DAN SPERBER

Modularity and Relevance

How Can a Massively Modular Mind Be Flexible and Context-Sensitive?

L et me start with a quotation from Randy Gallistel (1999, p. 1179, echoing Chomsky, 1975):

> Adaptive specialization of mechanisms is so ubiquitous and so obvious in biology, at every level of analysis, and for every kind of function, that no one thinks it necessary to call attention to it as a general principle about biological mechanisms. In this light, it is odd but true that most past and present contemporary theorizing about learning does not assume that learning mechanisms are adaptively specialized for the solution of particular kinds of problems. Most theorizing assumes that there is a general-purpose learning process in the brain, a process adapted only to solving the problem of learning. . . . From a biological perspective, this assumption is equivalent to assuming that there is a general-purpose sensory organ that solves the problem of sensing.

Gallistel's remark can be extended to cognition in general. It is odd but true that most past and present contemporary theorizing *about cognition* does not assume that cognitive mechanisms are adaptively specialized for the solution of particular kinds of problems. There is indeed a great divide today between a minority of cognitive scientists for whom mind-brains are best viewed as articulations of specialized modules and a majority for whom at least the *human* mind-brain is largely nonmodular. I belong to the minority and have argued the case for

massive modularity elsewhere.[1] What I want to do here is answer two questions: How can a massively modular mind be flexible? And: How can a massively modular mind be context-sensitive? The two questions are related: the context of cognitive processes is changing every fraction of a second, if only because it is modified by these very processes. In verbal comprehension, for instance, the interpretation of every utterance modifies the context in which the next utterance is interpreted. Context-sensitivity is the ability to take this ever-changing context into account. "Flexibility" (or "plasticity") is a metaphor that is best unpacked as meaning context-sensitivity in the longer run. An individual cognitive system is flexible if it can modify itself on the basis of experience. When humans in general are described as a particularly flexible species, it is even longer term context-sensitivity that is involved: over historical time, humans have adapted to very diverse natural and humanmade environments and have, for this, developed novel cognitive competencies. Clearly, a system that is flexible is in a better position to exhibit context-sensitivity in the short run.

1   Cognitive Modules Are a Type of Biological Module

Given that discussions of cognitive modularity often get bogged down in tedious terminological arguments, I might have been tempted to avoid the term "module" altogether if it were not that that there is much recent relevant work on biological modularity (e.g., Schlosser & Wagner, 2004), of which cognitive modularity is best seen, I want to argue, as a special case. It is hardly controversial that complex organisms are systems made up of many distinct subsystems—including but not limited to classical "organs," now often called "modules"—that may differ from one another functionally, structurally, ontogenetically, and phylogenetically. A modular organization is an effect of biological evolution, which responds in a piecemeal fashion to challenges presented by the environment. Arguably, modularity is also a condition of evolvability (Wagner & Altenberg, 1996). Because they are opportunistic responses to a great variety of problems and opportunities, it is in the nature of modules to be quite diverse in form, size, and function. Hence, one cannot both appreciate the role of modularity in biological systems and ask for a precise and rich definition of what a module is, or insist that a genuine module should resemble some prototype. Let me repeat, if you insist that a module should be defined in a narrow and rigid way, you are ignoring the evolutionary dimension of modularity.

Biological modules can be articulated in a variety of ways and can, in particular, contain submodules. For instance, the vertebrate digestive system is itself a complex module and contains as submodules various portions of the digestive tract such as the pharynx, the stomach, or the large intestine, glands such as the salivary glands or

---

1. See Sperber (1994), revised and expanded in Sperber (1996, 2001). It was under the influence of Chomsky that I was first led to argue that the human mind should be viewed as an articulation of autonomous domain-specific device (Sperber, 1974). Later, the work of Cosmides and Tooby (1992, 1994) convinced me that an evolutionary perspective, which I had taken as mere background, was crucial to developing such a view. Much of my thinking on the issue has, of course, been shaped by Fodor (1983), even when I disagree with him.

the liver, chemical modules such as hormones and enzymes produced by the glands, and so on. Inherited modules can evolve and both turn into and generate new modules in the lifetime of the organism. For instance, B lymphocytes are inherited cell-sized modules that evolve within the organism and generate antibodies, that is, new protein-sized modules the function of which is to bind to, and thereby neutralize, specific antigens. It may not be obvious at first to think of modules the size and character of freely moving short-lived cells and proteins, but, again, the point about a modular organization is that it may contain as modules any autonomously functioning device with a phylogenetic or ontogenetic history of its own.

If cognitive modules are real components of the cognitive system and not mere boxes in a nominalist flow-chart model, then they are a subtype of biological modules. They are characterized in particular by specific input conditions and by proprietary resources used to process inputs that meet these conditions. The inputs that happen to meet the input conditions of a given module constitute what I have called its *actual domain* (Sperber, 1994). In most cases, these input conditions are an imperfect but effective way of picking out items that belong to some objective category or domain of items in the environment. This objective domain then is the *proper domain* of the module. The function of the module is to inform the organism about items in its proper domain. It is with reference to such a proper domain that a module can be said to be domain specific. A module might, for instance, accept as inputs sounds exhibiting specific structural patterns, when, in the environment where this module operates, such sound patterns almost always correspond to speech in a given natural language. Then the proper domain of this module would be speech in that language (even if it might be activated by some nongenuinely linguistic sound pattern à la "Jabberwocky").

A cognitive module has its own procedures and may also have a database of its own. A face recognition module, for instance, has both data about the faces it is capable of recognizing and dedicated procedures to match perceptual inputs to these data. The fact that a module can draw only on a limited database, if any, to process its inputs is what Fodor (1983, 2000) calls "informational encapsulation," one of several criteria for modularity in his *Modularity of Mind* (1983) and the only one that plays a significant role in his book *The Mind Doesn't Work That Way* (2000). Because an informationally encapsulated device only has access to limited information, excluding some information that might in principle be pertinent to its producing the right outputs and that might be available elsewhere in the organism, it fails to exhibit the context-sensitivity that is characteristic of human cognition as a whole. Paradigm examples are provided by perceptual illusions: *I* (that is, a whole person) have the information that the two lines in the Müller-Lyer illusion are equal (say, I have measured them), but my visual perceptual device has no access to this information and keeps "seeing" them as unequal. Cognitive reflexes are, in this respect, extreme cases of encapsulation: given the proper input, they immediately deliver their characteristic output, whatever its appropriateness in the context.

It is important to distinguish *domain-specificity* from *encapsulation*. A device is *domain specific* if its function is to process only inputs belonging to some specific empirical domain (even if its input conditions do not perfectly pick out all and only items in this domain, so that there is a degree of mismatch between its proper

and its actual domain). For instance, a face recognition device has as its function to process faces (even if its operation can also be triggered by merely face-like stimuli, e.g., masks). An *encapsulated* device is one that uses a limited database to process its inputs. A word recognition device, for instance, takes as characteristic inputs phonetic representations of speech and uses as a database a dictionary. It is plausible that there are domain-general mental devices. Working memory, for instance, might be seen as a domain-general device that processes inputs whatever their contents, and manages their level of activation for the benefit of other, inferential devices. I cannot think, on the other hand, of a plausible example of a nonencapsulated mental device, that is, of a device that would use the whole mental encyclopedia as its database. Nonencapsulation is, tautologically, a property of the mind as a whole, but it does not seem to be a property of any autonomous subcomponent of it.[2]

What a cognitive module does at a given time (if it does anything at all) is determined by the inputs it is processing, by its procedures, and by its database, if any. It is not *directly* governed by what other modules of the cognitive system are doing, and does not *directly* draw on the informational resources available to these other components. I stress "directly" because there are, of course, indirect ways in which modules affect one another. Apart from sensory organs, all components of the cognitive system get their inputs from other components: roughly speaking, face recognition gets its input from visual perception, pragmatic interpretation of utterances gets part of its input from linguistic decoding, and so on. So a module's operations are typically triggered by being fed as input the output of some other module. Moreover, the triggering input typically has been informed by the procedures and data of the feeder module. Still, once it is performing its function, a module works on its own and is unable to take advantage of information that might be present in the system as a whole but that is found neither in the input nor in the proprietary database of the module.

Isn't there a risk, though, when allowing for a great variety of modules networked in complex ways, of trivializing the notion of modularity to the point of confusing modules with the boxes used in diagrams representing the flow of

2. Fodor, it is true, gives as an example of nonencapsulation the case of modus ponens inference, that is, an inference that takes as input any pair of beliefs of the form {P, [If P then Q]} and produces as output the belief that Q. Modus ponens, Fodor argues (2000, pp. 60–62), 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 this 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. However, the difference between the wholly general and the number-specific modus ponens devices is one of inputs, and therefore of domain specificity, not one of database, and therefore not of encapsulation. Both the general and the bridled modus ponens inferences apply a procedure to pairs of premises and do so without using any data. In particular, they ignore data that might cause a rational agent to refrain from performing the modus ponens inference and to question one or other of the premises instead (Harman, 1986). If there is a modus ponens inference procedure in the human mind, it is better viewed, I would argue, as cognitive reflex (Sperber, 2001).

information in cognitive processes? The risk is avoided, I maintain, by the modularist's commitment to biologically realistic interpretation of the boxes. A boxological flow chart can be interpreted as a mere algorithmic representation of a complex cognitive process showing how, in principle, the process could be materially realized but carrying no commitment regarding its actual implementation in mind-brains. The true modularist is interested in "boxes" that correspond to neurologically distinct devices. A neurologically distinct device, or module, need not occupy a single and continuous brain location all by itself—its boundaries need not be sharp—but still, it must be distinguishable not just functionally but also neurologically. This presupposes that a module has a distinct history in the development of the individual brain, and this in turn presupposes some genetic and evolutionary story about the conditions that make such an individual development possible.

The issue now is whether such an articulation of biologically real cognitive modules could exhibit the flexibility and context-sensitivity exhibited by the human mind as a whole.

2   Modularity and Flexibility

Modules are "rigid." The human mind, on the other hand, is "flexible." Since both "rigid" and "flexible" are metaphors, this raises not so much a serious objection to a modularist view of the human mind as an interesting question: How could flexibility be achieved in such a modular system? The answer is that most innate[3] cognitive modules are domain-specific *learning* mechanisms ("learning instincts," Marler, 1991, or "module templates," Sperber, 1994) that generate the working modules of acquired cognitive competence.

Even though the existence and many characteristics of mental modules are explained by biological evolution, this does not imply that modules are simply phenotypic expressions of genes, or that the development of each and every module is strongly canalized. On the contrary, it would be in the nature of modules to differ vastly from one another in this as in other respects. For some of the problems cognitive modules handle, "prewiring" may be appropriate. For other problems, an effective modular solution may involve adding data to the proprietary database of an otherwise predetermined module. In other cases still, the development of a module may involve drawing on information picked up from the environment not just to enrich the database but also to shape procedures.

There is, in fact, a full range of cases from innately specified modules to brain tissues that are merely ready to modularize competencies of a specific type. Here are five examples across this range:

- *Avoidance of vertical drops*: Human infants (and other baby animals also) perceive and avoid vertical drops in terrain, even if they have had no experience of falling before, as was demonstrated by means of the well-known "visual cliff" experiments initiated by Gibson and Walk

---

3. "Innate" in the sense of Samuels (2002).

(1960). This is an obvious modular adaptation to a serious hazard facing animals moving on the ground. To be efficient, this particular module had better not depend on learning. It is as good an example of an innate cognitive module as one may ever hope to find.

- *The Garcia effect* (Garcia & Koelling, 1966): Rats and other animals are innately equipped to develop an aversion to whatever type of food seems to have made them sick. This is a highly specialized one-pass-learning module. The outcome of such learning is a novel capacity, that of re-acting with aversion to a specific kind of food. If the rat develops, say, three such aversions, then it has three distinct abilities. It could be that the learning process and each specific aversive reaction are all carried out by the same module: learning consisting in adding to the initially empty proprietary database of the module data about specific foods to be avoided. Or it could be that the learning process results each time in the setting up of a new module or submodule dedicated to a specific aversive food. So which is it—one general food-aversion module with a growing database, or a learning module producing as many micro-modules as there are aversions? This is an empirical issue that might be decided by answering questions such as the following: Do aversive reactions to different foods employ different detection procedures (as opposed to the same procedure using different data)? Does a new aversion recruit distinct brain tissues? Can the more general ability to generate new aversions and each of the more specific aversions be selectively impaired? Positive answers to such questions would suggest that to each new aversion corresponds a new mini (sub) module.
- *Face recognition*: I assume that face recognition is modular (which is controversial, but see Kanwisher & Moscovitch, 2000). If so, we are deal-ing, as in the case of the Garcia effect, with two types of modular abil-ities: a *general* learning ability to form *specific* abilities to detect specific faces. Is there a general face recognition module that performs both functions or are individual-face-detectors developed as autonomous mini (sub) modules? This is an empirical question to which we do not have an answer. As in the case of the Garcia effect, these are nevertheless genuinely distinct possibilities involving subtle differences in the way these abilities may be carried out and impaired.
- *Language faculty and linguistic competences:* The language faculty is a complex learning module that, given proper linguistic and contextual inputs, yields one or, in the case of plurilinguals, several mental grammars. Each of these grammars is itself a complex module subserving both verbal coding and decoding in a given language. Each mental grammar has a distinct developmental story, and can selectively decay or be impaired. It is plausible that, say, the two mental grammars of a bilingual individual are submodules of a more general mental universal grammar and, as such, share some resources (Dehaene et al., 1997; Kim et al., 1997).
- *Reading*: Reading is too recent a cultural skill for a specialized innate module to have evolved. Yet reading systematically involves the same

brain site located in the left occipito-temporal sulcus and sometimes described as the "visual word form area." Dehaene speculates that "the human brain can learn to read because part of the primate visual ventral object recognition system spontaneously accomplishes operations closely similar to those required in word recognition, and possesses sufficient plasticity to adapt itself to new shapes, including those of letters and words. During the acquisition of reading, part of this system becomes highly specialized for the visual operations underlying location- and case-invariant word recognition....Thus, reading acquisition proceeds by selection and local adaptation of a preexisting neural region, rather than by *de novo* imposition of novel properties onto that region" (Dehaene, forthcoming). Reading skill can be viewed as resulting from a process of ad hoc modularization of already specialized brain tissues.

    With many innate modules being learning modules generating further modules, with brain areas ready to modularize, one may envisage that the human mind is characterized not only by massive modularity but also by *teeming modularity*. A great many highly specialized procedures—the size, say, of a specific concept or even of a particular inference rule—may be modular in the intended sense. That is, there may be a plethora of distinct biological devices emerging on some innate basis in the course of cognitive development, and functioning with a certain degree of autonomy in cognitive activity (a similar view, based on an analogy between cognitive modules and enzymes, is developed by Clark Barrett, forthcoming). I hope these remarks help one clarify how a massively modular mind may indeed be flexible, even if the detailed ways in which such flexibility is achieved are obviously a matter for empirical research.

3   How Can a Massively Modular Mind Be Context-Sensitive?

According to Fodor, in human cognition, only peripheral input systems are modular. One of the distinctive properties of modular input systems, he argues, is that their operations are mandatory. Supporters of the idea of massive modularity, not just at the input level but at all levels of cognitive activity, shouldn't lightly accept the idea that mandatoriness characterizes modular operations. If all the modules of a massively modular mind mandatorily processed any input available to them (including the outputs of other modules that meet their input conditions) there would be a computational explosion. Even if such a system could work at all, it is hard to see how it could exhibit the kind of context-sensitivity that is characteristic of human cognition. Every input would be processed in the same way in every situation. Of course, some limited context-sensitivity could still be built into such a system. The output of a given module could inhibit the operations of another module: the standard violent response to an apparently aggressive movement, for instance, can be inhibited by the perception of signs of playfulness. A danger detection module, acting as an "*and*-gate," may accept only complex inputs such as pairs of more elementary inputs, for instance a sound *and* a visual signal. In such cases, there is an in-built context-dependency, but it remains quite local, unlike the

context-dependency displayed by ordinary human cognition in, for example, verbal comprehension.

If one takes for granted that modularity implies mandatoriness, then one should reject the massive modularity hypothesis. My strategy will be to examine and question the idea that the operation of modules must be mandatory—even in the case of Fodorian input modules. I will then argue that the system as a whole exhibits context-sensitivity through the allocation of energy among modules.

There are two senses in which a cognitive procedure might be said to be mandatory. In a first sense—the only one in which I will use the term—a procedure is mandatory if, given the appropriate input, it will follow its course and produce its output whatever the rest of the mind-brain is doing (except in cases of pathological or accidental impairments). In other words, the procedure is mandatory in the sense that an appropriate input is sufficient to trigger it in such a manner that it will run its course (and not just to give it some initial activation). In a second sense, a procedure is "mandatory" if it cannot be voluntarily willed or blocked (except in an indirect way, for instance by acting on the availability of the inputs rather than on the procedure itself)—for this I will only use "involuntary." When Fodor argues that the operations of mental modules are "mandatory," he seems to have both senses in mind. It is self-evident that a procedure that is mandatory in the first sense, that is, automatically stimulus triggered, would be "mandatory" in the second sense, that is, involuntary. There are procedures that are indeed both mandatory (in the first sense) and involuntary. For instance, perceiving an object as colored is automatically triggered by the stimulus and cannot be willed or blocked. Similarly, being presented with a pair of numbers such as 50 and 100 automatically triggers (in a person familiar with numbers) a comparison of their size, before any decision could be taken to perform or not to perform such a comparison. Still, the two properties, that of being mandatory, that is, input triggered, and that of being involuntary are far from being coextensive. There are many cognitive procedures over which the individual has no voluntary control and that, in the course of ordinary cognitive activity, may be inhibited or enhanced both by mind-internal factors such as expectations and by mind-external factors such as distracting stimuli. These procedures are neither voluntary nor mandatory.

If I see just in front of me, in broad daylight, the face of my Paris dentist, Monsieur Durand, I cannot help but recognize him. My face recognition module (or my Monsieur-Durand-detection submodule) does its job. But suppose I am lecturing in London. Some 30 faces in front of me are each clearly visible. I look cursorily at all of them and I recognize some colleagues. Even though I have looked at his face as much as at those of the people I immediately recognized, it is only toward the end of the lecture that I suddenly recognize, sitting there in the second row, Monsieur Durand, whom I would never have expected to see in such a place.

The operations of input modules seem mandatory when you consider only cases where the stimulus is, and stays long enough, at fixation, and the perceiver is not actively tracking some other stimulus. Striking experimental demonstration of this is provided by work on "inattentional blindness." For instance, Simons and Chabris (1999) found that about 50 percent of participants asked to monitor a basketball passing event on a screen failed to notice a gorilla that walked across the screen in

full view, stopped in the middle of the players as the action continued all around it, turned to face the camera, thumped its chest, and then resumed walking. There are many, more banal cases, concerning most if not all input modules, where a stimulus is well within the field of perception but either is not in a focal position or is not sufficiently attended to, where the resources of the mind are invested in processing other competing stimuli, or inner thoughts, and where the module fails to process the stimulus (or at least fails to process it sufficiently): the familiar face is not recognized, the sentence structure is not parsed, the gorilla walks unnoticed. Let me insist, I am talking about cases where the psychophysical perceptual conditions for the operation of the module are satisfied and where, with less competition from other stimuli or other thoughts, or with appropriate expectations facilitating the process, the stimulus would have been processed. At least some of the procedures involved in perceiving the gorilla are not mandatory. There may well be an initial activation of the relevant procedures, but, when an individual's attention is focused on something else, they may not run their full course. I take it that the idea that visual perception is modular is not put in jeopardy by such data. Then, however, man-datoriness cannot be a defining trait of modules. (By the way, I am not trying to make a terminological but a substantive point. If these perceptual procedures that fail to deliver their expected output in the inattentional blindness experiments mentioned earlier are still "mandatory" by your definition, so be it. What matters here is that the availability of an appropriate input is not sufficient to cause these procedures to run their full course. The interesting issue then becomes: what other factors determine which procedures follow their course?)

The general point I am stressing here is this: mental modules in humans compete for energetic resources. Not all of them can operate simultaneously. This is true at all levels: perceptual, conceptual, and psychomotor. Contrast humans with simpler cognitive systems in this respect. Take a frog (or at least the idealized frog of philosophers—I am not making zoological claims). Here it sits, waiting for a fly moving within reach. There is no fly movement, no cognitive process other than the low-level monitoring of the visual field necessary to activate the get-the-fly module when appropriate. Is this a case of a wholly stimulus-driven module with mandatory operations? Almost, but not quite. Presumably the frog is also monitoring for possible predators and other dangers, and if a fly and a predator are sighted simultaneously, the operations of the get-the-fly module are preempted by those of the escape-the-predator module. This priority of the escape-the-predator module over all others (feeding and also mating modules) is clearly adaptive and is presumably built in. So the operations of the escape-the-predator module are fully mandatory, and those of the get-the-fly module are mandatory unless preempted. Frogs may well have a few more cognitive modules. Even so, it is plausible that the operations of each of them are mandatory except in the case of preemption, and that the order in which modules may preempt one another is fixed in the frog's nervous system. Moreover, cases of actual modular preemption are likely to be relatively rare (it is not very often that a frog is simultaneously presented with a possible prey, a possible predator, and a possible mate). The human predicament is quite different. If, as I have suggested, the human mind is teeming with modules, then, at all times, a number of modules have inputs available and must be competing for brain power to process them. Rather

than a fixed and global preemption order, which would not be adaptive in this case, some flexible, context-sensitive energy allocation procedure must be at work.

What should this energy allocation procedure be doing, that is, how might it contribute to the efficiency of the human cognitive system as a whole? Again, contrast with (philosophers') frogs. Presumably there are just a few categories of stimuli, such as flies, that frogs can discriminate, and only in restricted conditions. They monitor their environment to check whether any of these categories happen to be instantiated and then produce the prewired behavioral response. Humans can discriminate tens of thousands of categories in their environment, very few of which trigger automatic behavioral responses. At any one moment, humans are monitoring their environment through all their senses and establish perceptual contact with a great many potential inputs for further processing. Frogs have no memory to speak of. Humans have vast amounts of information stored in memory. When processing a new input, humans bring some of this stored information to bear on it. Attending to a given stimulus, activating memorized information, bringing the two together, and drawing inferences are effort-demanding mental activities. Effort is a cost that should be incurred only in the expectation of a benefit. Different trains of thought involve quite different evolving allocations of efforts and may produce quite different cognitive benefits.

What are the benefits of cognitive activity? The reply that comes most readily to mind is that cognition helps the organism recognize and react to opportunities and problems present in its environment; a more precise answer would consist in describing in much greater detail the various kinds of opportunities and problems that cognition helps the organism cope with. In the human case, a massive investment is made in cognition, and much knowledge is gathered, updated, and corrected without any specific practical goal. Presumably, what looks like—and often is—the pursuit of knowledge for its own sake helps prepare for an open range of future contingencies. Of course, knowledge is not equally pursued in all directions. Humans develop interests that guide their cognitive investments. Again, it seems, spelling out the benefit of cognition for humans would amount to describing in detail these diverse interests and possibly to explaining what makes their pursuit worth the effort. So, whereas it is natural to think of mental energy or effort in quantitative terms, one tends to approach cognitive benefit in qualitative terms. A philosopher might want to leave the matter there, but a psychologist cannot. The brain can be expected to allocate its energetic resources, not in a random but in a beneficial way. To achieve this, it does not have to be able to attribute an absolute value to the expected cognitive benefit of the processing of all available inputs, but it must be able to select, among the inputs and procedures actually competing for energy, some with relatively higher expected benefits.

Cognitive efficiency is a matter of investing effort in processing the right inputs. What are the right inputs? Do they have a characteristic property that the mind-brain can use in order to select them? Deirdre Wilson and I have argued that they do, and that this property is relevance, in a precise sense that we have tried to define and that I will briefly outline here (Sperber & Wilson, 1995; Wilson & Sperber, 2004).

Relevance is a property of inputs to cognitive processes. At a fairly abstract level, relevance can be defined *relative to an inferential procedure and a context*: a piece of information is relevant in a context for a given inferential procedure, if processing

the piece of information and the context together yields different conclusions from those that would be obtained by processing them separately. A bit more technically, a piece of information is relevant in a context for a given inferential procedure, just in case the set of conclusions that the inferential procedure derives from the union of this piece of information and the context, taken together as a single set of premises, is different from the union of the two sets of conclusions the inferential procedure would derive separately from the piece of information on the one hand and from the context on the other. For instance, if the procedure instantiates the elimination rules of propositional calculus, then (a) but not (b) is relevant in context (c):

(a) *p and r*

(b) *q and r*

(c) *{if p then s, if s then t}*

As can be easily verified, (a) in the context of (c) yields the two conclusions *s* and *t*, which are derivable neither from (a) alone nor from (c) alone, whereas (b) in the context of (c) yields no conclusions other than those derivable from (b) alone and from (c) alone.

This abstract definition is useful as a step toward defining relevance in a psychologically more pertinent way. A piece of information is relevant *to an individual at a time* only if there is a procedure and a context available to the individual at that time, relative to which the piece of information is relevant in the sense just proposed (this is just a necessary condition—for a fuller definition, see Sperber & Wilson, 1995, ch. 3).

Relevance is a property easily achieved: virtually any new piece of information that connects, however weakly, with what the individual already knows will be relevant by our definition. Relevance, however, is a matter of degree. Cognitive efficiency is not just a matter of processing relevant inputs; it is a matter of processing the most relevant inputs available. Everything else being equal, the greater the cognitive benefit yielded by the processing of an input, the greater its relevance. In addition—and this is quite specific to the approach taken by relevance theory—everything else being equal, the greater the cost of processing an input, the lesser its relevance. Here is a short artificial illustration. Being told by the doctor "You have flu" is likely to carry more cognitive effects, and therefore be more relevant, than being told "you are ill." Being told "you have flu" is also likely to be more relevant than being told "you have a disease spelled with the sixth, twelfth, and twenty-first letters of the alphabet," because the first of these two statements would yield the same cognitive effects as the second, but for less processing effort.

Cognitive efficiency, then, is a matter of maximizing the relevance of the inputs processed. There may well not be a unique way to maximize relevance and therefore to optimize cognitive efficiency. One input may be preferable to another in terms of benefits, the other in term of costs, and, in the absence of a common metric, there is no obvious way to decide between the two. Still, as long as some inputs are clearly more relevant and therefore preferable to others, it should be possible to enhance cognitive efficiency through input selection. In other words,

we should not expect the system to do more than tend to optimize. But how can even this be achieved? To try and answer, I will look first at costs, and then at benefits and then will put the two together.

How can the brain optimally allocate energy? The solution could, in principle, be a cognitive one. That is, the brain could represent its own energy consumption, compute the expected cost of various procedures, and use this as a criterion in deciding how much to invest in each procedure. In other words, the brain might automatically be taking, every fraction of a second, decisions similar to those we consciously take once in a while when, for instance, we choose to save our effort by using a pocket calculator rather than perform a mental calculation. Note, however, that this cognitive way of minimizing the energetic costs of cognitive processes would involve a significant cost of its own, which might make it self- defeating.

Are there noncognitive ways of minimizing effort in mental processes? Consider the comparable problem of minimizing energy consumption in muscular move-ment. Muscles get their energy from chemical reactions. This energy can be converted into work or into heat. The efficiency of the process (except when the function of the movement is to provide heat, as when shivering) depends on letting as little energy as possible degrade into heat. These local chemical reactions depend on a supply of oxygen and nutrients from blood vessels, a supply that has its own energy costs and that can be insufficient or excessive for optimal efficiency. Blood vessels also have the function of removing carbon dioxide and waste products such as lactate. The removal of lactate from the muscle is slower than its production, causing, in case of prolonged use of the muscle, a perception of fatigue. Only above this threshold is muscular effort *represented* in the cognitive system—and even then in a very coarse manner—often inducing intentional reallocation of muscular effort. The regulation of effort—the production of the right quantity of energy in muscle tissue, the adjustment of blood flow, and so on—is otherwise achieved not through computations over representations but through noncognitive physiological proce-dures that, one may assume, are to a very large extent genetically specified. I suggest that the regulation of effort in cognitive processes is likewise achieved, for the most part, through noncognitive brain processes that are also largely genetically specified.

That the flow of energy in the brain is guided by noncognitive mechanisms may seem easy enough to accept. Isn't it just an aspect of the neurological im-plementation of cognitive processes? How could this be relevant to an under-standing of cognition at a computational or algorithmic level, to use Marr's popular distinction? Well, I will argue that the regulation of this energy flow has cognitive, even epistemic, consequences.

Understanding how the brain is sensitive to the cost of various procedures may be difficult. Even more difficult is understanding how the brain could be sensitive to the size of the cognitive benefits resulting from the processing of various inputs.

To begin with, how can the brain distinguish, among all the cognitive changes that might be brought about by cognitive operations, those that are beneficial and those that are not, and which may even be costly (for instance, mistaken infer-ences)? Well, the brain has no other choice than to trust itself and be, so to speak, optimistic about its own procedures. That is, it should behave in a way consistent with the presumption that, in general, its perceptions are veridical and its inferences

rational. In normal conditions, the processing of new inputs yields positive cognitive effects, that is, it results in an improvement of the individual's knowledge of her world, be it by adding new pieces of knowledge, updating or revising old ones, updating degrees of subjective probability in a way sensitive to new evidence, or merely reorganizing existing knowledge so as to facilitate future use. There are many exceptions, of course—cases where less processing would have resulted in better knowledge— but procedures that have tended to produce more negative than positive cognitive effects are likely to have been selected out. The relevance of this is that the brain would be roughly right in treating any and every cognitive effect as a positive effect, in other words, as a cognitive benefit.

But then what? Supposing it treats all cognitive effects as cognitive benefits, how could the brain then calculate the size of these cognitive effects? Should it count the number of conclusions arrived at? Should it treat the value of each conclusion as depending on its complexity? Should it multiply the value of each conclusion by its subjective probability? Should it give greater value (and how much greater?) to conclusions that have practical consequences, or relate to standing interests? How should it evaluate revisions of previous beliefs? And so on. Or are these even the right questions? Actually, it is not at all obvious that the brain should *calculate* the size of cognitive effects. There may be physiological indicators of the size of cognitive effects in the form of patterns of chemical or electrical activity at specific locations in the brain. A module receives some degree of activation from other modules with which it is connected. It is activated by upstream feeder modules that present it with inputs. It may be activated by downstream client modules that are already mobilized and that would benefit from receiving new or further inputs from it. Suppose that these physiological indicators locally determine the ongoing allocation of brain energy to the processing of specific inputs. These indicators may be coarse. Nevertheless, they may be sufficient to cause energy to flow toward those processes that are likely to generate relatively greater cognitive effects at a given time. In other words, just as effort need not be computed, cognitive effect need not be computed either, and both effort and effect factors may steer the train of our thoughts without themselves being thought about at all.

Someone might object: suppose there are physiological indicators of effort and effect. All they can indicate are past or current effort and effect, whereas what should guide the allocation of brain resources is *expected* effort and effect.[4] Answer: It is not true that indicators can only indicate past and present states of affairs. Dark clouds may indicate that rain is probable. The current level of lactate concentration in a muscle may indicate that the muscle cannot continue to perform the same amount of work for long. The differences in the patterns of activity of two competing cognitive processes may indicate which has the highest expected cognitive utility. Suppose the processing of inputs A and B are both currently producing the same

4. As with "expected utility" in expected utility theory, I am speaking of "expected relevance" without presupposing a cognitive process involving the formation of mentally represented expectations. In fact, I am arguing that people tend to maximize expected relevance without, in most cases, representing it.

level of effect, but the processing of A is producing these effects with greater effort. Or suppose the processing of inputs A and B are both currently requiring the same level of effort, but the processing of B is resulting in greater effect. Of course, it is impossible to be sure how things would evolve, but in both cases, a greater cognitive utility should be expected from the continued processing of B rather than A. A better indication still may be given by the direction in which levels of effect and effort are moving. If the processing of inputs A and B are producing the same amount of effect for the same amount of effort, but the amount of effect produced by the processing of A is on the decrease, whereas that of B is constant or on the increase, or if the amount of effort required by the processing of A is on the increase and that of B constant or on the decrease, then, again, greater cognitive utility should be expected from the continued processing of B rather than A.

If we look at the issue in an evolutionary perspective, what does all this mean? Imagine a species investing more and more in cognition, monitoring in a more and more fine-grained way more and more aspects of the environment, constructing an ever richer memory, and achieving this by use of an ever greater variety of perceptual and conceptual modules. The result would be a kind of attentional bottleneck: only very few of the available inputs could be treated attentionally, and only very limited background information could be brought to bear on the treatment of these inputs. This bottleneck would in turn create a strong and constant selective pressure for optimizing the choice of inputs to be processed, which, in the picture I am presenting, is equivalent to optimizing the allocation of energy to modules. Such a selective pressure should result in the evolution of a variety of traits contributing to an optimal allocation. I am not excluding the possibility that, among these traits, there may be mental devices directly involved in internal administration of resources, but I find it implausible, both for evolutionary and efficiency reasons, to imagine that this allocation of resources might be wholly or even mostly controlled by some central specialized device. For the same kind of reasons that, whether we like it or not, market economies work better than centrally managed ones, competition for resources among modules seems more likely to yield good results than centrally controlled allocation.

There are a wide variety of small changes in the functioning and articulation of modules that may each have contributed to improving the allocation of resources over evolutionary time, or that may contribute to it in cognitive development. These include, as I have already suggested, the use of simple and approximate indicators of the ongoing and expected expenditure of energy, and of the ongoing and expected cognitive impact of specific procedures.

Different modules may be more or less easily mobilized in a way that reflects their general contribution to relevance. Modules that are specialized in processing inputs with high cognitive impact in the history of the species (and in particular with high practical impact) should be given a greater initial claim on brain resources, with the possibility of preempting other procedures in a bottom-up fashion (as the literature on attention shows us is typically the case with, for instance, potential danger signals). (Incidentally, given that the human environment changes much faster than the human genome, this may occasionally have counteradaptive results. For instance, people living in an urban environment are uselessly startled by

all-too-frequent sudden loud noises that would have deserved immediate attention in an ancestral environment.)

Inputs pertaining to an area of stable interest developed by the individual benefit from richer intramodular databases and from richer intermodular connections (the two ways in which richer background information is realized in a modular system). Modules processing such inputs should therefore be given a greater claim on energetic resources and mobilise more easily.

Inputs pertaining to ongoing cognitive processes also benefit, *ceteris paribus*, from a greater claim on resources, this time because of quantitative factors on the effort side: the devices and data needed to process these inputs are already mobilized, and therefore their processing is less costly than the processing of inputs for which inactive or less active devices must be given energy. Thus relevance to current cognitive activity is, *ceteris paribus*, greater relevance.

More generally, there are many different ways, some obvious, others still to be discovered, in which a massively modular system might improve the allocation of its energetic resources among its modules, doing so much better than random allocation. Some of the traits of human cognitive organization that tend to optimize relevance have emerged in the evolution of the species. Others emerge in cognitive development and throughout the cognitive life of the individual. These lifetime improvements are themselves made possible by the flexibility of the evolved modular system of human cognition. This flexibility, therefore, should not be seen as a mere ability to adjust cognitive capacities to the demands and opportunities of different environments. It also helps maximize the relevance achieved by ongoing cognitive processes. Flexibility, that is, long-term context-sensitivity, makes a critical contribution to short-term context-sensitivity.

4   Conclusion

The claim that the human cognitive system tends to allocate resources to the processing of available inputs according to their expected relevance is at the basis of relevance theory (where it constitutes the first, cognitive principle of relevance).[5] The main thesis of this chapter has been that this allocation can be achieved without

---

5. The cognitive principle of relevance has experimentally testable consequences, some of which are reviewed in Van der Henst & Sperber (2004). We have shown, for instance, with experiments on relational reasoning, that by manipulating contextual factors, people can be made either to derive logical implications from a given set of premises or to say that nothing follows from it (Van der Henst, Sperber, & Politzer, 2002). What the context does in this case, we claim, is raise or lower expectations of relevance that attach to the premises presented, thus triggering or, on the contrary, inhibiting, an inferential procedure. With experiments on the Wason selection task, we have shown that, by manipulating contextual factors, people can be made to apply one or other of several possible inferential procedures involved in the interpretation of conditionals and therefore to reach different conclusions from the same set of conditional premises (Sperber et al., 1995a; Girotto et al., 2001). What the context does in this case, we claim, is raise or lower expectations of relevance that attach to each of these procedures in their application to the premises. These experiments illustrate the main thesis of this chapter.

computing expected relevance. When an input meets the input condition of a given modular procedure, this gives this procedure some initial level of activation. Input-activated procedures are in competition for the energy resources that would allow them to follow their full course. What determines which of the procedures in competition get sufficient resources to trigger their full operation is the dynamics of their activation. These dynamics depend both on the prior degree of mobilization of a modular procedure and on the activation that propagates from other active modules. It is also quite conceivable that the mobilization of some procedures has inhibitory effects on some other procedures. The relevance-theoretic claim is that, at every instant, these dynamics of activation provide rough physiological indicators of expected relevance. The flow of energy in the system is locally regulated by these indicators. As a result, those input-procedure combinations that have the greatest expected relevance are the more likely ones to receive sufficient energy to follow their course. This is just a tendency, but it is strong enough to yield the kind of context-sensitivity that humans actually exhibit in their individual cognitive processes.[6]

I am well aware of the vague and speculative nature of the view outlined in this chapter. It calls both for greater empirical anchoring and for formal modeling. I nevertheless feel justified in putting forward this view, as it is supported by, paradoxically, an argument of Fodor himself. He writes:

> 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. . . . There 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*. (Fodor 2000, p. 23)

This should be a strong vindication of the massive modularity thesis. Fodor, however, goes on to say: "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" (p. 23). His main reason for claiming that the thesis is not true is the alleged inability of a massively modular system to exhibit context-sensitivity. This is why it seemed worth explaining, however tentatively, how such a system might be context-sensitive, contrary to Fodor's claim. Since the massive modularity thesis *might* be true, we can keep exploring "the only remotely plausible theory of the cognitive mind that we've got so far," and that, surely, is good news.

---

6. In collective intellectual endeavors that are pursued over generations, and in science in particular, greater context sensitivity and greater relevance can be achieved, but these achievements cannot be explained just by individual cognitive psychology, and, contrary to what Fodor tends to do, should not be taken as a benchmark to assess models of human cognition (Sperber & Wilson, 1996). Rather, the explanation of these achievements calls for a kind of epidemiology of representations that looks at the effect of the causal chaining of individual cognitive processes across populations (Sperber, 1996).