Précis of *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought* (Oxford University Press, 2006)

PETER CARRUTHERS

ABSTRACT: This article outlines the main themes and motivations of Carruthers (2006). Its purpose is to provide some background for the critical commentaries of Cowie, Machery, and Wilson (this volume).

1. Introduction

My book has three main goals. One is to make the strongest possible case in support of a massively modular conception of the human mind (or the strongest that can be made on the basis of current evidence, at least; the thesis is an empirical one, of course, and new data are coming in all the time), while also explicating how such an account should properly be understood.[1] The second goal is architectural: it is to lay out the basic framework of mind within which the multiple modules are located, and to sketch some of the components and their modes of connectivity and interaction with one another. And then the final goal is to show that such an account has the wherewithal to explain the various distinctive features of the human mind, especially those that are believed by many to be especially problematic for massively modular theories (and perhaps for *any* account), including creative and flexible thinking, and our capacity for science. These three main strands in the book are to some degree independent of one another, although they are also mutually supporting. I shall devote a section of this article to each in turn.[2]

Address for correspondence: Department of Philosophy, University of Maryland, College Park, MD 20742, USA.
Email: pcarruth@umd.edu

[1] I shall defer any discussion of the meaning of 'modularity' to my reply (this volume), since this provides a focus for quite a bit of the critical commentary.

[2] For references supporting the claims made in this article, readers are invited to consult the relevant portions of Carruthers (2006).

2. The Case for Massive Modularity

The evidence supporting a massively modular conception of the mind is multifaceted, and discussion of it is distributed over the first three chapters of the book. But three broad lines of argument can be distinguished. The first derives ultimately from Simon (1962), and concerns the design of complex functional systems quite generally, but in biology in particular. All such systems are constructed hierarchically out of dissociable sub-systems (each of which is made up of yet further sub-systems), in such a way that the whole assembly can be built up gradually, adding sub-system to sub-system. Each sub-system is characterized by significantly more inter-connections and interactions internally, within itself, than it has with others outside it. And the properties of sub-systems can be varied independently of one another, in such a way that the functionality of the whole is buffered, to a significant extent, from changes or damage occurring to the parts.

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, and that this sort of modular organization is in fact a pre-requisite of evolvability. This is true for the operations of genes, of cells, of cellular assemblies, of whole organs, of organ assemblies (like the respiratory system), of whole organisms, and of multi-organism units like a bee colony. And by extension, we should expect it to be true of cognition also, provided that (as I argue) it is appropriate to think of cognitive systems as biological ones, which have been subject to natural selection.

Moreover, just as this account would predict, there is now a great deal of evidence of a neuro-psychological sort that massive modularity is indeed a characteristic of the human mind. People can have their language system damaged while leaving much of the remainder of cognition intact (aphasia); people can lack the ability to reason about mental states while still being capable of much else (autism); people can lose their capacity to recognize just human faces; someone can lose the capacity to reason about cheating in a social exchange while retaining otherwise parallel capacities to reason about risks and dangers; someone can lose the capacity to name living things while retaining the capacity to name non-living things, or vice versa; someone can lose the capacity to name fruits and vegetables while retaining their ability to name animals; and so on and so forth.[3]

The second strand of argument supporting massive modularity is grounded in the modularity of non-human animal minds, on the assumption that this would have been conserved during the transition to the human mind, as adaptations characteristically are. In contrast with the account still commonly accepted by philosophers—that animal learning is a matter of associatively pairing stimuli with conditioned responses—it is argued that there is actually *no such thing as* general learning. Following Gallistel (1990), and reviewing a wide range of literature on animal cognition, Chapter 2 of my book argues that animals utilize an extensive set of specialized learning mechanisms—for calculating the solar ephemeris, for estimating distance traveled, for calculating temporal intervals, for doing dead reckoning, for constructing mental maps of the environment and storing and updating information about what is located in the places represented, for calculating rates of return when foraging, for figuring out causal dependencies amongst events, and many, many, more. Moreover, the evidence suggests that each of these learning systems possesses its own specialized memory system which is dissociable from the others.

In addition, Chapter 2 argues that there are multiple emotion-forming and desire-forming mechanisms in the minds of animals, each of which has been designed to take informational states of distinctive sorts as input (sometimes concerning properties of the animal's own body, sometimes concerning features of the current environment, sometimes both) and to generate an adaptive goal in consequence. These goal-states, when active, initiate a search for information that would enable the animal to satisfy its desire in the circumstances. But goals of a given type are generally restricted in the *kinds* of information that get searched for. So practical reasoning, too (that is: reasoning about what to do in the light of one's desires) fragments into a variety of distinct sub-systems. In fact, animal minds are modular through-and-through—from perceptual

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

systems right through to motor-control systems. It would be extraordinary, then, if the same were not true (at least for the most part) in the human mind.

The third line of support for massive modularity takes its start from Fodor (2000) and argues (a) that cognitive processes are computational in character, and (b) that complex computational systems, to be feasible, need to be constructed out of sets of specialized sub-systems, each of which is frugal in its use of resources and its need for information. This is then supplemented by evidence of the expensiveness of long-distance neural connections (whether measured in terms of processing time or energetic and other resources consumed), suggesting that brains should be organized, to a significant extent, into local networks of highly inter-connected neurons—which is just what we find. Indeed, the evidence is that as brains grow larger, they become *more* modular in their organization, with the proportion of local neural connections over long-distances ones increasing.

The upshot of all three sets of considerations, then, is that minds are massively modular in their organization. Whether the human mind, like animal minds, is *exclusively* modular depends on what got added in the course of the evolutionary transition from the common ancestor of the great ape lineages to ourselves. (I suggest that nothing significant would have been lost.) This will be addressed in Section 3.

3. The Architecture of the Modular Mind

I argue in Chapter 2 that almost all minds (including the minds of many types of insect) are organized into sets of perceptual systems which feed into belief-generating and goal-generating systems, and which also inform practical reasoning in light of the goals so generated. This provides the basic framework within which novel perceptual modules and sub-modules (e.g. for face recognition), as well as new belief modules and desire modules, can evolve. Moreover, in many minds each type of perceptual system (such as vision or audition) bifurcates into two distinct sub-systems, one of which is concerned with the fast on-line guidance of physical movement (and which further subdivides into modules that guide specific types of movement, or specific limbs), and the other of which broadcasts its outputs to the conceptual modules for forming beliefs and desires, and for reasoning in relation to the perceived environment. Neither of these types of perceptual system operates entirely ‘bottom–up’. On the contrary. Within the movement-guiding systems, efferent copies of motor instructions are used to form representations of the perceptual input that should occur if the action is executed as planned, which can then be matched against incoming data for even swifter and more fine-grained control. Likewise within the perceptual systems that feed to the conceptual modules: the latter are continually querying the incoming data, using appearance-templates associated with distinct concepts in an attempt to find a best match.

Importantly for the later themes of the book, the two sorts of perceptual system aren’t wholly independent of one another. For in monkeys and apes (at least) perceptual representations of actions generated within the action-guiding system can be transformed to create appropriately similar perceptual images within the other, conceptualizing, system (and vice versa). These can then be worked on by the conceptual belief-systems to create representations of the likely further consequences of the action, which can in turn be displayed imagistically for purposes of further inference. Amongst hominids, and perhaps also other great apes, these back-projecting pathways are extensively used in reflective forms of practical reasoning. A potential action schema can be mentally rehearsed (with overt movement suppressed) and used to generate predictions of the likely consequences. These in turn, when imagistically represented, can be taken as input, not only by the belief modules, but also by the various motivational modules, which react as if to the real thing. We then monitor our reactions to these imagined scenarios (our hearts sink, our spirits lift, and so forth), and our motivation to perform the envisaged action gets adjusted up or down accordingly.

Against this background, Chapter 3 argues that a number of additional modules were added in the course of human evolution (for the most part deepening and extending systems that were already partially present amongst other great apes). A number of these are discussed in my reply to my critics in the present symposium. But the most important additions (from the perspective of the arguments of later chapters) include: a mind-reading system, capable of attributing mental states to others and to oneself; a language-learning system, designed to build modular production and comprehension systems suited to the surrounding language; and a normative reasoning and motivational system, which forms judgments about what is required, forbidden, or permitted, and generates motivations (such as guilt or indignation) accordingly. In addition, Chapters 2 and 5 suggest that humans evolved an innate disposition to generate and rehearse action-plans *creatively*, utilizing a variety of constraints and heuristics. The first manifestation of this disposition in development is seen in early childhood in the form of episodes of pretend play.

4. Answering the Challenges

Most of the second half of the book (Chapters 4–7) is occupied with the task of showing how a massively modular mind of the sort sketched above could display all of the main properties distinctive of the human mind. Chapter 4 distinguishes a number of different kinds of cognitive *flexibility*. It shows how the context-sensitivity of our cognitive processes could be achieved via the workings of multiple modules, while explaining how the remarkable stimulus-independence of human thinking is achieved through cycles of mental rehearsal. The chapter then discusses how the language production system would be ideally positioned to take input from a variety of different belief modules, encoding these into sentences before unifying the results into a single sentence. The latter can then be uttered, or it can be mentally rehearsed in 'inner speech', hence giving rise to a kind of content-flexibility. Finally, the chapter discusses how sequences of mentally rehearsed sentences (together with other images) can give rise to a whole new *kind* of thinking and reasoning. (This idea is also pursued in Chapters 6 and 7.) For example, amongst our normative beliefs can be beliefs about how we *should* reason. And since inner speech is action-based, a normative belief of this sort can motivate us to token specific types of sentence having already tokened others.

Chapter 5 tackles the problem of creativity. Traditional accounts take creative thought to be prior to creative action, with the latter to be explained in terms of the former. This chapter argues the reverse. It suggests that it is creatively generated action schemata that, when rehearsed, give rise to novel thoughts. One advantage of this account is that we already know that many species of animal are capable of simple forms of creative action that don't depend upon prior thought, such as the 'protean behavior' that many animals exhibit when fleeing a predator. Our suggestion can therefore be that this capacity was deepened and extended with the advent of mental rehearsal of action (especially speech action). Indeed, I suggest that the proper function of childhood pretend play is to begin developing a set of fruitful heuristics for creative action rehearsal.

Chapter 6 discusses our capacity for science, which some have claimed to elude *any* cognitive explanation, let alone a massively modular one (Pinker, 1997; Fodor, 2000). This breaks down into three component capacities. One is the capacity for creative hypothesis generation, which has been explained, in outline, in Chapter 5. Another is the capacity to draw inferences from a hypothesis. This will utilize the resources of the conceptual modules, together with normative beliefs about how one *should* reason, as explained in Chapter 4. What remains is to explain our capacity for inference to the best explanation. This is shown to be an exaptation, utilizing principles involved in language comprehension and in the acceptance (or rejection) of people's verbal testimony.

Finally, Chapter 7 discusses a number of topics having to do with distinctively human *practical* reason. One strand involves explaining our tendency (in both our thought and our talk) to cast our desires in descriptive format ('P would be good/best', 'I want P'), rather than directly expressing them in the optative mood ('Would that P!'). My suggestion is that this enables our desires to be brought within the framework of distinctively human *theoretical* reason, as explained in Chapters 4 through 6. I also show how such descriptively-characterized desires can nevertheless possess a practical role in the initiation of action. Another strand in the chapter is to defend the belief–desire architecture adopted throughout the book against those philosophers who claim that it is perceptions of *reasons* rather than desires that fundamentally motivate human action (e.g. Scanlon, 1998). Finally, the chapter shows how some of the views developed in the book serve to vindicate Wegner's (2002) claim that conscious will is an illusion.

5. Conclusion

The Architecture of the Mind is an ambitious book. It surveys and integrates a wide range of empirical literature in cognitive science, both in support of a massively modular architecture for human and animal minds, as well as to develop an account that can explain the main elements of distinctively human cognition. The extent to which it does all this successfully is for others to judge. But I believe that it provides a well-articulated vision of the human mind, at least, which can serve as a focus for productive discussion and critique, as well as for future research.

Department of Philosophy
University of Maryland

References

Carruthers, P. 2006: *The Architecture of the Mind: massive modularity and the flexibility of thought*. Oxford University Press.

Fodor, J. 2000: *The Mind Doesn't Work That Way*. MIT Press.

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

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

Scanlon, T. 1998: *What We Owe to Each Other*. Harvard University Press.

Simon, H. 1962: The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467-482.

Wegner, D. 2002:. *The Illusion of Conscious Will*. MIT Press.