operations are mandatory, which are associated with specific neural structures, and whose internal operations may be both encapsulated from the remainder of cognition and inaccessible to it. 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 next two sections of this chapter.

### **3 Arguments for massively modular minds**

In this section I shall consider three of the main arguments that have been 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.

#### *3.1 The argument from biology*

The first argument derives from Simon (1962), and concerns the design of complex functional systems quite generally, and in biology in particular. According to this line of thought, we should expect such systems to be constructed hierarchically out of dissociable

sub-systems, in such a way that the whole assembly could be built up gradually, adding sub-system to sub-system; and in such a way that the functionality of the whole should be buffered, to some extent, from damage to the parts.

Simon (1962) uses the famous analogy of the two watch-makers to illustrate the point. One watch-maker assembles one watch at a time, adding micro-component to micro-component one at a time. 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 watch-maker first builds sets of sub-components out of the given micro-component parts, and then combines those into larger sub-components, until eventually the watches are complete. This helps organize and sequence the whole process, and makes it much less vulnerable to interruption.

Consistent with such an account, there is a very great deal of evidence from across many different levels in biology to the effect that complex functional systems are built up out of assemblies of sub-components. 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 cells, of cellular assemblies, of whole organs, 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. Accordingly, we will now spend some time examining this question.

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 those properties of the human organism that may have 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 different 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 3.2 below.) 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?<sup>5</sup> 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 genotypes. 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

---

<sup>5</sup> 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.

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 mind 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 dice 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 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.<sup>6</sup> 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.

The prediction of this line of reasoning, then, is that cognition will be structured out

---

<sup>6</sup> 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.

of dissociable systems, each of which has a distinctive function, or set of functions, to perform.<sup>7</sup> This gives us a notion of a cognitive ‘module’ that is pretty close to the everyday sense in which one can talk about a hi-fi system as ‘modular’ provided that the tape-deck can be purchased, and can function, 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; and so on and so forth (Sachs, 1985; Shallice, 1988; Tager-Flusberg, 1999; Stone *et al.*, 2002; Varley, 2002).

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. 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); how broad its base is; or 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 get implemented quite rapidly in neural processes. In contrast, only if the pyramid is high with a broad base and convex sides should we expect the mind to decompose into *many* components, each of which in turn consists of many components, and so on.

---

<sup>7</sup> 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.

There is more mileage to be derived from Simon's argument yet, however. For the complexity and range of functions that the overall system needs to execute will surely give us a direct measure of the manner in which the 'pyramid' will slope. (The greater the complexity, the greater the number of sub-systems into which the system will decompose.) This is because the hierarchical organization is there in the first place to ensure 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.<sup>8</sup> (So does learning, since once you have learned one skill, you need to be able to isolate and preserve it while you acquire others. See Manoel *et al.*, 2002.)

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 as evolutionary psychologists have often emphasized, these 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, woo and select 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 one's 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* distinct components, which is a (weak) form of massive modularity thesis.

### 3.2 *The argument from task specificity*

A second line of reasoning supporting massive modularity derives 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). It is one sort of

---

<sup>8</sup> Human software engineers and artificial intelligence researchers have hit upon the same problem, and the same solution, which sometimes goes under the name 'object-oriented programming'. In order that one part of a program can be improved and updated without any danger of introducing errors elsewhere, engineers now routinely modularize their programs. See the discussion towards the end of section 4.

task to learn the sun's azimuth (its height in the sky at any given the 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. 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 general-purpose conditioning is rare at best. Indeed, Gallistel (2000; Gallistel and Gibbon, 2001) has forcefully argued 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. For example, it is well established that *delay* of reinforcement has no effect on rate of acquisition, so long as the intervals between trials are increased by the same proportions. And the number of reinforcements required for acquisition of a new behavior isn't affected by interspersing a significant number of unreinforced trials. These facts are hard to explain if the animals are supposed to be building associations, since the delays and unreinforced trials should surely *weaken* those associations. But they can be predicted if what the animals are doing is estimating relative rates of return. For the rate of reinforcement per stimulus presentation *relative to* the rate of reinforcement in background conditions remains the same, whether or not significant

numbers of stimulus presentations remain unreinforced, for example.

What emerges from these considerations is a picture of the mind as containing a whole host of specialized learning systems (as well as systems charged with generating fitness-enhancing intrinsic desires). And this looks very much like *some* sort of thesis of massive modularity. Admittedly, it doesn't yet follow from the argument that the mind is composed *exclusively* of such systems. But when combined with the previous argument, outlined in section 3.1 above, the stronger conclusion would seem to be warranted.

There really is no reason to believe, however, that each processing system will employ a *unique* processing algorithm. On the contrary, consideration of how evolution generally operates suggests that the same or similar algorithms may be replicated many times over in the human mind / brain. (We could describe this by saying that the same *module-type* is tokened more than once in the human brain, with distinct input and output connections, and hence with a distinct functional role, in each case.) Marcus (2004) explains how evolution often operates by splicing and *copying*, followed by adaptation. First, the genes that result in a given micro-structure (a particular bank of neurons, say, with a given set of processing properties) is copied, yielding two or more instances of such structures. Then second, some of the copies can be adapted to novel tasks. *Sometimes* this will involve tweaking the processing algorithm that is implemented in one or more of the copies. But often it will just involve provision of novel input and/or output connections for the new system.

Samuels (1998) challenges the above line of argument for massive processing modularity, however, claiming that instead of a whole suite of specialized learning systems, there might be just a single general-learning / general-inferencing mechanism, but one operating on lots of organized bodies of innate information. (He calls this 'informational modularity', contrasting it with the more familiar form of *computational* modularity.) However, this would surely create a serious processing bottleneck. If there were really just one (or even a few) inferential systems – generating beliefs about the likely movements of the surrounding mechanical objects; about the likely beliefs, goals, and actions of the surrounding agents; about who owes what to whom in a social exchange; and so on and so forth – then it looks like there would be a kind of tractability problem here. It would be the problem of forming novel beliefs on all these different subject matters in real

time (in seconds or fractions of a second), using a limited set of inferential resources. Indeed (and in contrast with Samuel's suggestion) surely *everyone* now thinks that the mind / brain is massively parallel in its organization. In which case we should expect there to be distinct systems that can process each of the different kinds of information at the same time.

Samuels might try claiming that there could be a whole suite of distinct domain-general processing systems, all running the same general-learning / general-inferencing algorithms, but each of which is attached to, and draws upon the resources of, a distinct domain-specific body of innate information. This would get him the computational advantages of *parallel* processing, but without commitment (allegedly) to any *modular* processing. But actually this is just a variant on the massive-computational-module hypothesis. For there is nothing in the nature of modularity per se that requires modules to be running algorithms that are distinct from those being run by other modules, as we have just seen. What matters is just that they should be isolable systems, performing some specific function, and that their internal operations should be computationally feasible (as we will see in section 3.3 below). So one way in which massive modularity could be realized is by having a whole suite of processors, each of which performs some specific function within the overall architecture of the mind, and each of which draws on its own distinctive body of innate information relevant to that function, but where the algorithms being computed by those processors are shared ones, replicated many times over in the various different processing systems.

Although possible in principle, however, this isn't a very *likely* form of massive modularity hypothesis. For it does come with severe computational costs. This is because the difference between this 'informational module' hypothesis and the classic 'computational module' hypothesis just concerns whether or not the innate information is explicitly represented. The classical idea is that there will be, within the mind-reading faculty for example, an algorithm that takes the system straight from, 'x is seeing that P' to 'probably x believes that P' (say). The information that people believe what they see is implicitly represented in the algorithm itself. Samuel's view, in contrast, will be that there is an intermediate step. Domain-general inference mechanisms will draw on the explicitly represented belief that people believe what they see, in order to mediate the inference from

premise to conclusion. Imagine this multiplied again and again for all the different sorts of inferential transition that people regularly make in the domain of theory of mind, and it is plain that his proposal would come with serious computational costs. And it is equally clear that even if informational modules were the initial state of the ancestral mind / brain, over evolutionary time informational modules would be replaced by computational ones.

Combining the arguments of sections 3.1 and 3.2, then, we can predict that the mind should be composed entirely or almost entirely of modular components (in the everyday sense of ‘module’), many of which will be innate or innately channeled. All of these component systems should run task-specific processing algorithms, with distinct input and/or output connections to other systems, although some of them may replicate some of the same algorithm types in the service of distinct tasks. This looks like a thesis of massive modularity worth the name, even if there is nothing here yet to warrant the claims that the internal processing of the modules in question should be either encapsulated, on the one hand, or inaccessible, on the other.

### 3.3 *The argument from computational tractability*

Perhaps the best-known of the arguments for massive modularity, however – at least amongst philosophers – is the argument from computational tractability, which derives from Fodor (1983, 2000).<sup>9</sup> And it is generally thought that this argument, if it were successful, would license the claim that the mind is composed of *encapsulated* processing systems, thus supporting a far stronger form of massive modularity hypothesis than has been defended in this chapter so far (Carruthers, 2002a; Sperber, 2002).

The first premise of the argument is the claim that the mind is realized in processes that are computational in character. This claim is by no means uncontroversial, of course, although it is the guiding methodological assumption of much of cognitive science. Indeed, it is a claim that is denied by certain species of distributed connectionism. But in recent

---

<sup>9</sup> Fodor himself doesn’t argue for *massive* modularity, of course. Rather, since he claims that we know that central processes of belief fixation and decision making *can’t* be modular, he transforms what would otherwise be an argument for massive modularity into an argument for pessimism about the prospects for computational psychology. See Carruthers (2002a, 2002b, 2003, 2004a) for arguments that the knowledge-claim underlying such pessimism isn’t warranted.

years arguments have emerged against these competitors that are decisive, in my view (Gallistel, 2000; Marcus, 2001). And what remains is that computational psychology represents easily our best – and perhaps our only – hope for fully understanding how mental processes can be realized in physical ones (Rey, 1997). In any case, I propose just to *assume* the truth of this first premise for the purposes of the discussion that follows.

The second premise of the argument is the claim that if cognitive processes are to be realized computationally, then those computations must be *tractable* ones. What does this amount to? First of all, it means that the computations must be such that they can *in principle* be carried out within finite time. But it isn't enough that the computations postulated to take place in the human brain should be tractable in principle, of course. It must also be feasible that those computations could be executed (perhaps in parallel) in a system with the properties of the human brain, within time-scales characteristic of actual human performance. By this criterion, it seems likely that many computations that aren't strictly speaking intractable from the perspective of computer science, should nevertheless count as such for the purposes of cognitive science.

There is a whole branch of computer science devoted to the study of more-or-less intractable problems, known as 'Complexity Theory'. And one doesn't have to dig very deep into the issues to discover results that have important implications for cognitive science. For example, it has traditionally been assumed by philosophers that any candidate new belief should be checked for consistency with existing beliefs before being accepted. But in fact consistency-checking is demonstrably intractable, if attempted on an exhaustive basis. Consider how one might check the consistency of a set of beliefs via a truth-table. Even if each line could be checked in the time that it takes a photon of light to travel the diameter of a proton, then even after 20 billion years the truth-table for a set of just 138 beliefs ( $2^{138}$  lines) still wouldn't have been completed (Cherniak, 1986).

From the first two premises together, then, we can conclude that the human mind must be realized in a set of computational processes that are suitably tractable. This means that those processes will have to be *frugal*, both in the amount of information that they require for their normal operations, and in the complexity of the algorithms that they deploy when processing that information.

The third premise of the argument then claims that in order to be tractable,

computations need to be encapsulated; for only encapsulated processes can be appropriately frugal in the informational and computational resources that they require. As Fodor (2000) explains it, the constraint here can be expressed as one of *locality*. Computationally tractable processes have to be *local*, in the sense of only consulting a limited data-base of information relevant to those computations, and ignoring all other information held in the mind. For if they attempted to consult all (or even a significant subset) of the total information available, they would be subject to combinatorial explosion, and hence would fail to be tractable after all.

This third premise, in conjunction with the other two, would then (if it were acceptable) license the conclusion that the mind must be realized in a set of encapsulated computational processes. And when combined with the conclusions of the arguments of sections 3.1 and 3.2 above, this would give us the claim that the mind consists in a set of encapsulated computational systems whose operations are mandatory, each of which has its own function to perform, and many of which execute processing algorithms that aren't to be found elsewhere in the mind (although some re-use algorithms that are also found in other systems for novel functions). It is therefore crucial for our purposes to know whether the third premise is really warranted; and if not, what one might put in its stead. This will form the topic of the next section.

#### **4 What does computational frugality really require?**

I have claimed that the first two premises in the Fodorian argument sketched in section 3.3 are acceptable. So we should believe that cognition must be organized into systems of computational processes that are appropriately *frugal*. The question is whether frugality requires encapsulation, in the way that is stated by the third premise of the argument. The idea has an obvious appeal. It is certainly true that *one* way to ensure the frugality of a set of computational systems, at least, would be to organize them into a network of encapsulated processors, each of which can look only at a limited data-base of information in executing its tasks. And it may well be the case that evolution has settled on this strategy in connection with many of the systems that constitute the human mind. It is doubtful, however, whether this is the *only* way of ensuring frugality.

The assumption of encapsulation (at least, as it is normally understood – see below)

may derive from an older tradition in cognitive science and AI, in which information search had to be *exhaustive*, and in which algorithms were designed to be optimally reliable. But this tradition is now widely rejected. Most cognitive scientists now think that the processing rules deployed in the human mind have been designed to be *good enough*, not to be optimal. Given that speed of processing is always one constraint for organisms that may need to think and act swiftly in order to survive, evolution will have led to compromises on the question of reliability. Indeed, it will favor a *satisficing* strategy, rather than an optimal one. And likewise on information search: evolution will favor a variety of search heuristics that are good enough without being exhaustive.

These points are well illustrated by the research program pursued in recent years by Gigerenzer and colleagues (e.g. Gigerenzer *et al.*, 2000). They have investigated the comparative reliability and frugality of a variety of rules for use in information search and decision making, with startling results. It turns out that even very simple heuristics can be remarkably successful – such as choosing the only one of two options that you recognize, when asked which of two cities is larger, or when asked to predict which of two companies will do best in the stock market. In some cases these simple heuristics will even outperform much fancier and information-hungry algorithms, such as multiple regression. And a variety of simple heuristics for searching for information within a wider data-base, combined with stopping-rules if the search is unsuccessful within a specified time-frame, can also work remarkably well – such as accessing the information in the order in which it was last used, or accessing the information that is partially activated (and hence made salient) by the context.<sup>10</sup>

For a different sort of example, consider the simple practical reasoning system sketched in Carruthers (2002a). It takes as initial input whatever is currently the strongest desire, for *P*.<sup>11</sup> It then queries the various belief-generating modules, while also conducting

---

<sup>10</sup> See Carruthers (forthcoming) for an extended discussion of the relationship between the massive modularity hypothesis and the simple heuristics movement, and for elaboration and defense of a number of the points made in the present section.

<sup>11</sup> Note that *competition for resources* is another of the heuristics that may be widely used within our cognitive systems; see Sperber, 2005. In the present instance one might think of all activated desires as competing with one another for entry into the practical reasoning system.

a targeted search of long term memory, looking for beliefs of the form  $Q \supset P$ . If it receives one as input, or if it finds one from its own search of memory, it consults a data-base of action schemata, to see if  $Q$  is something doable here and now. If it is, it goes ahead and does it. If it isn't, it initiates a further search for beliefs of the form  $R \supset Q$ , and so on. If it has gone more than  $n$  conditionals deep without success, or if it has searched for the right sort of conditional belief without finding one for more than some specified time  $t$ , then it stops and moves on to the next strongest desire.

Such a system would be frugal, both in the information that it uses, and in the complexity of its algorithms. But does it count as encapsulated? This isn't encapsulation as that notion would generally be understood, which requires there to be a limited module-specific data-base that gets consulted by the computational process in question. For here, on the contrary, the practical reasoning system can search within the total set of the organism's beliefs, using structure-sensitive search rules. But for all that, there is *a* sense in which the system is encapsulated that is worth noticing.

Put as neutrally as possible, we can say that the idea of an encapsulated system is the notion of a system whose internal operations *can't* be affected by *most or all* of the information held elsewhere in the mind. But there is a scope ambiguity here.<sup>12</sup> We can have the modal operator take narrow scope with respect to the quantifier, or we can have it take wide scope. In its narrow-scope form, an encapsulated system would be this: concerning most of the information held in the mind, the system in question *can't* be affected by *that* information in the course of its processing. Call this 'narrow-scope encapsulation'. In its wide-scope form, on the other hand, an encapsulated system would be this: the system is such that it *can't* be affected by *most* of the information held in the mind in the course of its processing. Call this 'wide-scope encapsulation'.

Narrow-scope encapsulation is the one that is taken for granted in the philosophical literature on modularity. We tend to think of encapsulation as requiring some determinate (and large) body of information, such that *that* information can't penetrate the module.

---

<sup>12</sup> Modal terms like 'can' and 'can't' have wide scope if they govern the whole sentence in which they occur; they have narrow scope if they govern only a part. Compare: 'I can't kill everyone' (wide scope; equivalent to, 'It is impossible that I kill everyone') with, 'Everyone is such that I can't kill them' (narrow scope). The latter is equivalent to, 'I can't kill anyone'.

However, it can be true that the operations of a module can't be affected by most of the information in a mind, without there being some determinate sub-division between the information that can affect the system and the information that can't. For as we have just seen, it can be the case that the system's algorithms are so set up that only a limited amount of information is ever consulted before the task is completed or aborted. Put it this way: a module can be a system that *must* only consider a small sub-set of the information available. Whether it does this via encapsulation as traditionally understood (the narrow-scope variety), or via frugal search heuristics and stopping rules (wide-scope encapsulation), is inessential. The important thing is that the system should be *frugal*, both in the information that it uses and in the resources that it requires for processing that information.

The argument from computational tractability, then, does warrant the claim that the mind should be constructed entirely out of systems that are *frugal*; but it doesn't warrant a claim of encapsulation, as traditionally understood (the narrow-scope variety). It does, however, warrant a non-standard encapsulation claim (the wide-scope version). In addition, it supports the claim that the processing systems in question should have internal operations that are *inaccessible* elsewhere. Or so I shall now briefly argue by *reductio*, and by induction across current practices in AI.

Consider what it would be like if the internal operations of each system were accessible to all other systems. (This would be *complete* accessibility. Of course the notions of *accessibility* / *inaccessibility*, just like the notions of *encapsulation* / *lack of encapsulation*, admit of degrees.) In order to make use of that information, those other systems would need to contain a model of those operations, or they would need to be capable of simulating or replicating them. In order to use the information that a given processing system is currently undertaking such-and-such computations, the other systems would need to contain a representation of the algorithms in question. This would defeat the purpose of dividing up processing into distinct sub-systems running different algorithms for different purposes, and would likely result in some sort of combinatorial explosion. At the very least, we should expect that *most* of those processing systems should have internal operations that are inaccessible to all others; and that *all* of the processing systems that

make up the mind should have internal operations that are inaccessible to *most* others.<sup>13</sup>

Such a conclusion is also supported inductively by current practices in AI, where researchers routinely assume that processing needs to be divided up amongst distinct systems running algorithms specialized for the particular tasks in question. These systems can talk to one another and query one another, but not access one another's internal operations. And yet they may be conducting guided searches over the same memory database. (Personal communication: Mike Anderson, John Harty, Aaron Sloman.) That researchers attempting to build working cognitive systems have converged on some such architecture is evidence of its inevitability, and hence evidence that the human mind will be similarly organized.

This last point is worth emphasizing further, since it suggests a distinct line of argument supporting the thesis of massive modularity in the sense that we are currently considering. Researchers charged with trying to build intelligent systems 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, and provide input to each other, and many of which can access shared data-bases. But many of these systems will deploy processing algorithms that aren't shared by the others. And most of them won't know or care about what is going on within the others.

Indeed, the convergence here is actually wider still, embracing computer science more generally and not just AI. 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'. Many programming languages now enable 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

---

<sup>13</sup> One important exception to this generalization is as follows. We should expect that many modules will be composed out of other modules as parts. Some of these component parts may feed their outputs directly to other systems. (Hence such components might be shared between two or more larger modules.) Or it might be the case that they can be queried independently by other systems. These would then be instances where some of the intermediate *stages* in the processing of the larger module would be available elsewhere, without the intermediate *processing* itself being so available.

elsewhere, with all the attendant risks that this would bring. And the resulting architecture is regarded as well nigh inevitable once a certain threshold in the overall degree of complexity of the system gets passed. (Note the parallel here with Simon's argument from complexity, discussed in section 3.1 above.)

## 5 Conclusion

What emerges, then, is that there is a strong case for saying that the mind is very likely to consist of a great many different processing systems, which exist and operate to some degree independently of one another. Each of these systems will have a distinctive function or set of functions; each will have a distinct neural realization; and many will be significantly innate, or genetically channeled. Many of them will deploy processing algorithms that are unique to them. And all of these systems will need to be *frugal* in their operations, hence being encapsulated in either the narrow-scope or the wide-scope sense. Moreover, the processing that takes place within each of these systems will generally be inaccessible elsewhere.<sup>14</sup> Only the results, or outputs, of that processing will be made available for use by other systems.

Does such a thesis deserve the title of 'massive *modularity*'? It is certainly a form of massive modularity in the everyday sense that we distinguished at the outset. And it retains many of the important features of Fodor-modularity. Moreover, it does seem that this is the notion of 'module' that is used pretty commonly in AI, if not so much in philosophy or psychology (McDermott, 2001). But however it is described, we have here a substantive and controversial claim about the basic architecture of the human mind; and it is one that is supported by powerful arguments.

In any complete defense of massively modular models of mind, so conceived, we would of course have to consider all the various arguments *against* such models, particularly those deriving from the holistic and creative character of much of human thinking. This is a task that I cannot undertake here, but that I have attempted elsewhere (Carruthers, 2002a, 2002b, 2003, 2004a). If those attempted rebuttals should prove to be successful, then we can conclude that the human mind will, indeed, be massively modular

---

<sup>14</sup> As we already noted above, the notions of 'encapsulation' and 'inaccessibility' admit of degrees. The processing within a given system may be *more* or *less* encapsulated from and inaccessible to other systems.

(in one good sense of the term ‘module’).<sup>15</sup>

## References

- Carruthers, P. (2002a) The cognitive functions of language. & Author’s response: Modularity, language and the flexibility of thought. *Behavioral and Brain Sciences*, 25:6, 657-719.
- Carruthers, P. (2002b) Human creativity: its evolution, its cognitive basis, and its connections with childhood pretence. *British Journal for the Philosophy of Science*, 53, 1-25.
- Carruthers, P. (2003) On Fodor’s Problem. *Mind and Language*, 18, 502-523.
- Carruthers, P. (2004a) Practical reasoning in a modular mind. *Mind and Language*, 19, 259-278.
- Carruthers, P. (2004b) On being simple minded. *American Philosophical Quarterly*, 41, 205-220.
- Carruthers, P. (forthcoming) Simple heuristics meet massive modularity. In P.Carruthers, S.Laurence and S.Stich (eds.), *The Innate Mind: foundations and the future*. Oxford University Press.
- Cherniak, C. (1986) *Minimal Rationality*. MIT Press.
- Dawkins, R. (1986) *The Blind Watchmaker*. Norton.
- Fodor, J. (1983) *The Modularity of Mind*. MIT Press.
- Fodor, J. (2000) *The Mind doesn’t Work that Way*. MIT Press.
- Gallistel, R. (1990) *The Organization of Learning*. MIT Press.
- Gallistel, R. (2000) The replacement of general-purpose learning models with adaptively specialized learning modules. In M.Gazzaniga (ed.), *The New Cognitive Neurosciences* (second edition), MIT Press.
- Gallistel, R. and Gibson, J. 2001. Time, rate and conditioning. *Psychological Review*, 108:

---

<sup>15</sup> Thanks to Mike Anderson, Clark Barrett, John Harty, Edouard Machery, Richard Samuels, Aaron Sloman, Robert Stainton, Stephen Stich, and Peter Todd for discussion and/or critical comments that helped me to get clearer about the topics covered by this chapter. Stich and Samuels, in particular, induced at least one substantial change of mind from my previously published views, in which I had defended the idea that modules must be encapsulated (as traditionally understood). See Carruthers, 2002a, 2003.

289-344.

- Gardner, H. 1983. *Frames of Mind: the theory of multiple intelligences*. Heinemann.
- Gigerenzer, G., Todd, P., and the ABC Research Group. (2000) *Simple Heuristics that Make Us Smart*. Oxford University Press.
- Karmiloff-Smith, A. (1992). *Beyond Modularity*. MIT Press.
- Manoel E., Basso L., Correa U., and Tani G. (2002) Modularity and hierarchical organization of action programs in human acquisition of graphic skills. *Neuroscience Letters*, 335(2), 83-6.
- Marcus, G. (2001) *The Algebraic Mind*. MIT Press.
- Marcus, G. (2004) *The Birth of the Mind: how a tiny number of genes creates the complexities of human thought*. Basic Books.
- Marr, D. (1983) *Vision*. Walter Freeman.
- McDermott, D. (2001) *Mind and Mechanism*. MIT Press.
- Pinker, S. (1997) *How the Mind Works*. Penguin Press.
- Rey, G. (1997) *Contemporary Philosophy of Mind*. Blackwell.
- Sachs, O. (1985) *The Man who Mistook his Wife for a Hat*. Picador.
- Samuels, R. (1998) Evolutionary psychology and the massive modularity hypothesis. *British Journal for the Philosophy of Science*, 49, 575-602.
- Seeley, T. (1995) *The Wisdom of the Hive: the social physiology of honey bee colonies*. Harvard University Press
- Segal, G. (1998) Representing representations. In P. Carruthers and J. Boucher (eds.), *Language and Thought*, Cambridge University Press.
- Shallice, T. (1988) *From Neuropsychology to Mental Structure*. Cambridge University Press.
- Simon, H. (1962) The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467-482.
- Sperber, D. (1996) *Explaining Culture: a naturalistic approach*. Blackwell.
- Sperber, D. (2002) In defense of massive modularity. In I. Dupoux (ed.), *Language, Brain and Cognitive Development*. MIT Press.
- Sperber, D. (2005) Massive modularity and the first principle of relevance. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: structure and*

*contents*, Oxford University 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, 11531-11536.

Tager-Flusberg, H. (ed.) (1999) *Neurodevelopmental Disorders*. MIT Press.

Tooby, J. and Cosmides, L. (1992) The psychological foundations of culture. In J. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind*, Oxford University Press.

Varley, R. (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 University Press.