2 The modularity of thought and the epidemiology of representations

Dan Sperber

Ten years ago, Jerry Fodor published *The Modularity of Mind*, a book that received much well-deserved attention. His target was the then-dominant view according to which there are no important discontinuities between perceptual processes and conceptual processes. Information flows freely, "up" and "down," between these two kinds of processes, and beliefs inform perception as much as they are informed by it. Against this view, Fodor argued that perceptual processes (and also linguistic decoding) are carried out by specialized, rather rigid mechanisms. These "modules" each have their own proprietary data base, and do not draw on information produced by conceptual processes.

Although this was probably not intended and has not been much noticed, "modularity of mind" was a paradoxical title, for, according to Fodor, modularity is to be found only at the periphery of the mind, in its input systems.[1] In its center and bulk, Fodor's mind is decidedly *non*modular. Conceptual processes – that is, thought proper – are presented as a big holistic lump lacking joints at which to carve. Controversies have focused on the thesis that perceptual and linguistic decoding processes are modular, much more than on the alleged nonmodularity of thought.[2]

In this chapter, I have two aims. The first is to defend the view that thought processes might be modular too (what Fodor [1987: 27] calls "modularity theory gone mad" – oh well!). Let me however echo Fodor and say that, "when I speak of a cognitive system as modular, I shall . . . always mean 'to some interesting extent' " (Fodor, 1983: 37). My second aim is to articulate a modular view of human thought with the naturalistic view of human culture that I have been developing under the label "epidemiology of representations" (Sperber, 1985b). These aims are closely related: Cultural diversity has always been taken to show how plastic the human mind is, whereas the

I thank Lawrence Hirschfeld, Pierre Jacob, and Deirdre Wilson for their useful comments on an earlier version of this paper.

modularity of thought thesis seems to deny that plasticity. I want to show how, contrary to the received view, organisms endowed with truly modular minds might engender truly diverse cultures.

Two commonsense arguments against the modularity of thought

Abstractly and roughly at least, the distinction between perceptual and conceptual processes is clear: Perceptual processes have, as input, information provided by sensory receptors and, as output, a conceptual representation categorizing the object perceived. Conceptual processes have conceptual representations both as input and as output. Thus seeing a cloud and thinking "here is a cloud" is a perceptual process. Inferring from this perception "it might rain" is a conceptual process.

The rough idea of modularity is also clear: A cognitive module is a genetically specified computational device in the mind/brain (henceforth: the mind) that works pretty much on its own on inputs pertaining to some specific cognitive domain and provided by other parts of the nervous systems (e.g., sensory receptors or other modules). Given such notions, the view that perceptual processes might be modular is indeed quite plausible, as argued by Fodor. On the other hand, there are two main commonsense arguments (and several more technical ones) that lead one to expect conceptual thought processes not to be modular.

The first commonsense argument against the modularity of thought has to do with integration of information. The conceptual level is the level at which information from different input modules, each presumably linked to some sensory modality, gets integrated into a modality-independent medium: A dog can be seen, heard, smelled, touched, and talked about: The percepts are different; the concept is the same. As Fodor points out,

> the general form of the argument goes back at least to Aristotle: the representations that input systems deliver have to interface somewhere, and the computational mechanisms that affect the interface must ipso facto have access to information from more than one cognitive domain. (Fodor, 1983: 101–102)

The second commonsense argument against the modularity of thought has to do with cultural diversity and novelty. An adult human's conceptual processes range over an indefinite variety of domains, including party politics, baseball history, motorcycle maintenance, Zen Bhuddism, French cuisine, Italian opera, chess playing, stamp collecting, and Fodor's chosen example, modern science. The appearance of many of these domains in human cognition is very recent and not relevantly correlated with changes in the human genome. Many of these domains vary dramatically in content from one culture to another, or are not found at all in many cultures. In such conditions, it would be absurd to assume that there is an ad hoc genetically specified preparedness for these culturally developed conceptual domains.

These two commonsense arguments are so compelling that Fodor's more technical considerations (having to do with "isotropy," illusions, rationality, etc.) look like mere nails in the coffin of a dead idea. My goal will be to shake the commonsense picture and to suggest that the challenge of articulating conceptual integration, cultural diversity, and modularity may be met and turns out to be a source of psychological and anthropological insights.

Notice, to begin with, that both the informational integration argument and the cultural diversity argument are quite compatible with *partial* modularity at the conceptual level.

True, it would be functionally self-defeating to reproduce at the conceptual level the same domain partition found at the perceptual level, and have a different conceptual module treat separately the output of each perceptual module. No integration whatsoever would take place, and the dog seen and the dog heard could never be one and the very same mastiff Goliath. But who says conceptual domains have to match perceptual domains? Why not envisage, at the conceptual level, a wholly different, more or less orthogonal domain partition, with domain-specific conceptual mechanisms, each getting their inputs from several input mechanisms? For instance, all the conceptual outputs of perceptual modules that contain the concept MASTIFF might be fed into a specialized module (say a domain-specific inferential device handling living-kind concepts), which takes care (inter alia) of Goliath qua mastiff. Similarly, all the conceptual outputs of input modules that contain the concept THREE might be fed into a specialized module, which handles inference about numbers, and so forth. In this way, information from different input devices might get genuinely integrated, though not into a single conceptual system, but into several such systems.

Of course, if you have, say, a prudential rule that tells you to run away when you encounter more than two bellicose dogs, you would not really be satisfied to be informed by the living-kinds module that the category BELLICOSE DOG is instantiated in your environment, and by the numerical module that there are more than two of something. Some further, at least partial, integration had better take place. It might even be argued – though *that* is by no means obvious – that a plausible model of human cognition should allow for *full* integration of all conceptual information at some level. Either way, partial or full integration might take place further up the line, among the outputs of conceptual rather than of perceptual modules. Conceptual integration is not incompatible with at least some conceptual modularity.

Similarly, the conceptual diversity argument implies that some conceptual domains (chess, etc.) could not be modular. It certainly does not imply that none of them could be. Thus, in spite of superficial variations, living-kind classification exhibits strong commonalities across cultures (see Berlin, 1978) in a manner that does suggest the presence of a domain-specific cognitive module (see Atran, 1987, 1990).

The thesis that some central thought processes might be modular gets

support from a wealth of recent work (well illustrated in the present volume) tending to show that many basic conceptual thought processes, found in every culture and in every fully developed human, are governed by domain-specific competences. For instance, it is argued that people's ordinary understanding of the movements of an inert solid object, of the appearance of an organism, or of the actions of a person are based on three distinct mental mechanisms: a naive physics, a naive biology, and a naive psychology (see for instance Atran, 1987; Keil, 1989; Leslie, 1987, 1988; Spelke, 1988, and their contributions to this volume – see Carey, 1985, for a dissenting view). It is argued moreover that these mechanisms, at least in rudimentary form, are part of the equipment that makes acquisition of knowledge possible, rather than being acquired competences.

Accepting as a possibility some degree of modularity in conceptual systems is innocuous enough. Jerry Fodor himself recently considered favorably the view that "intentional folk psychology is, essentially, an innate, *modularized* database" (Fodor, 1992: 284 – italics added) without suggesting that he was thereby departing from his former views on modularity. But what about the possibility of *massive* modularity at the conceptual level? Do the two commonsense arguments, integration and diversity, really rule it out?

Modularity and evolution

If modularity is a genuine natural property, then what it consists of is a matter of discovery, not stipulation. Fodor himself discusses a number of characteristic and diagnostic features of modularity. Modules, he argues, are "domain-specific, innately specified, hardwired, autonomous" (1983: 36). Their operations are mandatory (p. 52) and fast (p. 61); they are "informationally encapsulated" (p. 64), that is, the only background information available to them is that found in their proprietary data base. Modules are "associated with fixed neural architecture" (p. 98). Fodor discusses still other features that are not essential to the present discussion.

There is one feature of modularity that is implied by Fodor's description, but that he does not mention or discuss. If, as Fodor argues, a module is innately specified, hardwired, and autonomous, then it follows that *a cognitive module is an evolved mechanism with a distinct phylogenetic history*. This is a characteristic, but hardly a diagnostic feature, because we know close to nothing about the actual evolution of cognitive modules. But I have been convinced by Leda Cosmides and John Tooby (see Cosmides, 1989; Cosmides & Tooby, 1987; Tooby & Cosmides, 1989, 1992, this volume)[3] that we know enough about evolution on the one hand and cognition on the other to come up with well-motivated (though, of course, tentative) assumptions as to when to expect modularity, what properties to expect of modules, and even what modules to expect. This section of the chapter owes much to their ideas.

Fodor himself does mention evolutionary considerations, but only in passing.

He maintains that, phylogenetically, modular input systems should have preceded nonmodular central systems:

> Cognitive evolution would thus have been in the direction of gradually freeing certain sorts of problem-solving systems from the constraints under which input analyzers labor – hence of producing, as a relatively late achievement, the comparatively domain-free inferential capacities which apparently mediate the higher flights of cognition. (Fodor, 1983: 43)

Let us spell out some of the implications of Fodor's evolutionary suggestion. At an early stage of cognitive evolution we should find modular sensory input analyzers directly connected to modular motor controllers. There is no level yet where information from several perceptual processes would be integrated by a conceptual process. Then there emerges a conceptual device, that is, an inferential device that is not itself directly linked to sensory receptors. This conceptual device accepts input from two or more perceptual devices, constructs new representations warranted by these inputs, and transmits information to motor control mechanisms.

Initially, of course, this conceptual device is just another module: It is specialized, innately wired, fast, automatic, and so forth. Then, so the story should go, it grows and becomes less specialized, possibly it merges with other similar conceptual devices, to the point where it is a single big conceptual system, able to process all the outputs of all the perceptual modules, and able to manage all the conceptual information available to the organism. This true central system cannot, in performing a given cognitive task, activate all the data accessible to it, or exploit all of its many procedures. Automaticity and speed are no longer possible. Indeed, if the central system automatically did what it is capable of doing, this would trigger a computational explosion with no end in sight.[4]

An evolutionary account of the emergence of a conceptual module in a mind that had known only perceptual processes is simple enough to imagine. Its demodularization would be much harder to explain.

A toy example might go like this: Organisms of a certain species, call them "protorgs," are threatened by a danger of a certain kind. This danger (the approach of elephants that might trample the orgs, as it might be) is signaled by the co-occurrence of a noise N and soil vibrations V. Protorgs have an acoustic perception module that detects instances of N and a vibration-perception module that detects instances of V. The detection either of N by one perceptual module, or of V by the other activates an appropriate flight procedure. Fine, except that when N occurs alone, or when V occurs alone, it so happens that there is no danger. So protorgs end up with a lot of "false positives," uselessly running away, and thus wasting energy and resources.

Some descendants of the protorgs, call them "orgs," have evolved another mental device: a conceptual inference mechanism. The perceptual modules no longer directly activate their flight procedure. Rather their relevant outputs,

that is, the identification of noise N and that of vibrations V, go to the new device. This conceptual mechanism acts essentially as an AND-gate: When, and only when both N and V have been perceptually identified, does the conceptual mechanism get into a state that can be said to represent the presence of danger, and it is this state that activates the appropriate flight procedure.

Orgs, so the story goes, competed successfully with protorgs for food resources, and that is why you won't find protorgs around.

The orgs' conceptual mechanism, though not an *input* module, is nevertheless a clear case of a module: It is a domain-specific problem solver; it is fast, informationally encapsulated, associated with fixed neural architecture, and so forth. Of course, it is a tiny module, but nothing stops us from imagining it becoming larger: Instead of accepting just two bits of information from two simple perceptual modules, the conceptual module could come to handle more from more sources, and to control more than a single motor procedure, but still be domain-specific, automatic, fast, and so on.

At this juncture, we have two diverging evolutionary scenarios on offer. According to the scenario suggested by Fodor, the conceptual module should evolve toward less domain specificity, less informational encapsulation, less speed, and so on. In other words, it would become less and less modular, possibly merge with other demodularized devices, and end up like the kind of central system with which Fodor believes we are endowed ("Quineian," "isotropic," etc.). There are two gaps in this scenario. The first gap has to do with mental mechanisms and is highlighted by Fodor himself in his "First Law of the Nonexistence of Cognitive Science." This law says in substance that the mechanisms of nonmodular thought processes are too complex to be understood. So, just accept that there are such mechanisms and don't ask how they work.

The second gap in Fodor's scenario has to do with the evolutionary process itself that is supposed to bring about the development of such a mysterious mechanism. No doubt, it might be advantageous to trade a few domain-specific inferential micromodules for an advanced all-purpose macrointelligence, if there is any such thing. For instance, superorgs endowed with general intelligence might develop technologies to eradicate the danger once and for all instead of having to flee again and again. But evolution does not offer such starkly contrasted choices. The available alternatives at any one time are all small departures from the existing state. Selection, the main force driving evolution, is near-sighted (whereas the other forces, genetic drift, etc., are blind). An immediately advantageous alternative is likely to be selected from the narrow available range, and this may bar the path to highly advantageous long-term outcomes. A demodularization scenario is implausible for this very reason.

Suppose indeed the conceptual danger analyzer is modified in some mutant orgs, not in the direction of performing better at its special task, but in

that of less domain specificity. The modified conceptual device processes not just information relevant to the orgs' immediate chances of escape, but also information about innocuous features of the dangerous situation, and about a variety of innocuous situations exhibiting these further features; the device draws inferences not just of an urgent practical kind, but also of a more theoretical character. When danger is detected, the new, less modular system does not automatically trigger flight behavior, and when it does, it does so more slowly – automaticity and speed go with modularity – but it has interesting thoughts that are filed in memory for the future . . . if there is any future for mutant orgs endowed with this partly demodularized device.

Of course, speed and automaticity are particularly important for danger analyzers, and less so for other plausible modules, for instance, modules governing the choice of sexual partners. However, the general point remains: Evolved cognitive modules are likely to be answers to specific, usually environmental problems. Loosening the domain of a module will bring about, not greater flexibility, but greater slack in the organism's response to the problem. To the extent that evolution goes toward improving a species' biological endowments, then we should generally expect improvements in the manner in which existing modules perform their task, emergence of new modules to handle other problems, but not demodularization.

True, it is possible to conceive of situations in which the marginal demodularization of a conceptual device might be advantageous, or at least not detrimental, in spite of the loss of speed and reliability involved. Imagine, for instance, that the danger the conceptual module was initially selected to analyze has vanished from the environment; then the module is not adapted any more and a despecialization would do no harm. On the other hand why should it do any good? Such odd possibilities fall quite short of suggesting a positive account of the manner in which, to repeat Fodor's words, "cognitive evolution would . . . have been in the direction of gradually freeing certain sorts of problem-solving systems from the constraints under which input analyzers labor." It is not that this claim could not be right, but it is poorly supported. In fact the only motivation for it seems to be the wish to integrate the belief that human thought processes are nonmodular in some evolutionary perspective, however vague. Better officialize the explanatory gap with a "Second Law of the Nonexistence of Cognitive Science," according to which the forces that have driven cognitive evolution can never be identified.[5] Just accept that cognitive evolution occurred (and resulted in the demodularization of thought) and don't ask how.

Instead of starting from an avowedly enigmatic view of homo sapiens's thought processes and concluding that their past evolution is an unfathomable mystery, one might start from evolutionary considerations plausible in their own right and wonder what kind of cognitive organization these might lead one to expect in a species of which we know that it relies heavily on its cognitive abilities for its survival. This yields our second scenario.

As already suggested, it is reasonable to expect conceptual modules to gain in complexity, fine-grainedness, and inferential sophistication *in the performance of their function.* As with any biological device, the function of a module may vary over time, but there is no reason to expect new functions to be systematically more general than old ones. It is reasonable, on the other hand, to expect new conceptual modules to appear in response to different kinds of problems or opportunities. Thus more and more modules should accumulate.

Because cognitive modules are each the result of a different phylogenetic history, there is no reason to expect them all to be built on the same general pattern and elegantly interconnected. Though most if not all conceptual modules are inferential devices, the inferential procedures that they use may be quite diverse. Therefore, from a modular point of view, it is unreasonable to ask what is the general form of human inference (logical rules, pragmatic schemas, mental models, etc.) as is generally done in the literature on human reasoning (see Manktelow & Over, 1990, for a recent review).

The "domains" of modules may vary in character and in size: There is no reason to expect domain-specific modules to handle each a domain of comparable size. In particular there is no reason to exclude micromodules the domain of which is the size of a concept rather than that of a semantic field. In fact, I will argue that many human concepts are individually modular. Because conceptual modules are likely to be many, their interconnections and their connections with perceptual and motor control modules may be quite diverse too. As argued by Andy Clark (1987, 1990), we had better think of the mind as kludge, with sundry bits and components added at different times, and interconnected in ways that would make an engineer cringe.

Modularity and conceptual integration

The input to the first conceptual modules to have appeared in cognitive evolution must have come from the perceptual modules. However, once some conceptual modules were in place, their output could serve as input to other conceptual modules.

Suppose the orgs can communicate among themselves by means of a small repertoire of vocal signals. Suppose further that the optimal interpretation of some of these signals is sensitive to contextual factors. For instance, an ambiguous danger signal indicates the presence of a snake when emitted by an org on a tree, and approaching elephants when emitted by an org on the ground. Identifying the signals and the relevant contextual information is done by perceptual modules. The relevant output of these perceptual modules is processed by an ad hoc conceptual module that interprets the ambiguous signals. Now, it would be a significant improvement if the conceptual module specialized in inferring the approach of elephants would accept as input not only perceptual information on specific noises and soil vibrations

but also interpretations of the relevant signals emitted by other orgs. If so, this danger-inferring conceptual module would receive input not just from perceptual modules but also from another conceptual module, the context-sensitive signal interpreter.

In the human case, it is generally taken for granted that domain-specific abilities can handle not just primary information belonging to their domain and provided by perception but also verbally or picturally communicated information. Thus experiments on the development of zoological knowledge use as material, not actual animals, but pictures or verbal descriptions. Though this practice deserves more discussion than it usually gets, it may well be sound. If so, its being sound is itself quite remarkable.

Then too, some conceptual modules might get *all* of their input from other conceptual modules. Imagine for instance that an org emits a danger signal only when two conditions are fulfilled: It must have inferred the presence of a danger on the one hand, and that of friendly orgs at risk on the other hand. Both inferences are performed by conceptual modules. If so, then the conceptual module that decides whether or not to emit the danger signal gets all of its input from other conceptual modules, and none from perceptual ones.

We are now envisaging a complex network of conceptual modules: Some conceptual modules get all of their input from perceptual modules, other modules get at least some of their input from conceptual modules, and so forth. Every information may get combined with many others across or within levels and in various ways (though overall conceptual integration seems excluded). What would be the behavior of an organism endowed with such complex modular thought processes? Surely, we don't know. Would it behave in a flexible manner like humans do? Its responses could at least be extremely fine-grained. Is there more to flexibility than this fine-grainedness? "Flexibility" is a metaphor without a clear literal interpretation, and therefore it is hard to tell. Still, when we think of flexibility in the human case, we particularly have in mind the ability to learn from experience. Can a fully modular system learn?

Imprinting is a very simple form of modular learning. What, for instance, do orgs know about one another? If orgs are nonlearning animals, they might be merely endowed with a conspecific detector and detectors for some properties of other orgs such as sex or age, but they might otherwise be unable to detect any single individual as such, not even, say, their own mothers. Or, if they are very primitive learners, they might have a mother-detector module that will be "initialized" (i.e., have its parameters fixed or its empty slots filled) once and for all by the newborn org's first encounter with a large moving creature in its immediate vicinity (hopefully its real mum). As a result of this encounter the initialized module becomes a detector for the particular individual who caused the imprinting.

If they are slightly more sophisticated learners, orgs may have the capacity to construct several detectors for different individual conspecifics. They might

have a template module quite similar to a mother-detector, except that it can be "initialized" several times, each time projecting a differently initialized copy of itself that is specialized for the identification of a different individual. Would the initialized copies of the template module be modules too? I don't see why not. The only major difference is that these numerous projected modules seem less likely to be hardwired than a single mother-detector module.[6] Otherwise, both kinds of modules get initialized and operate in exactly the same manner. Of our more sophisticated orgs, we would want to say, then, that they had a modular domain-specific ability to represent mentally conspecific individuals, an ability resulting in the generation of micromodules for each represented individual.

Consider in this light the human domain-specific ability to categorize living kinds. One possibility is that there is an initial template module for living-kind concepts that gets initialized many times, producing each time a new micromodule corresponding to one living-kind concept (the dog module, the cat module, the goldfish module, etc.).

Thinking of such concepts as modules may take some getting used to, I admit. Let me help: Concepts are domain-specific (obviously), they have a proprietary data-basis (the encyclopedic information filed under the concept), and they are autonomous computational devices (they work, I will argue, on representations in which the right concept occurs, just as digestive enzymes work on food in which the right molecule occurs). When, on top of all that, concepts are partly genetically specified (via some domain-specific conceptual template), they are modular at least to some interesting extent, no?

The template–copy relationship might sometimes involve more levels. A general living-kind-categorization metatemplate could project, not directly concepts, but other, more specific templates for different domains of living kinds. For instance, a fundamental parameter to be fixed might concern the contrast between self-propelled and non-self-propelled objects (Premack, 1990), yielding two templates, one for zoological concepts and another one for botanical concepts.

Another possibility still is that the initial metatemplate has three types of features: (1) fixed features that characterize living kinds in general, for instance, it might be an unalterable part of any living-kind concept that the kind is taken to have an underlying essence (Atran, 1987; Gelman & Coley, 1991; Gelman & Markman, 1986, 1987; Keil, 1989; Medin & Ortony, 1989); (2) parameters with default values that can be altered in copies of the template, for instance, "self-propelled" and "non-human" might be revisable features of the initial template; (3) empty slots for information about individual kinds. If so, then, the default-value template could serve as such for nonhuman animal concepts. To use the template for plant concepts, or to include humans in a taxonomy of animals would involve changing a default value of the initial template.

How is the flow of information among modules actually governed? Is there

a regulating device? Is it a pandemonium? A market economy? Many types of models can be entertained. Here is a simple possibility.

The output of perceptual and conceptual modules is in the form of conceptual representations. Perceptual modules categorize distal stimuli and must each have therefore the conceptual repertoire needed for the output categorizations of which they are capable. Conceptual modules may infer new output categorizations from the input conceptual representations they process; they must have an input and an output conceptual repertoire to do so. Let us assume that modules accept as input any conceptual representation in which a concept belonging to their input repertoire occurs. In particular single-concept micromodules process all and only representations where their very own concept occurs. These micromodules generate transformations of the input representation by replacing the concept with some inferentially warranted expansion of it. They are otherwise blind to the other conceptual properties of the representations they process (in the manner of the "calculate" procedure in some word processor, which scans the text but "sees" only numbers and mathematical signs). Generally, the presence of specific concepts in a representation determines what modules will be activated and what inferential processes will take place (see Sperber & Wilson, 1986, chap. 2).

A key feature of modularity in Fodor's description is informational encapsulation: A full-fledged module uses a limited data base and is not able to take advantage of information relevant to its task if that information is in some other data base. Central processes on the other hand are not so constrained: They are characterized, on the contrary, by free flow of information. Thus beliefs about Camembert cheese might play a role in forming conclusions about quarks, even though they hardly belong to the same conceptual domain. This is a fact, and I wouldn't dream of denying it. What does it imply regarding the modularity of conceptual processes? It implies that one particular modular picture cannot be right: Imagine a single layer of a few large mutually unconnected modules; then an information treated by one module won't find its way to another. If, on the other hand, the output of one conceptual module can serve as input to another one, modules can each be informationally encapsulated while chains of inference can take a conceptual premise from one module to the next and therefore integrate the contribution of each in some final conclusion. A holistic effect need not be the outcome of a holistic procedure.

Once a certain level of complexity in modular conceptual thought is reached, modules can emerge whose function it is to handle problems raised, not externally by the environment, but internally by the workings of the mind itself. One problem that a rich modular system of the kind we are envisaging would encounter as surely as Fodor's nonmodular central processes is the risk of computational explosion.

Assume that a device would have emerged, the function of which is to put up on the board, so to speak, some limited information for actual processing.

Call this device "attention." Think of it as a temporary buffer. Only representations stored in that buffer are processed (by the modules whose input conditions they satisfy), and they are processed only as long as they stay in the buffer. There is, so to speak, competition among representations for attention. The competition tends to work out so as to maximize cognitive efficiency, that is, it tends to select for a place in the buffer, and thus for inferential processing, the most relevant information available at the time. There is a much longer story to be told: read *Relevance* (Sperber & Wilson, 1986).

Attention is of course not domain-specific. On the other hand it is a clear adaptation to an internal processing problem: the problem encountered by any cognitive system able to identify much more information perceptually than it can fully process conceptually. Such a system must be endowed with a means of selecting the information to be conceptually processed. Relevance-guided attention is such a means. Whether or not it should be called a module does not really matter: Attention fits snugly into a modular picture of thought.

I don't expect these speculations to be convincing – I am only half convinced myself, though I will be a bit more by the end of this chapter – but I hope they are intelligible. If so, this means that one can imagine a richly modular conceptual system that integrates information in so many partial ways that it is not obvious any more that we, human beings, genuinely integrate it in any fuller way. The argument against the modularity of thought based on the alleged impossibility of modular integration should lose at least its immediate commonsense appeal.

Actual and proper domains of modules

Modules are domain-specific, and many, possibly most domains of modern human thought are too novel and too variable to be the specific domain of a genetically specified module. This second commonsense argument against the modularity of thought is reinforced by adaptationist considerations: In many domains, cultural expertise is hard to see as a biological adaptation. This is true not just of new domains such as chess, but also of old domains such as music. Expertise in these domains is unlikely therefore to be based on an ad hoc evolved mechanism. Of course, one can always try to concoct some story showing that, say, musical competence is a biological adaptation. However, merely assuming the adaptive character of a trait without a plausible demonstration is an all too typical misuse of the evolutionary approach.

Let me try an altogether different line. An adaptation is, generally, an adaptation to given environmental conditions. If you look at an adaptive feature just by itself, inside the organism, and forget altogether what you know about the environment and its history, you cannot tell what its function is, what it is an adaptation to. The function of a giraffe's long neck is to help

it eat from trees, but in another environment – make it on another planet to free your imagination – the function of an identical body part on an identical organism could be to allow the animal to see farther, or to avoid breathing foul air near the ground, or to fool giant predators into believing that its flesh was poisonous.

A very similar point – or, arguably, a special application of the very same point – has been at the center of major recent debates in the philosophy of language and mind between "individualists" and "externalists." Individualists hold that the content of a concept is in the head of the thinker, or, in other terms, that a conceptual content is an intrinsic property of the thinker's brain state. Externalists maintain – rightly, I believe – that the same brain state that realizes a given concept might realize a different concept in another environment, just as internally identical biological features might have different functions.[7]

The content of a concept is not an intrinsic but a relational property[8] of the neural realizer of that concept, and is contingent upon the environment and the history (including the phylogenetic prehistory) of that neural object. This extends straightforwardly to the case of domain-specific modules. A domain is semantically defined, that is, by a concept under which objects in the domain are supposed to fall. The domain of a module is therefore not a property of its internal structure (whether described in neurological or in computational terms).

There is no way a specialized cognitive module might pick its domain just in virtue of its internal structure, or even in virtue of its connections to other cognitive modules. All that the internal structure provides is, to borrow an apt phrase from Frank Keil (this volume), a *mode of construal*, a disposition to organize information in a certain manner and to perform computations of a certain form. A cognitive module also has structural relations to other mental devices with which it interacts. This determines in particular its *input conditions*: through which other devices the information must come, and how it must be categorized by these other devices. But, as long as one remains within the mind and ignores the connections of perceptual modules with the environment, knowledge of the brain-internal connections of a specialized cognitive module does not determine its domain.

Pace Keil, the fact that the mode of construal afforded by a mental module might fit many domains does *not* make the module any less domain-specific, just as the fact that my key might fit many locks does not make it any less the key to my door. The mode of construal and the domain, just like my key and my lock, have a long common history. How, then, do interactions with the environment over time determine the domain of a cognitive module? To answer this question, we had better distinguish between the *actual* and the *proper* domain of a module.

The *actual domain* of a conceptual module is all the information in the organism's environment that may (once processed by perceptual modules,

and possibly by other conceptual modules) satisfy the module's input conditions. Its *proper domain* is all the information that it is the module's biological function to process. Very roughly, the function of a biological device is a class of effects of that device that contributes to making the device a stable feature of an enduring species. The function of a module is to process a specific range of information in a specific manner. That processing contributes to the reproductive success of the organism. The range of information that it is the function of a module to process constitutes its proper domain. What a module actually processes is information found in its actual domain, whether or not it also belongs to its proper domain.

Back to the orgs. The characteristic danger that initially threatened them was being trampled by elephants. Thanks to a module, the orgs reacted selectively to various signs normally produced, in their environment, by approaching elephants. Of course, approaching elephants were sometimes missed, and other, unrelated and innocuous events did sometimes activate the module. But even though the module failed to pick out all and only approaching elephants, we describe its function as having been to do just that (rather than doing what it actually did). Why? Because it is its relative success at that task that explains its having been a stable feature of an enduring species. Even though they were not exactly coextensive, the actual domain of the module overlapped well enough with the approaching-elephants domain. Only the latter, however, was the proper domain of the module.

Many generations later, elephants had vanished from the orgs' habitat, while hippopotamuses had multiplied, and now *they* trampled absent-minded orgs. The same module that had reacted to most approaching elephants and a few sundry events now reacted to most approaching hippos and a few sundry events. Had the module's proper domain become that of approaching hippos? Yes, and for the same reasons as before: Its relative success at reacting to approaching hippos explains why this module remained a stable feature of an enduring species.[9]

Today, however, hippopotamuses too have vanished and there is a railway passing through the orgs' territory. Because orgs don't go near the rails, trains are no danger. However the same module that had reacted selectively to approaching elephants and then to approaching hippos now reacts to approaching trains (and produces a useless panic in the orgs). The *actual* domain of the module includes mostly approaching trains. Has its *proper* domain therefore become that of approaching trains? The answer should be "no" this time: Reacting to trains is what it does, but it is not its function. The module's reacting to trains does not explain its remaining a stable feature of the species. In fact, if the module and the species survive, it is in spite of this marginally harmful effect.[10]

Still, an animal psychologist studying the orgs today might well come to the conclusion that they have a domain-specific ability to react to trains. She might wonder how they have developed such an ability given that trains have

been introduced in the area too recently to allow the emergence of a specific biological adaptation (the adaptive value of which would be mysterious anyhow). The truth, of course, is that the earlier proper domains of the module, approaching elephants and then hippos, are now empty, that its actual domain is, by accident, roughly coextensive with the set of approaching trains, and that the explanation of this accident is the fact that the input conditions of the module, which had been positively selected in a different environment, happen to be satisfied by trains and hardly anything else in the orgs' present environment.

Enough of toy examples. In the real world, you are not likely to get elephants neatly replaced by hippos and hippos by trains, and to have each kind in turn satisfying the input conditions of some specialized module. Natural environments, and therefore cognitive functions, are relatively stable. Small shifts of cognitive function are more likely to occur than radical changes. When major changes occur in the environment, for instance as the result of a natural cataclysm, some cognitive functions are just likely to be lost: If elephants go, so does the function of your erstwhile elephant-detector. If a module loses its function, or equivalently if its proper domain becomes empty, then it is unlikely that its actual domain will be neatly filled by objects all falling under a single category, such as passing trains. More probably, the range of stimuli causing the module to react will end up being such an awful medley as to discourage any temptation to describe the actual domain of the module in terms of a specific category. Actual domains are usually not conceptual domains.

Cultural domains and the epidemiology of representations

Most animals get only highly predictable kinds of information from their conspecifics, and not much of it at that. They depend therefore on the rest of the environment for their scant intellectual kicks. Humans are special. They are naturally massive producers, transmitters, and consumers of information. They get a considerable amount and variety of information from fellow humans, and they even produce and store some for their own private consumption. As a result, I will argue, the actual domain of human cognitive modules is likely to have become much larger than their proper domain. Moreover these actual domains, far from being uncategorizable chaos, are likely to be partly organized and categorized by humans themselves. So much so, I will argue, that we should distinguish the *cultural domains* of modules from both their proper and actual domains.

Just a quick illustration before I give a more systematic sketch and a couple of more serious examples: Here is the infant in her cradle, endowed with a domain-specific, modular, naive physics. The proper domain of that module is a range of physical events that typically occur in nature, and the understanding of which is crucial to the organism's later survival. Presumably,

other primates are endowed with a similar module. The naive physics module of the infant chimp (and of the infant Pleistocene homo not-yet-sapiens) reacts to the odd fruit or twig falling, to the banana peel being thrown away, to occasional effects of the infant's own movement, and it may be challenged by the irregular fall of a leaf. Our human infant's module, on the other hand, is stimulated not just by physical events happening incidentally, but also by an "activity center" fixed to the side of her cradle, a musical merry-go-round just above her head, balls bounced by elder siblings, moving pictures on a television screen, and a variety of educational toys devised to stimulate her native interest in physical processes.

What makes the human case special? Humans change their own environment at a rhythm that natural selection cannot follow, so that many genetically specified traits of the human organism are likely to be adaptations to features of the environment that have ceased to exist or have greatly changed. This may be true not just of adaptations to the nonhuman environment, but also of adaptations to earlier stages of the hominid social environment.

In particular, the actual domain *of any* human cognitive module is unlikely to be even approximately coextensive with its proper domain. The actual domain of any human cognitive module is sure, on the contrary, to include a large amount of cultural information that meets its input conditions. This results neither from accident, nor from design. It results from a process of social distribution of information.

Humans not only construct individually *mental* representations of information, but they also produce information for one another in the form of *public* representations (e.g., utterances, written texts, pictures), or in the form of other informative behaviors and artifacts. Most communicated information, though, is communicated to one person or a few people on a particular occasion, and that is the end of it. Sometimes, however, addressees of a first act of communication communicate the information received to other addressees who communicate it in turn to others, and so on. This process of repeated transmission may go on to the point where we have a chain of mental and public representations both causally linked and similar in content – similar in content because of their causal links – instantiated throughout a human population. Traditions and rumors spread in this particular manner. Other types of representations may be distributed by causal chains of a different form (e.g., through imitation with or without instruction, or through broadcast communication). All such causally linked, widely distributed representations are what we have in mind when we speak of culture.

I have argued (Sperber, 1985b, 1990a, 1992) that to explain culture is to explain why some representations become widely distributed: A naturalistic science of culture should be an *epidemiology of representations*. It should explain why some representations are more successful – more contagious – than others.[11]

In this epidemiological perspective, all the information that humans

introduce into their common environment can be seen as competing[12] for private and public space and time, that is, for attention, internal memory, transmission, and external storage. Many factors affect the chances of some information being successful and reaching a wide and lasting level of distribution, of being stabilized in a culture. Some of these factors are psychological, others are ecological. Most of these factors are relatively local, others are quite general. The most general psychological factor affecting the distribution of information is its compatibility and fit with human cognitive organization.

In particular, relevant information, the relevance of which is relatively independent from the immediate context, is *ceteris paribus*, more likely to reach a cultural level of distribution: Relevance provides the motivation both for storing and for transmitting the information, and independence from an immediate context means that relevance will be maintained in spite of changes of local circumstances, that is, it will be maintained on a social scale. Relevance is, however, always relative to a context; independence from the immediate context means relevance in a wider context of stable beliefs and expectations. On a modular view of conceptual processes, these beliefs, which are stable across a population, are those that play a central role in the modular organization and processing of knowledge. Thus information that either enriches or contradicts these basic modular beliefs stands a greater chance of cultural success.

I have argued (Sperber, 1975, 1980, 1985b) that beliefs that violate head-on module-based expectations (e.g., beliefs in supernatural beings capable of action at a distance, ubiquity, metamorphosis, etc.) thereby gain a salience and relevance that contribute to their cultural robustness. Pascal Boyer (1990) has rightly stressed that these violations of intuitive expectations in the description of supernatural beings are in fact few and take place against a background of satisfied modular expectations. Kelly and Keil (1985) have shown that cultural exploitation of representations of metamorphoses are closely constrained by domain-based conceptual structure. Generally speaking, we should expect culturally successful information essentially to resemble that found in some proper domain, and at the same time to exhibit sufficient originality so as to avoid mere redundancy.

A cognitive module stimulates in every culture the production and distribution of a wide array of information that meets its input conditions. This information, being artifactually produced or organized by the people themselves, is from the start conceptualized and therefore belongs to conceptual domains that I propose to call the module's *cultural domain(s)*. In other terms, cultural transmission causes, in the actual domain of any cognitive module, a proliferation of parasitic information that mimics the module's proper domain.

Let me first illustrate this epidemiological approach with speculations on a nonconceptual case, that of music. This is intended to be an example of a way of thinking suggested by the epidemiological approach rather than a serious scientific hypothesis, which I would not have the competence to develop.

Imagine that the ability and propensity to pay attention to, and analyze certain complex sound patterns became a factor of reproductive success for a long enough period in human prehistory. The sound patterns would have been discriminable by pitch variation and rhythm. What sounds would have exhibited such patterns? The possibility that springs to mind is human vocal communicative sounds. It need not be the sounds of homo sapiens speech, though. One may imagine a human ancestor with much poorer articulatory abilities and relying more than modern humans do on rhythm and pitch for the production of vocal signals. In such conditions, a specialized cognitive module with the required disposition might well have evolved.

This module would have had to combine the necessary discriminative ability with a motivational force to cause individuals to attend to the relevant sound patterns. The motivation would have to be on the hedonistic side: pleasure and hopeful expectation rather than pain and fear. Suppose that the relevant sound pattern co-occurred with noise from which it was hard to discriminate. The human ancestor's vocal abilities may have been quite poor, and the intended sound pattern may have been embedded in a stream of parasitic sounds (a bit like when you speak with a sore throat, a cold, and food in your mouth). Then the motivational component of the module should have been tuned so that detecting a low level of the property suffices to procure a significant reward.

The proper domain of the module we are imagining is the acoustic properties of early human vocal communications. It could be that this proper domain is now empty: Another adaptation, the improved modern human vocal tract, may have rendered it obsolete. Or it may be that the relevant acoustic properties still play a role in modern human speech (in tonal languages in particular) so that the module is still functional. The sounds that the module analyzes thereby causing pleasure to the organism of which it is a part – that is, the sounds meeting the module's input conditions – are not often found in nature (with the obvious exception of bird songs). However, such sounds can be artificially produced. And they have been, providing this module with a particularly rich cultural domain: music. The relevant acoustic pattern of music is much more detectable and delectable than that of any sound in the module's proper domain. The reward mechanism, which was naturally tuned for a hard-to-discriminate input, is now being stimulated to a degree that makes the whole experience utterly addictive.

The idea is, then, that humans have created a cultural domain, music, which is parasitic on a cognitive module, the proper domain of which preexisted music and had nothing to do with it. The existence of this cognitive module has favored the spreading, stabilization, and progressive diversification and growth of a repertoire meeting its input conditions: First pleasing sounds were serendipitously discovered, then sound patterns were deliberately produced and became music proper. These bits of culture compete for mental and public space and time, and ultimately for the chance to stimulate

the module in question in as many individuals as possible for as long as possible. In this competition, some pieces of music do well, at least for a time, whereas others are selected out, and thus music, and musical competence, evolve.

In the case of music, the cultural domain of the module is much more developed and salient than its proper domain, assuming that it still has a proper domain. So much so that it is the existence of the cultural domain and the domain-specificity of the competences it manifestly evokes that justifies looking, in the present or in the past, for a proper domain that is not immediately manifest.

In other cases, the existence of a proper domain is at least as immediately manifest as that of a cultural one. Consider zoological knowledge. The existence of a domain-specific competence in the matter is not hard to admit, if the general idea of domain specificity is accepted at all. One way to think of it, as I have suggested, is to suppose that humans have a modular template for constructing concepts of animals. The biological function of this module is to provide humans with ways of categorizing animals they may encounter in their environment and of organizing the information they may gather about them. The proper domain of this modular ability is the living local fauna. What happens however is that you end up, thanks to cultural input, constructing many more animal concepts than there are animals with which you will ever interact. If you are a twentieth-century Westerner, you may, for instance, have a well-stocked cultural subdomain of dinosaurs. You may be a dinosaur expert. In another culture you might have been a dragon expert.

This invasion of the actual domain of a conceptual module by cultural information occurs irrespective of the size of the module. Consider a micromodule such as the concept of a particular animal, say the rat. Again, you are likely to have fixed, in the data base of that module, culturally transmitted information about rats, whether of a folkloristic or of a scientific character, that goes well beyond the proper domain of that micromodule, that is, well beyond information derivable from, and relevant to, interactions with rats (though, of course, it may be of use for your interactions with other human beings, e.g., by providing a data base exploitable in metaphorical communication).

On the macromodular side of things, accept for the sake of this discussion that the modular template on which zoological concepts are constructed is itself an initialized version (maybe the default version) of a more abstract living-kinds metatemplate. That metatemplate is initialized in other ways for other domains (e.g., botany), projecting several domain-specific templates, as I have suggested here. What determines a new initialization is the presence of information that (1) meets the general input conditions specified in the metatemplate, but (2) does not meet the more specific conditions found in the already initialized templates. That information need not be in the proper domain of the metatemplate module. In other words, the metatemplate might

get initialized in a manner that fits no proper domain at all but only a cultural domain. A cultural domain that springs to mind in this context is that of representations of supernatural beings (see Boyer, 1990, 1993, this volume). But there may also be less apparent cases.

Consider in this light the problem raised by Hirschfeld (this volume; see also Hirschfeld, 1988, 1993). Children are disposed to categorize humans into "racial" groups conceived in an essentialist manner. Do children possess a domain-specific competence for such categorization? In other terms, are humans naturally disposed to racism? In order to avoid such an unappealing conclusion, it has been suggested (Atran, 1990; Boyer, 1990) that children transfer to the social sphere a competence that they have first developed for living kinds, and that they do so in order to make sense of the regularities in human appearance (e.g., skin color) that they have observed. However, Hirschfeld's experimental evidence shows that racial categorization develops without initially drawing on perceptually relevant input. This strongly suggests that there is a domain-specific competence for racial classification.

What the epidemiological approach adds is the suggestion that racial classification might result from an ad hoc template derived from the living-kinds metatemplate, through an initialization triggered by a cultural input. Indeed, recent experiments suggest that, in certain conditions, the mere encounter with a nominal label used to designate a living thing is enough to tilt the child's categorization of that thing toward an essentialist construal (Davidson & Gelman, 1990; Gelman & Coley, 1991; Markman, 1990; Markman & Hutchinson, 1984). It is quite possible then that being presented with nominal labels for otherwise undefined and undescribed humans is enough (given an appropriate context) to activate the initialization of the ad hoc template. If so, then perception of differences among humans is indeed not the triggering factor.

There is, as Hirschfeld suggested, a genetically specified competence that determines racial classification without importing its models from another concrete domain. However, the underlying competence need not have racial classification as its proper domain. Racial classification may be a mere cultural domain, based on an underlying competence that does not have any proper domain. The initialization of an ad hoc template for racial classification could well be the effect of parasitic, cultural input information on the higher-level learning module the function of which is to generate ad hoc templates for genuine living-kind domains such as zoology and botany. If this hypothesis is correct – mind you, I am not claiming that it is, merely that it may be – then no racist disposition has been selected *for* (Sober, 1984) in humans. However the dispositions that have been selected for make humans all too easily susceptible to racism given minimal, innocuous-looking cultural input.

The relationship between the proper and the cultural domains of the same module is not one of transfer. The module itself does not have a preference

between the two kinds of domains, and indeed is blind to a distinction that is grounded in ecology and history.

Even when an evolutionary and epidemiological perspective is adopted, the distinction between the proper and the cultural domain of a module is not always easy to draw. Proper and cultural domains may overlap. Moreover, because cultural domains are things of this world, it can be a function of a module to handle a cultural domain, which ipso facto becomes a proper domain.

Note that the very existence of a cultural domain is an effect of the existence of a module. Therefore, initially at least, a module cannot be an adaptation to its own cultural domain. It must have been selected because of a preexisting proper domain. In principle, it might *become* a function of the module to handle its own cultural domain. This would be so when the ability of the module to handle its cultural domain contributed to its remaining a stable feature of an enduring species. The only clear case of an adaptation of a module to its own effects is that of the linguistic faculty. The linguistic faculty in its initial form cannot have been an adaptation to a public language that could not exist without it. On the other hand it seems hard to doubt that language has become the proper domain of the language faculty.[13]

If there are modular abilities to engage in specific forms of social interaction (as claimed by Cosmides, 1989), then, as in the case of the language faculty, the cultural domains of these abilities should at least overlap with their proper one. Another interesting issue in this context is the relationship between numerosity – the proper domain of a cognitive module – and numeracy, an obvious cultural domain dependent on language (see Dehaene, 1992; Gallistel & Gelman, 1992; Gelman & Gallistel, 1978). In general, however, there is no reason to expect the production and maintenance of cultural domains to be a biological function of all, or even most, human cognitive modules.

If this approach is correct, it has important implications for the study of domain specificity in human cognition. In particular it evaporates, I believe, the cultural diversity argument against the modularity of thought. For even if thought were wholly modular, we should nevertheless find many cultural domains, varying from culture to culture, and whose contents are such that it would be preposterous to assume that they are the proper domain of an evolved module. The cultural idiosyncrasy and lack of relevance to biological fitness of a cognitive domain leaves entirely open the possibility that it might be a domain of a genetically specified module: its cultural domain.

Metarepresentational abilities and cultural explosion

If you are still not satisfied that human thought could be modular through and through, if you feel that there is more integration taking place than I have allowed for so far, if you can think of domains of thought that

don't fit with any plausible module, well then we agree. It is not just that beliefs about Camembert cheese might play a role in forming conclusions about quarks, it is that we have no trouble at all entertaining and understanding a conceptual representation in which Camembert and quarks occur simultaneously. You have just proved the point by understanding the previous sentence.

Anyhow, with or without Camembert, beliefs about quarks are hard to fit into a modular picture. Surely, they don't belong to the actual domain of naive physics; similarly, beliefs about chromosomes don't belong to the actual domains of naive biology, beliefs about lycanthropy don't belong to the actual domain of folk zoology, beliefs about the Holy Trinity or about cellular automata seem wholly removed from any module.

Is this to say that there is a whole range of extramodular beliefs, of which many religious or scientific beliefs would be prime examples? Not really. We have not yet exhausted the resources of the modular approach.

Humans have the ability to form mental representations of mental representations; in other words, they have a metarepresentational ability. This ability is so particular, both in terms of its domain and of its computational requirements that anybody willing to contemplate the modularity of thought thesis will be willing to see it as modular. Even Fodor does (Fodor, 1992). The metarepresentational module[14] is a special conceptual module, however, a second-order one, so to speak. Whereas other conceptual modules process concepts and representations of things, typically of things perceived, the metarepresentational module processes concepts of concepts and representations of representations.

The actual domain of the metarepresentational module is clear enough: It is the set of all representations of which the organism is capable of inferring or otherwise apprehending the existence and content. But what could be the proper domain of that module? Much current work (e.g., Astington et al., 1989) assumes that the function of the ability to form and process metarepresentations is to provide humans with a naive psychology. In other terms, the module is a "theory of mind module" (Leslie, this volume), and its proper domain is made of the beliefs, desires, and intentions that cause human behavior. This is indeed highly plausible. The ability to understand and categorize behavior, not as mere bodily movements, but in terms of underlying mental states, is an essential adaptation for organisms that must cooperate and compete with one another in a great variety of ways.

Once you have mental states in your ontology, and the ability to attribute mental states to others, there is but a short step, or no step at all, to your having desires about these mental states – desiring that she should believe this, desiring that he should desire that – and to forming intentions to alter the mental states of others. Human communication is both a way to satisfy such metarepresentational desires, and an exploitation of the metarepresentational abilities of one's audience. As suggested by Grice (1957) and developed

by Deirdre Wilson and myself (1986), a communicator, by means of her communicative behavior, is deliberately and overtly helping her addressee to infer the content of the mental representation she wants him to adopt (Sperber & Wilson, 1986).

Communication is, of course, radically facilitated by the emergence of a public language. A public language is rooted in another module, the language faculty. We claim, however, that the very development of a public language is not the cause, but an effect of the development of communication made possible by the metarepresentational module.

As a result of the development of communication, and particularly of linguistic communication, the actual domain of the metarepresentational module is teeming with representations made manifest by communicative behaviors: intentions of communicators and contents communicated. Most representations about which there is some interesting epidemiological story to be told are communicated in this manner and therefore enter people's minds via the metarepresentational module.

As already suggested, many of the contents communicated may find their way to the relevant modules: What you are told about cats is integrated with what you see of cats, in virtue of the fact that the representation communicated contains the concept CAT. But now you have the information in two modes: as a representation of cats, handled by a first-order conceptual module, and as a representation of a representation of cats, handled by the second-order metarepresentational module. That module knows nothing about cats but it may know something about semantic relationships among representations; it may have some ability to evaluate the validity of an inference, the evidential value of some information, the relative plausibility of two contradictory beliefs, and so forth. It may also evaluate a belief, not on the basis of its content, but on the basis of the reliability of its source. The metarepresentational module may therefore form or accept beliefs about cats for reasons that have nothing to do with the kind of intuitive knowledge that the CAT module (or whatever first-order module handles cats) delivers.

An organism endowed with perceptual and first-order conceptual modules has beliefs delivered by these modules, but has no beliefs about beliefs, either its own or those of others, and no reflexive attitude to them. The vocabulary of its beliefs is limited to the output vocabulary of its modules, and it cannot conceive or adopt a new concept nor criticize or reject old ones. An organism also endowed with a metarepresentational module can represent concepts and beliefs qua concepts and beliefs, evaluate them critically, and accept them or reject them on metarepresentational grounds. It may form representations of concepts and of beliefs pertaining to all conceptual domains, of a kind that the modules specialized in those domains might be unable to form on their own, or even to incorporate. In doing so, however, the better endowed organism is merely using its metarepresentational module within the module's own domain, that is, representations.

Humans, with their outstanding metarepresentational abilities, may thus have beliefs pertaining to the same conceptual domain rooted in two quite different modules: The first-order module specialized in that conceptual domain, or the second-order metarepresentational module, specialized in representations. These are, however, two different kinds of beliefs, "intuitive beliefs" rooted in first-order modules, and "reflective beliefs" rooted in the metarepresentational module (see Sperber, 1985a, chap. 2, 1985b, 1990a). Reflective beliefs may contain concepts (e.g., "quarks," "Trinity") that do not belong in the repertoire of any module, and that are therefore available to humans only reflectively, via the beliefs or theories in which they are embedded. The beliefs and concepts that vary most from culture to culture (and that often seem unintelligible or irrational from another culture's perspective) are typically reflective beliefs and the concepts they introduce.

Reflective beliefs can be counterintuitive (more exactly, they can be counterintuitive with respect to our intuitions about their subject matter, while, at the same time, our metarepresentational reasons for accepting them are intuitively compelling). This is relevant to the most interesting of Fodor's technical arguments against the modularity of central processes. The informational encapsulation and mandatory character of perceptual modules is evidenced, Fodor points out, by the persistence of perceptual illusions, even when we are apprised of their illusory character. There is, he argues, nothing equivalent at the conceptual level. True, perceptual illusions have the feel, the vividness of perceptual experiences, that you won't find at the conceptual level. But what you do find is that we may give up a belief and still feel its intuitive force, and feel also the counterintuitive character of the belief we adopt in its stead.

You may believe with total faith in the Holy Trinity, and yet be aware of the intuitive force of the idea that a father and son cannot be one and the same. You may understand why black holes cannot be seen, and yet feel the intuitive force of the idea that a big solid, indeed dense object cannot but be visible. The case of naive versus modern physics provides many other blatant examples.[15] What happens, I suggest, is that the naive physics module remains largely unpenetrated by the ideas of modern physics, and keeps delivering the same intuitions, even when they are not believed any more (or at least not reflectively believed).

More generally the recognition of the metarepresentational module, of the duality of beliefs that it makes possible, and of the gateway it provides for cultural contagion, plugs a major gap in the modular picture of mind I have been trying to outline. The mind is here pictured as involving three tiers: a single thick layer of input modules, just as Fodor says, then a complex network of first-order conceptual modules of all kinds, and then a second-order metarepresentational module. Originally, this metarepresentational module is not very different from the other conceptual modules, but it allows the development of communication and triggers a cultural explosion of such

magnitude that its actual domain is blown up and ends up hosting a multitude of cultural representations belonging to several cultural domains.

This is how you can have a truly modular mind playing a major causal role in the generation of true cultural diversity.

Notes

1. Fodor also mentions the possibility that output, i.e., motor systems might be modular too. I assume that it is so, but will not discuss the issue here.
2. Howard Gardner's *Frames of Mind* (1983) defends a modular theory of central processes with a concern that I share for the cultural aspect of the issue. My approach is otherwise quite different from his.
3. See also Barkow (1989), Barkow, Cosmides, & Tooby (1992), Brown (1991), Rozin (1976), Rozin & Schull (1988), and Symons (1979).
4. This is, of course, the "frame problem," the very existence of which Fodor (1987) sees as indissolubly linked to the nonmodularity and to the rationality of thought. The frame problem, qua psychological problem, is being overestimated. Two psychological hypotheses allow us to reduce it to something tractable. First the modularity of thought hypothesis, as pointed out by Tooby & Cosmides (1992) considerably reduces the range of data and procedures that may be invoked in any given conceptual task. Second, the hypothesis that cognitive processes tend to maximize relevance (Sperber & Wilson, 1986) radically narrows down the actual search space for any conceptual task.
5. The point cannot just be that the forces that have driven cognitive evolution cannot be identified for certain; that much is trivially true. The claim must be that these forces cannot be even tentatively and reasonably identified, unlike the forces that have driven the evolution of, say, organs of locomotion. See Piatelli-Palmarini (1989) and Stich (1990) for clever but unconvincing arguments in favor of this Second Law.
6. Note that if apparent lack of hardwiring was an obstacle to acknowledging modularity, this would be an obstacle in the case of Fodor's linguistic input modules too. Take the case of a bilingual. Surely she has two modules, one for each language. Both result from fixing parameters and filling a lexicon in a template module, the language acquisition device. However we should be reluctant to imagine that there were (at least) two hardwired templates in place, waiting to be initialized. Hence, at least one of the initialized templates results from a projection of the initial structure onto new sites.
7. Burge (1979) and Putnam (1975) offered the initial arguments for externalism (I myself am convinced by Putnam's arguments but not by Burge's). For a sophisticated discussion, see Recanati (1993).
8. Arguably, content is a biological function in an extended sense – see Dretske (1988), Millikan (1984), and Papineau (1987). My views have been influenced by Millikan's.
9. There are of course conceptual problems here (see Dennett, 1987; Fodor, 1988). It could be argued, for instance, that the module's proper domain was neither elephants nor hippos, but something else, say, "approaching big animals that might trample orgs." If so, we would want to say that its proper domain had *not*

changed with the passing of the elephants and the coming of the hippos. I side with Dennett in doubting that much of substance hinges on which of these descriptions we choose: The overall explanation remains exactly the same.

10. That is why it would be a mistake to say that the function of a device is to react to whatever might satisfy its input conditions and to equate its actual and proper domains. Though there may be doubt about the correct assignment of the proper domain of some device (see note 9), the distinction between actual and proper domains is as solid as that between effect and function.
11. Comparable evolutionary or epidemiological views of culture have been put forward by Boyd and Richerson (1985), Cavalli-Sforza and Feldman (1981), Dawkins (1976), and myself (in addition to some very different evolutionary approaches by many others). The epidemiology of representations that I have been advocating differs from other approaches (1) by stressing the importance of individual cognitive mechanisms in the overall explanation of culture, and (2) by arguing that information is transformed every time it is transmitted to such an extent that an analogy with biological reproduction or replication is inappropriate. See also Tooby and Cosmides (1992) for important new developments in this area.
12. Here, as in talk of representations competing for attention, the term "competition" is only a vivid metaphor. Of course, no intention or disposition to compete is implied. What is meant is that, out of all the representations present in a human group at a given time, some, at one extreme, will spread and last, whereas, at the opposite extreme, others will occur only very briefly and very locally. This is not a random process, and it is assumed that properties of the information itself play a causal role in determining its wide or narrow distribution.
13. See Pinker and Bloom (1990) and my contribution to the discussion of their paper (Sperber 1990b).
14. The capacity to form and process metarepresentations could be instantiated not in a single, but in several distinct modules, each, say, metarepresenting a different domain or type of representations. For lack of space and compelling arguments, I will ignore this genuine possibility.
15. And a wealth of subtler examples have been analyzed in a proper cognitive perspective by Atran (1990).

References

Astington, J. W., Harris, P., & Olson, D. (1989). *Developing theories of mind.* New York: Cambridge University Press.

Atran, S. (1987). Ordinary constraints on the semantics of living kinds. *Mind & Language, 2*(1), 27–63.

Atran, S. (1990). *Cognitive foundations of natural history.* New York: Cambridge University Press.

Barkow, J. H. (1989). *Darwin, sex and status: Biological approaches to mind and culture.* Toronto: University of Toronto Press.

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

Berlin, B. (1978). Ethnobiological classification. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Boyd, Robert, & Richerson, Peter J. (1985). *Culture and the evolutionary process.* Chicago: The University of Chicago Press.

Boyer, P. (1990). *Tradition as truth and communication.* New York: Cambridge University Press.

Boyer, P. (1993). *The naturalness of religious ideas.* Berkeley: University of California Press.

Brown, D. (1991). *Human universals.* New York: McGraw-Hill.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 5,* 73–122.

Carey, S. (1985). *Conceptual development in childhood.* Cambridge, MA: MIT Press.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach.* Princeton: Princeton University Press.

Clark, A. (1987). The kludge in the machine. *Mind and Language, 2*(4), 277–300.

Clark, A. (1990). *Microcognition: Philosophy, cognitive science, and parallel distributed processing.* Cambridge, MA: MIT Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,* 187–276.

Cosmides, L., & 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.

Davidson, N. S., & Gelman, S. (1990). Induction from novel categories: The role of language and conceptual structure. *Cognitive Development, 5,* 121–152.

Dawkins, Richard. (1976). *The selfish gene.* Oxford: Oxford University Press.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1–2), 1–42.

Dennett, D. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

Dretske, F. (1988). *Explaining behavior.* Cambridge, MA: MIT 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. (1988). *Psychosemantics.* Cambridge, MA: MIT Press.

Fodor, J. (1992). A theory of the child's theory of mind. *Cognition, 44,* 283–296.

Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1–2), 43–74.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gelman, S., & Coley, J. D. (1991). The acquisition of natural kind terms. In S. Gelman & J. Byrnes (Eds.), *Perspectives on language and thought.* New York: Cambridge University Press.

Gelman, S., & Markman, E. (1986). Categories and induction in young children. *Cognition, 23,* 183–209.

Gelman, S., & Markman, E. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58,* 1532–1541.

Grice, H. P. (1957). Meaning. *Philosophical Review, 66,* 377–388.

Hirschfeld, L. (1988). On acquiring social categories: Cognitive development and anthropological wisdom. *Man, 23*, 611–638.

Hirschfeld, L. (1993). Discovering social difference: The role of appearance in the development of racial awareness. *Cognitive Psychology, 25*, 317–350.

Kelly, M., & Keil, F. C. (1985). The more things change . . .: Metamorphoses and conceptual development. *Cognitive Science, 9*, 403–416.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: Bradford Books/MIT Press.

Leslie, A. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review, 94*, 412–426.

Leslie, A. (1988). The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language.* Oxford: Clarendon Press.

Manktelow, K., & Over, D. (1990). *Inference and understanding: A philosophical and psychological perspective.* London: Routledge.

Markman, E. M. (1990). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. Gelman & J. Byrnes (Eds.), *Perspectives on language and thought.* New York: Cambridge University Press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*, 1–27.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.

Millikan, R. G. (1984). *Language, thought, and other biological categories.* Cambridge, MA: MIT Press.

Papineau, D. (1987). *Reality and representation.* Oxford: Blackwell.

Piatelli-Palmarini, M. (1989). Evolution, selection and cognition: From "learning" to parameter setting in biology and the study of language. *Cognition, 31*, 1–44.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*(4), 703–784.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1–16.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.

Putnam, H. (1975). The meaning of "meaning." In *Mind, language and reality: Philosophical papers, volume II.* Cambridge: Cambridge University Press.

Recanati, F. (1993). *Direct reference, meaning and thought.* Oxford: Blackwell.

Rozin, P. (1976). The evolution of intelligence and access to the cognitive unconscious. In J. M. Sprague & A. N. Epstein (Eds.), *Progress in psychobiology and physiological psychology.* New York: Academic Press.

Rozin, P., & Schull, J. (1988). The adaptive-evolutionary point of view in experimental psychology. In R. Atkinson, R. Herrnstein, G. Lindzey, & R. Luce (Eds.), *Steven's handbook of experimental psychology.* New York: John Wiley & Sons.

Sober, E. (1984). *The nature of selection.* Cambridge, MA: MIT Press.

Spelke, E. S. (1988). The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without language.* Oxford: Clarendon Press.

Sperber, D. (1975). *Rethinking symbolism.* Cambridge: Cambridge University Press.

Sperber, D. (1980). Is symbolic thought prerational? In Mary Foster & Stanley Brandes (Eds.), *Symbol as sense*. New York: Academic Press.

Sperbér, D. (1985a). *On anthropological knowledge*. New York: Cambridge University Press.

Sperber, D. (1985b). Anthropology and psychology: Towards an epidemiology of representations (The Malinowski Memorial Lecture 1984). *Man* (N.S.) *20*, 73–89.

Sperber, D. (1990a). The epidemiology of beliefs. In C. Fraser & G. Gaskell (Eds.), *The social psychological study of widespread beliefs*. Oxford: Clarendon Press.

Sperber, D. (1990b). The evolution of the language faculty: A paradox and its solution. *Behavioral and Brain Sciences*, *13*(4), 756–758.

Sperber, D. (1992). Culture and matter. In J.-C. Gardin & C. S. Peebles (Eds.), *Representations in archeology*. Bloomington: Indiana University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

Stich, S. (1990). *The fragmentation of reason*. Cambridge, MA: MIT Press.

Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, Part I: Theoretical considerations. *Ethology & Sociobiology*, *10*, 29–49.

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