Simple heuristics meet massive modularity

Peter Carruthers

This chapter investigates the extent to which claims of massive modular organization of the mind (espoused by some members of the evolutionary psychology research program) are consistent with the main elements of the simple heuristics research program. A number of potential sources of conflict between the two programs are investigated and defused. However, the simple heuristics program turns out to undermine one of the main arguments offered in support of massive modularity, at least as the latter is generally understood by philosophers. So one result of the argument will be to force us to re-examine the way in which the notion of *modularity* in cognitive science should best be characterized, if the thesis of massive modularity isn't to be abandoned altogether. What is at stake in this discussion, is whether there is a well-motivated notion of 'module' such that we have good reason to think that the human mind must be massively modular in its organization. I shall be arguing (in the end) that there is.

1 Introduction

Two distinct research programs in cognitive science have developed and burgeoned over the last couple of decades, each of which is broadly evolutionary or adaptationist in orientation, and each of which is nativist in the claims that it makes about the architecture and much of the contents of the human mind. One is the evolutionary psychology program and its associated claim of massive mental modularity (Gallistel, 1990, 2000; Tooby and Cosmides, 1992; Sperber, 1996; Pinker, 1997). The other is the simple heuristics movement and its associated claim of an 'adaptive toolbox' of cognitive procedures (Gigerenzer *et al.*, 1999). Each is, in addition, committed to explaining the variability of culture in terms of the flexible application of modules / heuristics in local conditions.

My question is this: are these *competing* research programs, or do they complement one another? The proponents of each these programs don't often mention the existence of the other. Yet both are in the business of constructing explanations that are plausible, not only in evolutionary terms, but in relation to data from comparative psychology. And it would seem that both are in the business of explaining (or trying to explain) how cognition can be realized in processes that are computationally tractable. However, there is some reason to think that these programs offer explanations of human cognition that are inconsistent with one another, or that otherwise undermine each other, as we shall see.

I shall begin by briefly elaborating and elucidating the twin theses that cognition is massively modular, and that it is constructed out of an adaptive toolbox of simple heuristics. I shall then turn to the question of consistency, arguing (in the end) that the two research programs should best be seen as natural bedfellows and mutual supporters, rather than as competitors. But in order for this to be the case, the thesis of massive mental modularity will have to undergo a (well-motivated) transformation.

2 Massive modularity

Modularists claim that evolutionary thinking about the mind delivers specific predictions about the latter's architecture, the most important of which is that the mind is massively modular in its organization. This conclusion can be reached via a number of distinct (but mutually supporting) lines of reasoning. I shall sketch two of them here. (For more extensive discussion, see Carruthers, 2005.)

2.1 The argument from biology

The first argument derives from Simon (1962), and concerns the evolution 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 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 a set of sub-components out of the given micro-component parts, and then combines those into larger sub-components, until eventually the watch is 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 which is constructed out of further sub-components and has a distinctive role to play in the functioning of the whole, and many of which 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.

The prediction of this line of reasoning, then, is that cognition will be structured out of dissociable systems, each of which has a distinctive function, or set of functions, to perform. (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.) 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).

2.2 The argument from computational tractability

The second line of reasoning supporting massive modularity begins from the assumption that mental processes must be realized *computationally* in the brain.[1] And the argument, deriving

[1] This assumption is common to all of classical cognitive science. It has been challenged more recently by proponents of distributed connectionism. But whatever the successes of connectionist networks in respect of pattern recognition, there are principled reasons for thinking that such networks are incapable of the kind of one-

from Fodor (1983, 2000), is that computational processes need to be encapsulated if they are to be tractable.

The argument goes something like this.[2] If a processing system can look at any arbitrary item of information in the course of its processing, then the algorithms on which that system runs will have to be arbitrarily complex also. For those algorithms will have to specify, in respect of each item of information that it could access, what step should be taken next – presumably different for each such item of information, if the system is to be a context-sensitive one. So, the more items of information a program can look at while processing, the more complex its algorithms will need to be. So conversely, if a system's algorithms are to be computationally tractable, limits will need to be placed on the set of items of information that it can look at.

Consistent with Fodor's argument, what more than a quarter century of research in artificial intelligence appears to have taught us is that computational processes need to be divided up amongst a suite of modular sub-systems if they are to be tractable (Bryson, 2000; McDermott, 2001). Note that this line of argument doesn't start from commitment to some particular agent architecture (e.g. Brooks' 1986 subsumption architecture) and say, 'Hey, this system is modular; therefore cognition is modular'. Rather, the argument is a meta-induction across recent trends in AI. The claim is something like this: over the last half-dozen years virtually everyone in AI has converged on modular architectures of one sort or another. This has been forced on them by the experience of trying to design systems that actually work. So this gives us good reason to think that any real agent must have a modular internal organization.

Now it may well be, as we shall see, that the notion of 'module' at work in AI isn't quite the same as Fodor's. But the main premise in the above meta-induction can be challenged more directly. For there are some agent-architectures on the market that are avowedly a-modular in character, such as Newell's (1990) SOAR architecture. However, it can be claimed that there is a confusion over terminology lying in back of these avowals. Although everyone can agree that a system is a module only if it is domain-specific (and hence that an architecture is a-modular if the systems within it *aren't* domain-specific), different research traditions mean different things

shot learning and updating of variables of which humans and other animals are manifestly capable. See Gallistel, 2000; Marcus, 2001.

[2] I don't mean to endorse this argument exactly as stated. Some of its assumptions will get unpacked and challenged, and the argument will get rebuilt, as our discussion proceeds.

by a 'domain'. So when someone coming from one tradition *says* that their architecture is an a-modular one, it might actually be modular in the sense deployed within the other tradition. Let me elaborate.

Developmental and cognitive psychologists tend to think of domains in terms of kinds of content, or kinds of subject matter. When they say that development is a domain-specific process, what they mean is that it proceeds at a different pace and follows a different trajectory in the different areas of cognitive competence that adult humans display (folk-psychology, folk-physics, folk-biology, and so on). Call this a 'content-domain'. Evolutionary psychologists and/or massive modularity theorists, in contrast, think of domains as characterized by a *function*. In this latter sense, the domain of a module is what it is *supposed to do* within the overall architecture of the cognitive system. Call this a 'task-domain'. The confusion arises quite naturally, and may easily pass unnoticed, because *many* of the task-specific modules postulated by evolutionary psychology are *also* content-specific in nature. (The folk-psychology module is targeted on mental states; the folk-physics module is about physical movements of inanimate objects; the cheater-detection module is about costs and benefits in exchange; and so on.) But there is nothing in the notion of a module *per se* to require this, from an evolutionary-psychology perspective.

When someone coming out of the cognitive psychology tradition says, 'I've built a system that is a-modular in its architecture', what they probably mean is, 'I've built a system that doesn't operate on any specific type of subject matter'. And it is true that Newell's SOAR, for example, which is designed for decision-making, can acquire the ability to make decisions concerning many different kinds of subject-matter. But it may still be a modular system from the perspective of the evolutionary psychologist (it may consist of isolable systems with specific functions whose internal operations are encapsulated). You have to actually look and see. And when you do look at SOAR, it does *seem* to be modularly organized (despite advertising itself as a-modular). For different goals and sub-goals come with 'frames' of relevant information attached. When figuring out what to do in pursuit of a goal, the program is only allowed to look at what is in the relevant frame. So its operations would seem to be encapsulated.[3]

It should be noted that the information contained in a given 'frame' can change with time, however. This requires us to distinguish between *weakly modal* and *strongly modal* construals of

[3] Quite what *sense* of 'encapsulated' is involved here will loom large in later sections, especially in section 8.

encapsulation. In the strong sense, to say that a given system is encapsulated from all but the information in its proprietary data-base is to say that it cannot access any other information at *any* time during its existence, no matter what else happens. This is one sense in which SOAR's 'frames' aren't encapsulated, since they can and do alter with time. In a weaker sense, however, we can say that a system is encapsulated provided it can only access whatever information is contained it its proprietary data-base at *that* time.

There seems no good reason to insist on the strongly modal construal of modularity. For the weaker construal gives us all that we *need* from modularity, which is that the system's operations should be computationally tractable. And think of the language-faculty, for example, which is considered by Fodor (1983) to be one of the archetypal modules. The processing data-base for this system surely isn't frozen for all time. New irregular verbs can always be learned, for instance, and these would surely then be counted as belonging to the system's processing data-base.

Putting together the above two lines of reasoning, then (from biology and from computational tractability), what we get is the prediction that the human mind should consist of a whole host of functional and multi-functional systems and sub-systems, which are to some degree dissociable from one another, and whose internal computational processes are encapsulated from most of the information held elsewhere in the mind at that time. This is the thesis of massive mental modularity, as generally conceived.[4]

3 Simple heuristics

Where evolutionary psychology starts from reflection upon biological systems generally, and proposes a research program for uncovering the elements of a modular mind, the initial focus of the simple heuristics movement is more limited. It starts from consideration of the rationality or irrationality of human reasoning.

Psychologists have been claiming since the 1970s that humans are extremely *bad* at many

[4] There are, of course, many objections to the thesis of massive modularity, too. Most of them have to do with the apparent *holism* of human central cognitive processes of inference and belief-formation (Fodor, 1983, 2000), and with the distinctive flexibility and creativity of the human mind. These facts make it hard to see how the mind can consist entirely (or even largely) of modules. This is not the place to pursue and reply to these difficulties. See Carruthers, 2002a, 2002b, 2002c, 2003, 2004.

kinds of reasoning. For example, numerous studies involving the Wason conditional selection task suggest that people are quite poor at discerning under what circumstances a simple conditional statement would be true or false (Wason and Evans, 1975; Evans and Over, 1996). And human reasoners commit frequent fallacies, especially when reasoning about probability, where they commit the conjunction fallacy, base-rate neglect, and so on (Kahneman and Tversky, 1982). But it is evident that for us to move beyond these factual claims to make a judgment about human irrationality may well require us to make some assumptions about the *nature* of rationality.

In fact, the question, ‘What is rationality?’ is the *same* as the question, ‘How *should* we reason?’ Philosophers and psychologists alike have traditionally assumed that we should reason *validly*, where possible; and in *reliable* ways more generally (e.g. in domains of non-demonstrative inference, such as science). But in fact truth cannot be our only goal. We also need *enough* truths in a short enough time-frame to enable decision-making and action. Moreover, reasoning and decision-making have to be realized in computationally *tractable* processes, if they are to be computationally realized at all.

For example, it has traditionally been assumed 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).

There is a whole field of computer science devoted to the study of such problems, called ‘complexity theory’. But it is important to realize that computational intractability, for the purposes of *cognitive* science, can include problems that wouldn’t be characterized as intractable by computer scientists. This is because our goal is to explain processes that happen in real time (in seconds or minutes rather than millennia), and because we need to operate with assumptions about the speed of processing of brains (*significantly* slower than modern computers, albeit with much of that processing being conducted in parallel), as well as making assumptions about memory power. In effect this means that the idea of computational *intractability*, for the purposes of cognitive science, doesn’t admit of a formal definition. But that is just as it should be, since we are dealing here with an *empirical* science.

This line of thinking leads to the idea of *naturalized rationality*. We need reasoning processes that are reliable *enough*, but also *fast* enough and *frugal* enough (in terms of the temporal and computational resources required) given the demands of a normal human life. And of course, what counts as 'fast' or 'frugal' isn't something that can be specified by philosophers *a priori*. Rather, these things will depend upon the properties of the computational systems employed by mammalian brains generally, and by the human brain in particular; and on the task demands that our ancestors regularly faced.

This is the background against which the simple heuristics research program has been developed. The goal is to find computational procedures that are fast and frugal, but that are reliable enough in a given environment for them to be worthwhile having. For example, one heuristic explored by Gigerenzer *et al*. (1999) is *recognition* – if asked which of two German cities is the larger, the heuristic tells you to select the one that you recognize, if you recognize only one. This heuristic proves remarkably successful, even when pitted against much fancier (and much more computationally and informationally demanding) choice procedures like Bayes' rule, multiple regression, etc.; and it proves successful across a wide range of decision types (including the selection of companies that are most likely to do well in the stock market).

Note that there is one important point of similarity between the simple heuristics movement and the evolutionary psychology program, then. This is that each lays a similar emphasis on computational tractability amongst cognitive mechanisms. But each then appears to follow a different strategy in pursuit of computationally tractable processes. One postulates a set of simple heuristics; the other postulates a set of encapsulated modules. These seem like distinct – perhaps inconsistent – approaches to the same problem. We will pursue these issues in the sections that follow.

4 An inconsistent pair?

Are the massive modularity and simple heuristics research programs inconsistent, then? At the very least, it would seem that each can incorporate elements from the other without inconsistency, and perhaps to their mutual benefit. Thus a massive modularist might believe that some of the processes that occur internally within a module are heuristic in character. For example, rather than searching exhaustively through all the information in its proprietary data-base, a module might adopt the satisficing heuristic of stopping search when it has found an item of information that is *good enough* for use in its current task. Likewise, a modularist might accept that simple heuristics play a role in orchestrating the *interactions amongst* modules and their influence upon behavior. Similarly, believers in simple heuristics could surely accept that at least some of the processes that issue in belief or that lead to a decision are modular in character.

Moreover, massive modularity theorists emphasize that part of the point of a modular architecture is that different modules can be structured in such a way as to embody information about the content-domains that they concern and/or can deploy algorithms that are tailored to a specific set of task demands. A similar idea seems to be at work within the simple heuristics framework, in the notion of *ecological rationality*. The idea is that, in connection with any given heuristic, there will be a range of different environments and environment types within which that heuristic will operate with a significant degree of reliability. And we can think of the heuristic as having been selected (by evolution, by individual learning, or through the success of a particular culture) to operate in those environments, thereby (in a sense) embodying information about them.

Nevertheless, an impression of inconsistency between the two research programs might remain. For it might appear that they offer competing models of the overall innate architecture of the mind. Much of what goes on within the simple heuristics program is an attempt to model various aspects of *decision-making*; and many people assume that the decision-making system has to be an a-modular one. (It certainly can’t be domain specific in the content-specific sense of ‘domain’, anyway.) Moreover, many of the heuristics discussed by those working within the simple heuristics program would seem to apply in widely diverse and evolutionarily distinct domains; and some of them might be learned, too. By contrast, the hypothesis of massive modularity is generally thought to suppose that the mind consists of a set of evolved modular systems, targeted on domains of special evolutionary significance.

Practical reasoning can actually be thought of as modular, however, in the relevant sense of ‘module’. For recall that modularity is about encapsulation, and not necessarily about domain- (in the sense of content-) specificity. A practical reasoning module would be a system that could take any belief or desire as input, but which was nevertheless encapsulated in respect of its processing of that input. As sketched in Carruthers (2002a), such a system might be set up (in animals if not in us) to take whatever is the currently strongest desire, for *P*, as initial input, and then to query other modular belief-generating systems and/or initiate a memory search for beliefs of the form, $Q \supset P$. When a conditional of this form is received as input, it checks Q against a data-base of action-schemata to see if it is something doable; of so, it goes ahead and does it; if not, it initiates a further search for beliefs of the form, $R \supset Q$. And so on. Perhaps it also has a simple stopping rule: if you have to go more than *n* conditionals deep, or if more than time *t* has elapsed without success, stop and move on to the next strongest desire. This looks like it would be an encapsulated system, alright; but not a content-specific one.

One can easily imagine, then, that the operations of such a system might be supplemented by a set of heuristics, such as: if you want something, first approach it. Or for another example, much of the literature on navigation suggests that children and other animals operate with a nested set of heuristics when disoriented (Shusterman and Spelke, 2005). The sequence appears to be something like this: if you don't know where you are, seek a directional beacon (e.g. a distant landmark, or the position of the sun); if there is no directional beacon, seek to match the geometric properties of the environment; if geometric information is of no help, seek a local landmark. Likewise, it is plausible that the practical reasoning system might employ a variety of heuristics for ending search (e.g. when deciding whom to marry; Gigerenzer *et al*., 1999). And so on. None of this seems inconsistent with the claim that the practical reasoning system is modular.

As for the fact that heuristics seem to apply across what are evolutionary distinct domains, recall the metaphor of an 'adaptive toolbox', which appears central to the way of thinking about the mind adopted by the simple heuristics program. One way to spell this out would be to propose the existence of heuristic procedures that can be multiply-instantiated within a range of distinct modular systems. (So on this account, a given heuristic is a *type* of processing rule, which can be instantiated many times over within different systems in the brain, rather than a processing system in its own right.) For there is no reason to think that each module has to deploy a *unique* set of algorithms for processing items in its domain. There might be a range of algorithm types / heuristic strategies that have been adopted again and again by different modular systems.[5] Nor is there any reason to think that modular systems should use 'maximizing' or

[5] 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

exhaustive algorithms. On the contrary, pressures of speed and time should lead to the evolution of 'quick and dirty' intra-modular decision-rules, just as they lead to such rules for cognition as a whole.

Equally (and in addition), massive modularity certainly isn't inconsistent with learning. Many modules are best characterized as learning modules, indeed. And some modules are likely to be built by other learning mechanisms, rather than being innate. (The example of *reading* comes to mind.) Moreover, while some of these mechanisms might be targeted on just a single content-domain, some might involve the interactions of a variety of different modules and modular learning mechanisms (hence giving he appearance of domain generality). And then it may well be that there exists a suite of heuristic operating principles that can be selected from amongst some sort of pre-existing 'toolbox' for implementation in one of these modules, if appropriate. The learning process would partly involve the selection of the appropriate tool from the toolbox.

5 Is the argument from computational tractability undermined?

We have seen that the massive modularity hypothesis seems to be fully consistent with the claims made by the simple heuristics program. It appears, nevertheless, that the successes of the latter program must undermine one of the main arguments *in support of* massive modularity – specifically the argument from computational tractability. For it looks like heuristic-based computational mechanisms offer a way for computations to be rendered tractable *without* the need for informational encapsulation. If so, then cognitive processes can be computationally tractable (because heuristic-based) without being structured along modular lines, and it will turn out that the argument, 'Cognition must be modular in order that it should realized in a computationally tractable form', collapses.

In order to evaluate this objection, we will need to undertake a closer examination of the notion of encapsulation. But I propose to approach this initially by going back one step further: to the considerations of computational tractability that supposedly give rise to the demand that cognition should be constructed out of encapsulated systems.

We can distinguish two different ways in which a system might fail to be computationally

or more of the copies. But often it will just involve provision of novel input and/or output connections for the new system.

tractable. One is that its algorithms might require it to consult *too much information* to reach a solution in real time. For example, consider a very simple consistency-checking device. Given a candidate new belief, it crawls through the total set of the subject's beliefs, looking for an explicit contradiction. Although the algorithm being executed here might be an extremely simple one (essentially it just looks for any pair of beliefs of the form, *P*, ~*P*), in attempting to take every belief as input (whether sequentially or in parallel) it almost certainly wouldn't be feasible for mind/brains like ours. Call the corresponding demand on computationally tractable systems, *information-frugality*. We can say that cognition needs to be realized in information-frugal systems if it is to be tractable.

The other way in which a system might fail to be computationally tractable, is if the algorithms that it runs are *too complex* to be feasibly executed in real time. Consider, for example, a consistency-checker that operates using the following crude heuristic, which only requires it to consider relatively small sets of beliefs. For any candidate new belief, it randomly selects a smallish set of a hundred or so other beliefs and generates a truth-table for the total set of propositions, checking each line to see if there is a complete set of 'F's on any line. It is easy to see that the amount of time and working memory that this system would need in order to complete its task would go up exponentially with the size of its input-set. As we noted earlier, even if it checks each line of the truth-table in the time that it takes light to travel the diameter of a proton, it would take the system longer than the time that has now elapsed since the beginning of the universe to check the consistency of just 138 beliefs – and note that this doesn't include the time needed to generate the truth-table in the first place! Call the corresponding demand on computationally tractable systems, *processing-frugality*. We can say that cognition needs to be realized in processing-frugal systems if it is to be tractable.[6]

The argument from computational tractability, then, leads us to think that cognition

[6] It should be stressed that the notions of *too much* information, and of processing that is *too complex*, as deployed here, remain highly indeterminate. Common-sense reflection on the circumstances of human life can get us some sense of the sorts of time-scales within which cognitive systems have to perform their tasks, of course. But most of the other parameters that would need to be taken account of are lacking. We don't know much about the *speed* of processing of brain-systems, when described at a cognitive as opposed to a neurological level. Nor do we know very much about the memory capacity of the various systems that might be involved. So any judgment that we make to the effect that a given system is or isn't sufficiently *frugal* will have to be tentative, and hostage to future discoveries in cognitive science.

should consist of systems that are *both* information-frugal *and* processing-frugal. Now one way of making a system information-frugal would be to deny it access to any stored information at all. This gives us the archetype of an input-module, as explored by Fodor (1983). This would be a system that can receive and process sensorily-transduced information, but which can't access any of the stored information held elsewhere in the mind. But of course this can't be a general model of what a module should look like, if we are seeking to extend the notion of modularity to central systems that operate on beliefs (and desires) as input. Once we shift away from considering *input*-modules to putative *central*-modules, then we can no longer think of encapsulation as a matter of isolating the system from stored information. For central modules will often need to operate on stored information as input.

The natural way forward, at this point, involves distinguishing between the *input to* a module and the *processing data-base of* a module (Sperber, 2002; Carruthers, 2003). A non-content-specific central module would be a system that could take any stored information as input, but that would be encapsulated in respect of its processing – either it can access *no* stored information in executing its algorithms (in which case the system is wholly encapsulated), or it can only access a limited data-base of information that is relevant to the execution of those algorithms (in which case the system is encapsulated to a degree inversely proportional to the size of the data-base).

With this rough suggestion on the table, the issue comes down to this. If computational tractability (and hence *frugality*) requires informational encapsulation, then for each computational system and sub-system: (a) we must be able to identify its input, and distinguish this from its processing data-base (if any); and (b) its processing data-base must be a small sub-set of the total amount of information available. If the simple heuristics program suggests a way in which one can have frugality in the absence of either (a) or (b), in contrast, then the argument from computational tractability to massive modularity would seem to be undermined.

6 Heuristics and processing data-bases

In order to see whether or not the simple heuristics program undermines the argument for processing encapsulation, then, we need to examine whether particular applications of that research program – such as the Recognition heuristic, Take the Best, Take the Last, and so on – can support a suitable division between a system's *input* and its *processing data-base*.

The Recognition heuristic operates somewhat as follows. When required to decide which of two items scores higher along some dimension (e.g., which of two German cities is the larger), if you only recognize one of the two items, then select that one. (If both items are recognized, then some other heuristic must be employed.) For our purposes, the important point to notice is the recognition heuristic is fully encapsulated in its operation. No other information in the mind either does or can influence the outcome, except perhaps information that is somehow implicated in the recognition process itself. Once the system has received a judgment-task to process, it just has to look to determine which of the objects presented to it evokes recognition.[7] No other information needs to be consulted (nor can it be, indeed, or at least not internally to the operation of recognition heuristic itself), and the inferential procedure involved is a very simple one. So it would appear that instantiations of the recognition heuristic deserve to be counted as *modules* in the traditional sense.

Now consider the Take the Best heuristic. Unlike Recognition, this heuristic does require the system to search for and consult some further information concerning the items in question. But it doesn't look at *all* the information concerning those items. Specifically, it searches for the item of information concerning the two target items that has most often been found in the past to discriminate between items of that type along the required dimension. Gigerenzer *et al*. (1999) have shown that this heuristic can perform almost as well as a bunch of fancier processing algorithms, but it can do so while being much more frugal in the information that it uses and the demands that it places on the computational resources of the system.

In this case it isn't easy to see how the distinction between *input* and *processing data-base* should be drawn. One might try saying that the relevant sub-set of total information, that a system instantiating Take the Best can consult during processing, consists of its beliefs about relative cue validity together with its further beliefs concerning the cues in question. When it gets a query about the relative size of two German cities, for example, it must look first at its beliefs about which properties of cities have correlated best with size in the past. If having a top-division soccer team was the best predictor, then it will query the wider data-base: does either of these teams have a top-division soccer team? If it receives back the information that just one of them

[7] Hence the processing data-base for the system would consist in the set of concepts possessed, together with any information required for object recognition. This is likely to be a small sub-set of the total information contained within a mind.

does, it selects that one as the larger. If neither or both do, it moves on to the next best predictor of size listed in its processing data-base. And so on.

Note, however, that which beliefs such a system *can* consult in the course of its processing is a function of what its beliefs actually *are*. If the system had believed that having a high crime rate was the best predictor of city size, then *that* is the information that it would have sought out. And in principle any belief *could* have had an impact on processing. So it seems that our best hope of finding a place for the notion of 'encapsulation', here, would be to adopt an idea from our discussion of the SOAR planning architecture. We could regard the specific beliefs that the system instantiating Take the Best happens to acquire as carving out a functionally-individuated processing data-base from the wider set of stored information in relation to each dimension of comparison, such that the system *can* only consider that narrower set in answer to a given question. But it has to be admitted that this looks pretty forced and unnatural.

Now consider heuristic processes that rely upon such phenomena as the *salience* of a piece of information in a context, or the *accessibility* of that information given the recent history of its activation. (This latter is closely related to the heuristic that Gigerenzer *et al*., 1999, call 'Take the Last'.) Consider language comprehension, for example, on the sort of model provided by Sperber and Wilson (1996), in which accessibility of beliefs plays a major role. On their account, one of the factors in interpretation is saliency in the present environment, and another is relative recency (e.g. whether or not an item of information has been activated earlier in the conversation).

Might the comprehension process nevertheless count as an encapsulated one, although in principle *any* belief *might* be made salient by the present environment, or *might have been* activated previously? If so, we shall have to think of the comprehension process, as it unfolds in the course of a set of linguistic exchanges, as creating a sort of local comprehension module 'on the fly', whose encapsulation-conditions are continually modified as the conversation continues. But what becomes of the idea that there is some sub-set of the total information available that the comprehension system can look at, if *any* item of information *could* have been salient?

It might be replied, however, that we are dealing here with a *briefly-existing* encapsulated system, created out of the resources of a longer-lasting comprehension system by facts about the recent environment. *Given* the previous history of the conversation, then some items of information are much more accessible than others. So a search process that operates on principles of accessibility *can* only look at that information, and other information in the mind *can't* influence the comprehension process. Granted, if the earlier facts about the conversation had been different, then other information could have had an influence on the comprehension of the sentence in question. But this doesn't alter the fact that, the previous history of the conversation having been what it was, that information *cannot* now have an influence.

Although there is a sense in which this reply works, the victory is a Pyrrhic one. For the resulting notion of modularity is highly problematic. Cognitive science, like any other science, is in the business, *inter alia*, of discovering and studying the properties of the set of *natural kinds* within its domain. And a natural kind, in order to be a worthwhile object of study, must have a certain sort of *stability*, or regular recurrence. In contrast, the state of a comprehension system that has undergone a specific conversational history, and hence that has a particular distribution of degrees of accessibility amongst its representations, is something that might exist just once in the history the universe. That particular combination of processing principles and accessibility (yielding the 'processing data-base' of an on-the-fly module) might never recur again.

If cognitive science is to attain the sort of generality that one expects of a science, it needs to carve its kinds at *recurring* joints. This requires us to think of the comprehension system as a *single* system over time, operating partly on principles of accessibility that help to make its operations information-frugal. Likewise, even in the case of SOAR (to return to an example that occurred much earlier in our discussion; similar things could be said about Take the Best, discussed more recently): we should probably think of this as being the *same* system that is employed in pursuit of a variety of different goals, in which information-frugality is ensured by organizing its data-base into 'frames' linked to each type of goal. We shouldn't think of it as a whole set of overlapping encapsulated systems (one for each processing-system-and-'frame' pair), that share a common set of algorithms.

7 Input information *versus* processing data-base

Some examples of processing drawn from the simple heuristics program appear to put severe pressure on the notion of an *encapsulated* system, then, where the latter is explicated in terms of an intuitive distinction between *input* and *processing data-base*. It is worth asking directly how this distinction is to be explicated in turn, however. The arguments above attempt to *use* the distinction between input and processing data-base, without saying what that distinction amounts to. But how is this distinction to be drawn?[8]

One way to do it, would be to think of the input to a system as *whatever turns the system on*. But this doesn't seem a very plausible proposal. Surely we would want to allow that, once 'turned on' by something (e.g. by one desire winning out over others in the competition to control the resources of the practical reasoning system), a system might query other systems for information without all such information thereby being counted as belonging to the processing data-base of the system in question. AI, as currently practiced, is full of networks of distinct systems that can query each other for information after they have been 'turned on'. But if we were to adopt the above proposal, then we should have to say that there was a sense in which they were all really *one* big system, since information produced by each one would form part of the processing data-base of each of the others.

Another way we might go, would be to say that the processing data-base of a system, to count as such, must be a *dedicated* data-base, whose contents aren't available to other systems. This fits quite well with the way in which people think about the language module. One might regard the processing data-base for the language faculty as a set of acquired language-specific items of information – concerning grammatical and phonological rules, for example – which isn't available to any other system to make use of. It doesn't seem well-motivated, however, to insist that memory systems and processing systems should line up one-for-one in the way that this suggestion postulates. For modularity is a thesis about *processing*, not a thesis about storage. It is quite unclear why there shouldn't be multi-purpose information storage systems that can be accessed by any number of processing systems in the course of their processing. Nor is it clear why the modular status of those processing systems would have to be compromised as a result.

Another alternative is to think of the processing data-base of a system as the body of information that it *must* consult in order to execute its algorithms. The *input* to the system (if the system isn't a content-specific one) could in principle come from anywhere. But once the system is turned on, it would be required to start consulting some part of its processing data-base in order to handle its input. (The system needn't attempt to consult *all* of the information in its

[8] Note that the distinction is only problematic in respect of central modules, whose input can include beliefs or other stored propositional states. Where the module in question is an input-module, the distinction is straightforward: the input to the system is the information that reaches it from the sensory transducers, and any other information that gets used in processing can be counted as belonging to the processing data-base.

processing data-base, of course. This is one of those places where it is helpful to imagine various heuristics and search-rules operating within a given module.)

This proposal seems to fit quite neatly with the ways in which people tend to think about a language module, or a theory of mind module, for example. When linguistic input is received, the language module *must* start consulting its data-base of grammatical and phonological rules, its lexicon, and so forth. In the course of processing it *might* also send out requests to other systems for information concerning the speaker's likely interests or knowledge, for example; and the replies to these queries can be thought of as further inputs to the system. Likewise when a description of an item of behavior is received by the theory of mind system: it *must* start consulting its data-base of acquired information (e.g. concerning the conventional significance of a handshake in the surrounding culture, or concerning the mental states previously attributed to the actor). And in the course of its processing it, too, might send out requests to other systems for information concerning the likely physical consequences of the subject's observed behavior, for instance.

There are good reasons why one can't explain encapsulation in terms of the information that the system *must* consult, however. Consider the practical reason module sketched above. Once it has been turned on by a desire, it's algorithms *require* it to seek information of a certain conditional form. But these conditional beliefs can in principle come from anywhere and be about anything. So if we said that the processing data-base for a system is the set of beliefs that it is *required* to consult, then almost all beliefs might belong to this data-base, and practical reason would turn out to be radically unencapsulated and a-modular after all.[9]

A final option is to make use of the distinction between conducting an information-search oneself, and sending out a query for information from elsewhere. We could say that the processing data-base for a module is the stored information that it (the module) searches, rather than the search being devolved to other systems. But now, focus on that aspect of practical reason's requirement for conditional beliefs that involves a search amongst *stored* conditional

[9] Am I begging the question by *assuming* here that practical reason is modular? No. For what is in question is whether there is a notion of module 'encapsulation' that can be put to the service of a massive modularity thesis; and the latter must surely maintain that practical reason is modular. Moreover the practical reason system, as initially sketched above, did seem like it might be computationally tractable. So if tractability is supposed to require encapsulation, then again we need a notion of encapsulation that can fit the operations of such a system.

information (as opposed to requests to other systems to see if they can *generate* such information in the circumstances). Why *shouldn't* this be conducted by the practical reason system itself? If memory is content-addressable, or addressable by syntactic form, one might be able to search under the description, 'conditional with *P* as a consequent'. And is there any reason to think that conducting such a search would render practical reason intractable? (Or any more intractable than if there were some other system to which this search was devolved?)

It appears that the distinction between *input* and *processing data-base* can't do the work required of it in the context of a thesis of massive mental modularity, then – at least, not if we want to allow for modular systems that can take unrestricted input, that can query other systems for information, that can conduct searches for information themselves at various points during their processing, and so forth. At this point, the notion of *encapsulation*, and with it the notion of *modularity*, appears to be under severe pressure, in the context of a thesis of massive mental modularity.

8 Whither modularity?

The simple heuristics program places considerable pressure on the claim that cognition must be constructed out of *encapsulated* systems, then, if an encapsulated system is one that might be capable of receiving anything as input, but which can only access a limited data-base of information in the course of its processing. But is this how we *must* understand the notion of encapsulation? Are there any alternatives open to us?

Put as neutrally as possible, we can say that the idea of an encapsulated system is the idea of a system whose operations *can't* be affected by *most or all* of the information held elsewhere in the mind. But there is a scope ambiguity here. 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, following Fodor (1983).[10] Most of us naturally picture modularity in terms of there being some determinate body of information, such that *that* information can't penetrate the module. And this way of looking at the matter is only reinforced if we explicate encapsulation in terms of the distinction between the input to a system and its processing data-base – for here there is supposed to be some determinate body of information (the information in the processing data-base) that *can* affect the operations of the module; implying that all other information *can't* (except by being taken as input).

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 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 to be both information-frugal and processing-frugal.

In the end, then, the following argument is a failure: if cognitive processes are to be tractably realized, then the mind must be constructed out of networks of processing systems that are encapsulated in the narrow-scope sense. So the argument for massive modularity as traditionally conceived of (by philosophers) fails too. But we still have the argument that computational tractability requires *wide*-scope encapsulation – we still have the argument that if cognitive processes are to be tractably realized, then the mind must be constructed out of systems whose operations are both information-frugal and processing-frugal; and this means that those systems must only access a small sub-set of the total available information while executing their tasks.

Does this mean that the thesis of massive mental modularity is insufficiently supported,

[10] The use of the term 'module' in the AI literature is probably rather different, however (personal communication: Joanna Bryson, Jack Copeland, John Horty, Aaron Sloman). It may be closer to a combination of the everyday sense of 'module' meaning 'functionally-individuated processing component', together with a requirement of what I am here calling 'wide-scope encapsulation'. If so, then the argument for massive modularity from recent trends in AI, sketched in section 2, can still hold up, given the intended sense of 'modularity'.

and should be rejected? Well that, of course, depends upon what we continue to mean by ‘a module’. We still have in play the argument from biology that we should expect cognition to be built out of separable task-specific systems (this is the everyday meaning of ‘module’). And we still have the argument from computational tractability, that these systems need to be both information-frugal and processing-frugal. This requires that those systems should be wide-scope encapsulated. (They need to be systems that can’t access more than a small sub-set of the total information available before completing their tasks.) And it is open to us to say that this is how the thesis of massive modularity should properly be understood.

Moreover, we still have in play the meta-inductive argument from recent trends in AI. 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 systems, 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. The fact of such convergence is then good evidence that this is how the mind, too, will be organized.[11]

The term ‘module’ has been used in *many* different ways within the cognitive science literature, of course, from Fodor (1983) onwards. One reaction to this mess of different usages would be to urge that the term should be dropped, and that people should describe in other words what it is that they believe. But it is actually quite handy to have a single term to express what one means. And provided that one is explicit about how that term is being used, no confusion should result. I propose, then, that by ‘module’ we should mean something along the lines of, ‘a distinct task-specific processing system whose operations are both information-frugal and

---

[11] Indeed, the convergence is actually wider still, embracing computer science more generally. 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 like C++ and Java now require a total processing system to treat some of its parts as ‘objects’ which can be queried or informed, but where the processing that takes place within those objects isn’t accessible elsewhere. And the resulting architecture is regarded as well nigh inevitable whenever a certain threshold in the overall degree of complexity of the system gets passed.

processing-frugal (and hence which is wide-scope encapsulated)'.[12] And the thesis of massive modularity then becomes the claim that cognition must be built up out of such systems. Thus understood, the thesis of massive mental modularity is both well-supported, and fully consistent with the insights of the simple heuristics research program.

What really matters in the end, of course, isn't what the systems in question get *called*, but rather what we can claim to know about the architecture of the human mind. There are a range of different arguments (not all of which could be surveyed in this chapter – for further examples see Carruthers, 2005), together with a set of progressive research programs in cognitive science and AI, all of which suggest that the mind is, indeed, composed of a multitude of distinct processing systems. These systems will talk to one another and query one another, but will to a significant degree operate independently of one another. And their processing will be *frugal* (either using algorithms tailored specifically to the task demands, or using heuristics or short-cuts of one sort or another, or both). I am myself inclined to express this result by saying that the mind is massively modular in its organization; but what matters is the result itself, not the manner in which it is described.

Acknowledgments

I am grateful to Stephen Stich for a conversation that prompted the main set of thoughts in this chapter; and to Richard Samuels for an earlier exchange that also proved fruitful. Thanks to Mike Anderson, Clark Barrett, Stephen Laurence, Rui Mata, Stephen Stich, and Peter Todd for comments on an earlier draft.

References

Brooks, R. 1986. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2: 14-23.

Bryson, J. 2000. Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12:165-190.

Carruthers, P. 2002a. The cognitive functions of language. *Behavioral and Brain Sciences*, 25:6, 657-674.

[12] For consideration of a wider set of arguments in support of massive modularity, and a resulting notion of 'module' that isn't quite the same as the one outlined here (incorporating the idea of *inaccessibility*), see Carruthers, 2005.

Carruthers, P. 2002b. Modularity, language, and the flexibility of thought. *Behavioral and Brain Sciences*, 25:6, 705-719.

Carruthers, P. 2002c. 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. 2004. Practical reasoning in a modular mind. *Mind and Language*, 19, 259-278.

Carruthers, P. 2005. The case for massively modular models of mind. In R. Stainton (ed.), *Contemporary Debates in Cognitive Science*. Blackwell.

Cheng, K. 1986. A purely geometric module in the rat's spatial representation. *Cognition*, 23, 149-178.

Cherniak, C. 1986. *Minimal Rationality*. MIT Press.

Evans, J. and Over, D. 1996. *Rationality and Reasoning*. Psychology Press.

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.

Gigerenzer, G., Todd, P. and the ABC Research Group. 1999. *Simple Heuristics that Make us Smart*. Oxford University Press.

Hermer, L. and Spelke, E. 1996. Modularity and development: the case of spatial reorientation. *Cognition*, 61, 195-232.

Hermer-Vasquez, L., Spelke, E. and Katsnelson, A. 1999. Sources of flexibility in human cognition: dual-task studies of space and language. *Cognitive Psychology*, 39, 3-36.

Kahneman, D., Slovic, P. and Tversky, A. (eds.) 1982. *Judgment under Uncertainty: heuristics and biases*. Cambridge University Press.

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.

McDermott, D. 2001. *Mind and Mechanism*. MIT Press.

Newell, A. 1990. *Unified Theories of Cognition*. Harvard University Press.

Pinker, S. 1997. *How the Mind Works*. Penguin Press.

Sachs, O. 1985. *The Man who Mistook his Wife for a Hat.* Picador.

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.

Seeley, T. 1995. *The Wisdom of the Hive: the social physiology of honey bee colonies*. Harvard University Press.

Shusterman, A. and Spelke, E. 2005. Investigations in the development of spatial reasoning. In P. Carruthers, S. Laurence and S. Stich (eds.), *The Innate Mind: structure and contents*, Oxford University Press.

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. and Wilson, D. 1996. *Relevance*. Blackwell. (2nd Edition.)

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.

Wason, P. and Evans, J. 1975. Dual processes in reasoning? *Cognition*, 3, 141-154.