Dan Sperber

Massive Modularity and the First Principle of Relevance

(Incomplete draft)

Introduction: Two questions:

- How can a massively modular mind be flexible?
- How can a massively modular mind be context-sensitive?

1. Modularity and flexibility

What would be a non-modular mind ?

- "Classical" view:
    - A central computing device
    - A data-base

- Connectionist view:
    - A single neural network

Modules are characterized by

- Specific inputs
- Proprietary resources
    - procedures
    - data-base

Modularity in a connectionist perspective

[Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness. A connectionist perspective on development*] distinguish among different things that can be innate in a neural network: the connection weights…, architectural constraints…, and chronotopic constraints…. One could also add that there may genetically inherited constraints on [connections weights], for example their maximum value or their "sign" (for excitatory or inhibitory connections) may be genetically specified or the genotype may encode the value of learning parameters such as the learning rate and momentum

> Raffaele Calabretta and Domenico Parisi (in press) "*Evolutionary Connectionism and Mind/Brain Modularity*"

Modularity in evolution and development

- Evolved modules may be more or less fully specified:
  Comparison between visual cliff, Garcia effect, face recognition, language faculty

Many modules are "module templates" that generate modules through initialisation, i.e.

- Parameters fixing
- Slots filling

- Module templates can be repeatedly projected into new modules
  (examples of imprinting, face recognition, language, living kind concepts)

Modularity and modularisability

There may be a continuum of cases between fully specified modules and preparedness for modularisation

Example: speech/comprehension vs. writing/reading

Higher-level or metarepresentational modules

Metarepresentational modules (e.g. ToM, verbal comprehension) provide a kind of virtual domain-generality, while being strictly modular

Flexibility of a massively modular mind

All the above properties show how a truly modular mind can be flexible

[NB: This first part goes over grounds I have covered before, and could therefore be shortened, or even reduced to an introduction, if the second part, which is novel, gets sufficiently developed, or if a relatively short paper is OK]

- <u>How can a massively modular mind be context-sensitive?</u>

[Here I will start from Fodor's central argument that a modular mind cannot be context sensitive. What will follow in in part written in a very drafty form. Here it comes:]

Few if any of the operations of the human mind are mandatory. Here, for instance, is a pair of numbers:

17 68

You are equipped with procedures to derive a variety of conclusions about such a pair. For instance:

(a) The second number is bigger than the first

(b) Both are two-digit numbers

(c) The sum of the two is 85

(d) The difference between the two is 51

(e) The first number is one fourth of the second one

Chances are that, when you read these numbers, only conclusions (a) and (b), if any, came to your mind. If you had been told that these numbers were two amounts of money you had just won, you might have computed the sum. If you had been told they were the ages of a wife and her husband, you might have computed the difference. If you had been told that 17 was the number of syllables in each verse of a 68 syllables poem, you might have computed the ratio. In the absence of such incentives, you probably did not derive these conclusions. Clearly, we don't derive all the inferences we could from the inputs that are active in our mind at a given time. It is obvious anyhow that we could not derive jointly all the inferences we can derive singly if only because the conclusions of any inferential procedure (apart from end-of-the line practical inferences that have motor command as their output) could serve as premises in further inferences, and so ad infinitum. So, which inferences we actually perform and where we stop in a series of inferences must somehow be regulated. Fodor (1983) has argued that the operations of mental modules are mandatory. Doesn't the example we have just considered show that inferential processes are not modular (unlike input level processes which, according to Fodor, are)?

However, even the operations of input modules are far from being mandatory. Yes, if I see just in front of me, in broad daylight, the face of my Paris neighbour, the painter Gérard, I cannot help but recognise him. My face recognition module does its job. But suppose I am lecturing in London. Some thirty faces in front of me are each clearly visible. I look cursorily at all of them and I do recognise some colleagues. Even though I had looked at his face as much as at that of the people I had immediately recognised, it is only towards the end of the lecture that I suddenly recognise, sitting there on the second row, Gérard, whom I would never have expected to see in such a place.

The operations of input modules seem mandatory when you just consider cases where the stimulus is, and stays long enough, at the centre of attention. There are cases however, with most if not all input modules, where a stimulus is well within the field of perception but 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: the face is not recognised, the sentence structure is not parsed, the colour is not identified. 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 thought, the stimulus would have been processed. The general point is this: humans' mental modules compete for resources. Not all of them can operate simultaneously. This is true at all levels: perceptual, conceptual, psychomotor. As I will soon argue, the allocation of resources among modules has wide-ranging cognitive and epistemic consequences.

Contrasts humans with simpler cognitive systems. Take a frog (or at least the idealised frog of philosophers – I am not making zoological claims). Here it sits waiting for a fly moving within reach. 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 then a case of a wholly stimulus-driven module with mandatory operations? 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 pre-empted 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 pre-empted. Frogs may well have a few more modules, but, even so, it is plausible that the operations of each of them are mandatory except in the case of pre-emption, and that the order in which modules may pre-empt one another is fixed in the frog's nervous system. Moreover, cases of actual modular pre-emption are likely to relatively rare (not that often is our frog simultaneously presented with a possible meal, a possible predator, and a possible mate).

If, as I have suggested, the human mind is teeming with modules, then, at all time, a number of modules have available inputs and must be competing for brain power to process them. Rather than a fixed and global pre-emption order which would not be adaptive in this case, some flexible, context-sensitive resource allocation procedure must be at work.

What should this resource allocation procedure be doing, that is, how might it contribute to the efficiency of the human cognitive system as a whole? Again, compare with 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 behavioural response. Humans can discriminate tens of thousands of categories in their environment very few of which trigger automatic behavioural 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 their memory. When processing a new input they bring to bear on it some of this stored information. Attending to a given stimulus, activating memorised information, bringing the two together and drawing inferences form their union are effort-demanding mental activities. Effort is a cost that should be incurred only in the expectation of a benefit. Different course of thought involve quite different ongoing allocation of resources 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 recognise opportunities and problems present in its environment and react to them; 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 prepares for an open range of future contingencies. Of course, knowledge is not pursued in all directions. Humans develop interests that guide their cognitive investments. Again it seems, spelling out the benefit of cognition for human would amount to describing in detail these diverse interests and possibly in explaining what makes their pursuit worth the effort. So, whereas it is natural to think of mental effort in a quantitative manner, 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 resources not in a random but in a beneficial way. For this, it needn't 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 resources, some with relatively high expected benefits.

How could the brain invest its resources in the processing of inputs likely to yield higher cognitive benefits? To begin with, the brain should 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 reorganising existing knowledge so as to facilitate future use. There are many exceptions of course, but procedures that 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 terms as a cognitive benefit.

Some cognitive effects are larger than others. PARAGRAPH TO BE DEVELOPED WITH EXAMPLES.

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 tried to define and that I will outline again here. Relevance is a relational property. It can be defined in relationship to an inferential procedure and a context: a piece of information is relevant in a context for a given inferential procedure just in case the conclusions that the inferential procedure derives from this piece of information and the context taken together as a single set of premises are different from the conclusions the inferential procedure derives from the piece of information on the one hand and from the context on the other taken as two separate sets of premises. 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

Here are the different pertinent sets of conclusions:

(conclusions of a) p, r

(conclusions of b) q, r

(conclusions of c) if p then t

(conclusions of a and c) p, r, if p then t, s, t

(conclusions of b and c) q, r, if p then t

As can be seen, (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 of (b) and those of (c).

Relevance can also be defined relative to an individual at a time. 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 and relative to which the piece of information is relevant in the previous sense considered (this is just a necessary condition – for a fuller definition, see Sperber & Wilson 1995, chapter 3).

Relevance is a property easily achieved: any new piece of information that connects, however weakly, with what the individual already knows will be relevant by our definition. Cognitive efficiency is not a matter of just processing relevant inputs, it is a matter of processing the most relevant inputs available. Relevance is a matter of degree. Everything else being equal, the greater the cognitive benefit yielded by the processing of an input, the greater its relevance. Also – and this is quite specific to relevance theory's approach– everything else being equal, the greater the processing cost of an input, the lesser its relevance.

Cognitive efficiency, then, is a matter of maximising the relevance of the inputs processed. There may well not be a unique way to maximise relevance and therefore to optimise 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, it is enough that some inputs be clearly more relevant and therefore preferable to others for it to be possible to enhance cognitive efficiency through input selection. In other words, we should not expect the system to do more than tend to optimise. But how can even this be achieved? To try and answer, I will look first at costs, then at benefits, and then will put the two together.

How can the brain allocate its resources so as to minimise its consumption of energy or its efforts (I am using the two expressions as synonyms)? 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 terms, the brain might be automatically taking, every fraction of a second, decisions similar to those we consciously take once in a while when we choose, for instance, to use a pocket calculator rather than perform a mental calculus that would take too much effort. Note however that this cognitive way of minimising the energetic costs of cognitive processes would involve a significant cost of its own which might render it self-defeating.

Are there non-cognitive ways of minimising effort in mental processes? Consider the comparable problem of minimising energy consumption in muscular movement. 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 supply of oxygen and nutrient by blood vessels, a supply which has its own energy costs and which therefore can be insufficient or excessive for optimal efficiency. Blood vessels have also the function of removing carbondioxide 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 central nervous system – and even then in a very coarse manner –, causing it to modify its demands on the muscles. 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 non-cognitive physiological procedures which, 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 non-cognitive brain processes that are also largely genetically specified.

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

Understanding how the brain is sensitive to t cost of various procedure may be difficult. Even more difficult is understanding how the brain could calculate the size of the cognitive effects resulting from the processing of some input. Should it count the number of conclusions arrived at? Should it ponder the value of each conclusion in terms of 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 having practical consequences, or relating to standing interests? How should it evaluate revisions of previous beliefs? And so on. But are these the right questions? Actually, it not at all obvious *that the brain should calculate* the size of cognitive effects. Suppose that there are physiological indicators of the size of cognitive effects, such as chemical changes or patterns of electrical activity at specific locations in the brain, and that these indicators influence the allocation of brain resources to the processing of specific inputs. 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.

Before proceeding, I must answer an obvious objection: Say there are physiological indicators of effort and effect. All they can indicate, so the objection goes, are past or current effort and effect, whereas what should guide the allocation of brain resources is *expected* effort and effect. Answer: It is not true that indicators can indicate only past and present state of affairs. Dark clouds may indicate that rain is probable. The current level of lactate concentration in a muscle may indicate that it cannot continue for long to perform the same amount of work. 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 level of effect but the processing of A does so with greater effort. Or suppose the processing of inputs A and B are both currently expending the same level of effort, but the processing of B does so to 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 continuation of the continuation of the processing of B than from that of A. A better indication still may be given by the direction in which effect and effort level 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 expended 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 continuation of the processing of B procedure than from that A.

[The paper goes on to explain how the selection procedure envisaged would typically be context sensitive, selecting inputs with the greatest expected relevance in the situation. It describe how this would have cognitive and epistemic effects. The general idea then is that a massively modular computational system combined with a non-representational, non-computational physiological or metabolic input-selection system would be a modular system with just te right kind of context sensitivity. There will also be some experimental evidence]