# 3

## In Defense of Massive Modularity

Dan Sperber

In October 1990, a psychologist, Susan Gelman, and three anthropologists whose interest in cognition had been guided and encouraged by Jacques Mehler, Scott Atran, Larry Hirschfeld, and myself, organized a conference on "Cultural Knowledge and Domain Specificity" (see Hirschfeld and Gelman, 1994). Jacques advised us in the preparation of the conference, and while we failed to convince him to write a paper, he did play a major role in the discussions.

A main issue at stake was the degree to which cognitive development, everyday cognition, and cultural knowledge are based on dedicated domain-specific mechanisms, as opposed to a domain-general intelligence and learning capacity. Thanks in particular to the work of developmental psychologists such as Susan Carey, Rochel Gelman, Susan Gelman, Frank Keil, Alan Leslie, Jacques Mehler, Elizabeth Spelke (who were all there), the issue of domain-specificity—which, of course, Noam Chomsky had been the first to raise—was becoming a central one in cognitive psychology. Evolutionary psychology, represented at the conference by Leda Cosmides and John Tooby, was putting forward new arguments for seeing human cognition as involving mostly domain- or task-specific evolved adaptations. We were a few anthropologists, far from the mainstream of our discipline, who also saw domain-specific cognitive processes as both constraining and contributing to cultural development.

Taking for granted that domain-specific dispositions are an important feature of human cognition, three questions arise:

1. To what extent are these domain-specific dispositions based on truly autonomous mental mechanisms or "modules," as opposed to being

domain-specific articulations and deployments of more domain-general abilities?

2. What is the degree of specialization of these dispositions, or equivalently, what is the size of the relevant domains? Are we just talking of very general domains such as naive psychology and naive physics, or also of much more specialized dispositions such as cheater-detection or fear-of-snakes?

3. Assuming that there are mental modules, how much of the mind, and which aspects of it, are domain-specific and modular?

As a tentative answer to these three questions, I proposed in some detail an extremist thesis, that of "massive modularity" (Sperber, 1994). The expression "massive modularity" has since served as a red cloth in heated debates on psychological evolution and architecture (e.g., see Samuels, 1998; Murphy and Stich, 2000; Fodor, 2000). I was arguing that domain-specific abilities were subserved by genuine modules, that modules came in all formats and sizes, including micromodules the size of a concept, and that the mind was modular through and through. This was so extremist that archmodularist John Tooby gently warned me against going too far. Jacques Mehler was, characteristically, quite willing to entertain and discuss my speculations. At the same time, he pointed out how speculative indeed they were, and he seemed to think that Fodor's objections against "modularity thesis gone mad" (Fodor, 1987, p. 27) remained decisive. I agreed then, and I still agree today, that our understanding of cognitive architecture is way too poor, and the best we can do is try and speculate intelligently (which is great fun anyhow). I was not swayed, on the other hand, by the Fodorian arguments. To his objections in *The Modularity of Mind* (1983), Fodor has now added new arguments in *The Mind Doesn't Work That Way* (2000). I, however, see these new considerations as weighing on the whole for, rather than against, massive modularity.

Modularity—not just of the mind but of any biological mechanism—can be envisaged at five levels:

1. At a morphological or *architectural level,* what is investigated is the structure and function of specific modules, and, more generally, the extent to which, and the manner in which, the organism and its subparts,

in particular the mind/brain, are an articulation of autonomous mechanisms.

2. At the *developmental level,* modules are approached as phenotypic expressions of genes in an environment. Cognitive modules in particular are hypothesized to explain why and how children develop competencies in specific domains in ways that could not be predicted on the basis of environmental inputs and general learning mechanisms alone.

3. At the *neurological level,* modules are typically seen as dedicated brain devices that subserve domain-specific cognitive functions and that can be selectively activated, or impaired.

4. At the *genetic level,* what is at stake are the pleiotropic effects among genes such that relatively autonomous "gene nets" (Bonner, 1988) get expressed as distinct phenotypic modules. Genetic modularity is more and more seen as crucial to explaining, on the one hand, phenotypic modularity, and on the other the evolution of specific modules (Wagner, 1995, 1996; Wagner and Altenberg, 1996).

5. At the *evolutionary level,* hypotheses are being developed about the causes of the evolution of specific modules, and of genetic modularity in general. Understanding the causes of the evolution of modules helps explain the known features of known modules and also aids the search for yet-to-be discovered features and modules.

   In cognitive science, discussions of mental modularity received their initial impetus and much of their agenda from Fodor's pioneering work (Fodor, 1983). Fodor's take on modularity had two peculiar features. First, while he was brilliantly defending a modularist view of input systems, Fodor was also—and this initially attracted less attention—decidedly antimodularist regarding higher cognitive processes. Incidentally, whereas his arguments in favor of input systems modularity relied heavily on empirical evidence, his arguments against modularity of central processes were mostly philosophical, and still are. Second, Fodor focused on the architectural level, paid some attention to the developmental and neurological levels, and had almost nothing to say about the genetic and evolutionary levels. Yet the very existence of modularity and of specific modules begs for an evolutionary explanation (and raises difficult and important issues at the genetic level). This is uncontroversial in the case of nonpsychological modular components of the organisms, for ex-

ample, the liver, eyes, endocrine glands, muscles, or enzyme systems, which are generally best understood as adaptations (how well developed, convincing, and illuminating are available evolutionary hypotheses varies, of course, with cases). When, however, an evolutionary perspective on psychological modularity was put forward by evolutionary psychologists (Barkow, Cosmides, and Tooby, 1992; Cosmides and Tooby, 1987, 1994; Pinker, 1997; Plotkin, 1997; Sperber, 1994), this was seen by many as wild speculation.

I will refrain from discussing Fodor's arguments against the evolutionary approach. They are, anyhow, mostly aimed at grand programmatic claims and questionable illustrations that evolutionary psychologists have sometimes been guilty of (just as defenders of other ambitious and popular approaches). As I see it, looking from an evolutionary point of view at the structure of organisms, and in particular at the structure of the mind of organisms, is plain ordinary science and does not deserve to be attacked or need to be defended. Like ordinary science in general, it can be done well or badly, and it sometimes produces illuminating results and sometimes disappointing ones.

There is, moreover, a reason why the evolutionary perspective is especially relevant to psychology, and in particular to the study of cognitive architecture. Apart from input and output systems, which, being linked to sensory and motor organs, are relatively discernible, there is nothing obvious about the organization of the mind into parts and subparts. Therefore, all sources of insight and evidence are welcome. The evolutionary perspective is one such valuable source, and I cannot imagine why we should deprive ourselves of it. In particular, it puts the issue of modularity in an appropriate wider framework. To quote a theoretical biologist, "The fact that the morphological phenotype can be decomposed into basic organizational units, the homologues of comparative anatomy, has . . . been explained in terms of modularity. . . . The biological significance of these semi-autonomous units is their possible role as adaptive 'building blocks' " (Wagner, 1995). In psychology, this suggests that the two notions of a mental module and of a psychological adaptation (in the biological sense), though definitely not synonymous or coextensive, are nevertheless likely to be closely related. Autonomous mental mechanisms that are also very plausible cases of evolved adapta-

tions—face recognition or mind reading, for instance—are prime examples of plausible modularity.

Fodor is understandably reluctant to characterize a module merely as a "functionally individuated cognitive mechanism," since "anything that would have a proprietary box in a psychologist's information flow diagram" would thereby "count as a module" (Fodor 2000, p. 56). If, together with being a distinct mechanism, being plausibly a distinct adaptation with its own evolutionary history were used as a criterion, then modularity would not be so trivial. However, Fodor shuns the evolutionary perspective and resorts, rather, to a series of criteria in his 1983 book, only one of which, "informational encapsulation," is invoked in his new book.

The basic idea is that a device is informationally encapsulated if it has access only to limited information, excluding some information that might be pertinent to its producing the right outputs and that might be available elsewhere in the organism. Paradigm examples are provided by perceptual illusions: *I* have the information that the two lines in the Müller-Lyer illusion are equal, but my visual perceptual device has no access to this information and keeps "seeing" them as unequal. Reflexes are in this respect extreme cases of encapsulation: given the proper input, they immediately deliver their characteristic output, whatever the evidence as to its appropriateness in the context. The problem with the criterion of encapsulation is that it seems too easy to satisfy. In fact, it is hard to think of any autonomous mental device that would have unrestricted access to all the information available in the wider system.

To clarify the discussion, let us sharply distinguish the informational inputs to a cognitive device (i.e., the representations it processes and to which it associates an output) from its database, that is, the information it can freely access in order to process its inputs. For instance, a word-recognition device takes as characteristic inputs phonetic representations of speech and uses as a database a dictionary. Nonencapsulated devices, if there are any, use the whole mental encyclopedia as their database. Encapsulated devices have a restricted database of greater or lesser size. A reflex typically has no database to speak of.

In his new book, Fodor attempts to properly restrict the notion of encapsulation. For this, he discusses the case of modus ponens (Fodor,

2000, pp. 60–62). Imagine an autonomous mental device that takes as input any pairs of beliefs of the form {$p$, [If $p$, then $q$]} and produces as output the belief that $q$. I would view such a modus ponens device as a cognitive reflex, and therefore a perfect example of a module. In particular, as a module would, it ignores information that a less dumb device would take into account: suppose that the larger system has otherwise information that $q$ is certainly false. Then the smarter device, instead of adding $q$ to the belief box, would consider erasing from the belief box one of the premises, $p$, or [If $p$, then $q$]. But our device has no way of using this extra information and of adjusting its output accordingly.

Still, Fodor would have this modus ponens device count as unencapsulated, and therefore as nonmodular. Why? Modus ponens, he argues, applies to pairs of premises in virtue of their logical form and is otherwise indifferent to their informational content. An organism with a modus ponens device can use it across the board. Compare it with, say, a bridled modus ponens device that would apply to reasoning about number, but not about food, people, or plants, in fact about nothing other than numbers. According to Fodor, this latter device would be encapsulated. Yet, surely, the logical form of a representation contributes to its informational content. [If $p$, then $q$] does not have the same informational content as [$p$ and $q$] or [$p$ or $q$], even though they differ only in terms of logical form and their nonlogical content is otherwise the same. Moreover, the difference between the wholly general and the number-specific modus ponens devices is one of inputs, not one of database. Both are cognitive reflexes and have no database at all (unless you want to consider the general or restricted modus ponens rule as a piece of data, the only one then in the database).

The logic of Fodor's argument is unclear, but its motivation is not. Fodor's main argument against massive modularity is that modules, given their processing and informational restrictions, could not possibly perform the kind of general reasoning tasks that human minds perform all the time, drawing freely, or so it seems, on all the information available. An unrestricted modus ponens device looks too much like a tool for general reasoning, and that is probably why it had better not be modular. Still, this example shows, if anything, that a sensible definition of informational encapsulation that makes it a relatively rare property of mental devices is not so easily devised.

Naively, we see our minds—or simply ourselves—as doing the thinking and as having unrestricted access to all our stored information (barring memory problems or Freudian censorship). There is in this respect a change of gestalt when passing from naive psychology to cognitive science. In the information-processing boxological pictures of the mind, there is no one box where the thinking is done and where information is freely accessible. Typically, each box does its limited job with limited access to information. So who or what does the thinking ? An optimist would say, "It is the network of boxes that thinks (having, of course, access to all information in the system), and not any one box in particular." But Fodor is not exactly an optimist, and, in this case, he has arguments to the effect that a system of boxes cannot be the thinker, and that therefore the boxological picture of the mind cannot be correct.

Our best theory of the mind, Fodor argues, is the Turing-inspired computational theory according to which mental processes are operations defined on the syntax of mental representations. However such computations are irredeemably local, and cannot take into account contextual considerations. Yet, our best thinking (that of scientists) and even our more modest everyday thinking are highly sensitive to context. Fodor suggests various ways in which the context might be taken into consideration in syntactic processes, and shows that they fall short, by a wide margin, of delivering the kind of context-sensitivity that is required. He assumes that, if the computational theory of mind is correct, and if, therefore, there are only local operations, global contextual factors cannot weight on inference, and in particular cannot contribute to its rationality and creativity. Therefore the computational theory of mind is flawed.

Has Fodor really eliminated all possibilities? Here is one line that he has not explored. Adopt a strong modularist view of the mind, assume that all the modules that have access to some possible input are ready to produce the corresponding output, but assume also that each such process takes resources, and that there are not enough resources for all processes to take place. All these potentially active modules are like competitors for resources. It is easy to see that different allocations of resources will have different consequences for the cognitive and epistemic efficiency of the system as a whole.

I wake and see through the window that the street is wet. I normally infer that it has been raining. For the sake of simplicity, let us think of this as a modus ponens deduction with [if the street is wet, it has been raining] and [the street is wet] as premises. This is a process readily described in syntactic terms. Of course, the major premise of this deduction is true only by default. On special occasions, the street happens to be wet because, say, it has just been washed. If I remember that the street is washed every first day of the month and that today is such a day, I will suspend my acceptance of the major premise [if the street is wet, it has been raining] and not perform the default inference. Again, this process is local enough and easily described in syntactic terms. Whether or not I remember this relevant bit of information and make or suspend the inference clearly affects the epistemic success of my thinking.

Does there have to be a higher-order computational process that triggers my remembering the day and suspending the default inference? Of course not. The allocation of resources among mental devices can be done in a variety of non computational ways without compromising the computational character of the devices. Saliency is an obvious possible factor. For instance the premise [the street is washed on the first day of the month] may be more salient when both the information that it is the first day of the month, and that the street is wet are activated. A device that accepts this salient premise as input is thereby more likely to receive sufficient processing resources.

It is not hard to imagine how the mere use of saliency for the allocation of computing resources might improve the cognitive and epistemic efficiency of the system as a whole, not by changing local processes but by triggering the more appropriate ones in the context. The overall inferential performance of the mind would then exhibit some significant degree of context-sensitivity without any of the computational processes involved being themselves context-sensitive. Saliency is an obvious and obviously crude possibility. Deirdre Wilson and I have suggested a subtler, complex, noncomputational factor, relevance, with two subfactors, mental effort and cognitive effect (Sperber and Wilson, 1995, 1996). Relevance as we characterize it would in particular favor simplicity and conservatism in inference, two properties that Fodor argues cannot be

accommodated in the classical framework. This is not the place to elaborate. The general point is that a solution to the problems raised for a computational theory of mind by context-sensitive inference may be found in terms of some "invisible hand" cumulative effect of noncomputational events in the mind/brain, and that Fodor has not even discussed, let alone ruled out, this line of investigation.

Of course, the fact that a vague line of investigation has not been ruled out is no great commendation. It would not, for instance, impress a grant-giving agency. But this is where, unexpectedly, Fodor comes to the rescue. Before advancing his objections to what he calls the "New Synthesis," Fodor presents it better than anybody else has, and waxes lyrical (by Fodorian standards) about it. I quote at some length:

> Turing's idea that mental processes are computations . . . , together with Chomsky's idea that poverty of the stimulus arguments set a lower bound to the information a mind must have innately, are half of the New Synthesis. The rest is the "massive modularity" thesis and the claim that cognitive architecture is a Darwinian adaptation. . . . [T]here are some very deep problems with viewing cognition as computational, but . . . these problems emerge primarily in respect to mental problems that *aren't* modular. The real appeal of the massive modularity thesis is that, if it is true, we can either solve these problems, or at least contrive to deny them center stage pro tem. The bad news is that, since massive modularity thesis pretty clearly *isn't* true, we're sooner or later going to have to face up to the dire inadequacies of the only remotely plausible theory of the cognitive mind that we've got so far. (Fodor, 2000, p. 23)

True, Fodor does offer other arguments against massive modularity, but rejoinders to these will be for another day (anyhow, these arguments were preempted in my 1994 paper, I believe). However, the crucial argument against computational theory and modularity is that it cannot be reconciled with the obvious abductive capabilities of the human mind, and I hope to have shown that Fodor's case here is not all that airtight. Now, given Fodor's praises for the new synthesis and his claim that this is "the only remotely plausible theory" we have, what should our reaction be? Yes, we might—we should—be worried about the problems he raises. However, rather than giving up in despair, we should decidedly explore any line of investigation still wide open.

As a defender of massive modularity, I am grateful to Fodor for giving us new food for thought and new arguments. I wish to express also my

gratitude to Jacques without whose demanding encouragement, in spite of my anthropological training, I would never have ventured so far among the modules.

## References

Barkow, J., Cosmides, L., and Tooby, J. (Eds.) (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Bonner, J. T. (1988). *The evolution of complexity*. Princeton, NJ: Princeton University Press.

Cosmides, L., and Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality*. Cambridge MA: MIT Press.

Cosmides, L., and Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld and S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). New York: Cambridge University Press.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In J. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 26–36). Cambridge, MA: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge MA: MIT Press.

Hirschfeld, L. A., and Gelman, S. A. (Eds.) (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Murphy, D., and Stich, S. (2000). Darwin in the madhouse: Evolutionary psychology and the classification of mental disorders. In P. Carruthers and A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language and metacognition*, Cambridge, UK: Cambridge University Press.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Plotkin, H. (1997). *Evolution in mind*. London: Alan Lane.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *British Journal for the Philosophy of Science, 49*, 575–602.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld and S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39–67). New York: Cambridge University Press. [Revised version in Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford, Blackwell.]

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and cognition,* 2nd ed. Oxford: Blackwell.

Sperber, D., and Wilson, D. (1996). Fodor's frame problem and relevance theory. *Behavioral and Brain Sciences, 19,* 530–532

Wagner, G.P. (1995). Adaptation and the modular design of organisms. In F. Morán, A. Morán, J. J. Merelo, and P. Chacón (Eds.), *Advances in artificial life* (pp. 317–328). Berlin: Springer Verlag.

Wagner, G. P. (1996). Homologues, natural kinds and the evolution of modularity. *American Zoologist, 36,* 36–43.

Wagner, G. P. and Altenberg, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution, 50,* 967–976.