Modularity and the Possibility of a Cognitive Neuroscience of Central Systems

Vinod Goel
Jordan Grafman
Cognitive Neuroscience Section
Medical Neurology Branch
Bldg. 10; Rm 5S209
NINDS/NIH
Bethesda, MD 20892
goel@helix.nih.gov
jgr@cu.nih.gov

Abstract

The methodology of cognitive neuroscience presupposes that cognitive functions are modular. Fodor (1983) offered an interesting characterization of various forms of modularity and an argument to the effect that while language and input systems are probably modular, higher cognitive processes such as problem solving probably are not. If this is the case, there will be methodological obstacles in developing a neuroscience of higher cognitive functions. We offer an analysis of the issue of modularity as it affects the cognitive sciences, evaluate Fodor's characterization with respect to this analysis, and suggest that his argument for the nonmodularity of central systems has a very narrow scope. It is not something that neuroscience needs to *necessarily* worry about.

Introduction

In the *Modularity of Mind*, Jerry Fodor (1983) offered a wonderful and lively exposition of modularity in cognitive systems and argued that there are compelling reasons to expect input systems and language to be modular, and equally compelling reasons to expect the central systems involved in higher cognitive functions not to be modular. This latter argument lead Fodor to a rather gloomy conclusion about the prospects of a neuroscience of central systems.

There is considerable consensus that cognitive neuroscience does need *some* notion of modularity (McCarthy & Warrington, 1990; Shallice, 1988; Vallar, 1991).[1] The methodology of cognitive neuroscience requires a mapping of behavioral deficits onto computational procedures and/or cortical structures. The way to do this is to show that damage to a specific region a leads to a specific behavioral deficit x but leaves other behavior x' intact (single dissociation), while damage to cortical region a' leads to specific behavioral deficit x' but leaves x intact (double dissociation). A search for dissociations is of course a search for modules. So, much hangs on the modularity issue.

A few neuropsychologists have taken up the challenge posed by Fodor (Marshall, 1984; Shallice, 1984; Shallice, 1988) but no adequate analysis of the issues or a coherent response have been forthcoming from the neuroscience community. Such an analysis and response is attempted here.

At first pass, the issue of modularity seems to be reasonably clear and unproblematic. The underlying intuition is that, while the world may be such that:

All things by immortal power,
Near or far,
Hiddenly,
To each other linked are,
That thou canst not stir a flower
Without troubling of a star.[2]

It is also true that the degree, intensity, and number of "links" among "things" vary. This heterogeneity in the distribution of links or relations results in stable clusters which we recognize as subsystems or modules.

Another way of saying this is that the world seems to be constructed of complex systems that can be decomposed – *to some extent* – into stable subsystems such that inter-system interactions are stronger (more robust, abundant, coherent, etc.) than intra-system interactions. In fact, it is this varying degree of relations/ interactions – along with our goals and interests – that allows us to individuate systems/subsystems in the first place. Individuated as such, systems correspond to modules.

How pervasive a phenomenon decomposability, or modularity, turns out to be is an open empirical question but there is converging evidence from at least three sources: the natural sciences, design/ engineering practice, and arguments about the evolution of complex systems (Simon, 1962; Simon, 1973), suggesting that it is a deep fact about the world.

[1] See Farah (in press) for an alternative view.

[2] Francis Thompson

So far so good. The main problem that is typically encountered in the individuation of systems is that relations/interactions among components can be of many types, including spatial, temporal, logical, conceptual, causal, etc. In fact relations are down right cheap, making the notion of systems (and thus modules) equally cheap. But for the purposes of natural science, we generally restrict ourselves to a subset of all possible relations, namely the ones that are *causally efficacious* in the functioning of the system *under the aspect that we are interested in the system.* There is generally just one such aspect (per discipline) and the causal properties involved are of a pedestrian sort. This is usually sufficient to render the notion of systems and modules substantive and useful.

However, when we try to apply this seemingly simple concept to the study of human *cognitive systems*, a number of complications quickly arise. Some of the complicating factors are the following: First of all, we are interested in cognitive systems under three different aspects; under the aspect of *functions*, of *computational procedures*, and *computational mechanisms.*[3] Secondly, cognitive neuroscience is far from discovering the relevant causal relations to carve cognitive systems at their "natural joints." So we are relegated to individuating cognitive systems functionally based on behavior. Unlike the natural sciences, we cannot, in the first instance, appeal to causal relations to guide and constrain our individuations. This leaves our functional individuations radically under-constrained. So, a major issue for us is "what are the correct or coherent functional individuations or systems?" One way of testing functional individuations is to see how deep they run? That is, will they survive a shift to an aspect where causal relations exist, or will this shift result in cross-classification? If the latter occurs, we would be justified in questioning the functional individuation. Thirdly, system decomposition is hierarchical and occurs along two dimensions. These factors require some sorting out.

Individuation of Cognitive Systems & the Resulting Notions of Modularity

Systems as Functions

Individuation of Functions: Cognitive neuroscience is interested in cognitive systems under several different aspects. One such aspect is that of *functions.* So the problem of individuating cognitive systems becomes, on the first instance, the problem of individuating functions. Strictly speaking, functions are *rules of correspondence* that *map* elements of a *domain* set onto unique *images* in a *range* set. They are typically individuated by specifying the set of ordered pairs of domain elements and their corresponding images in the range. The set of ordered pairs are both necessary and sufficient to define the function.

However, given different goals, there are other possibilities for individuating functions. Consider the cognitive functions **I, I', I''**, and **I'''**. **I** maps from domain D (which consists of fragments of natural language utterances) to the range R (which consists of the ability to understand/generate novel sentences of that language). **I'** maps from domain D' (which consists of fragments of set theory) to the range R' (which consists of the ability to prove previously unencountered theorems). **I''** maps from domain D'' (which consists of light energy values impinging on the retina) to the range R'' (which consists of a visual experience). **I'''** maps from domain D (which consists of fragments of natural language utterances) to the range R'' (which consists of a visual experience).

Given that $D \neq D' \neq D''$ and $R \neq R' \neq R''$ the resulting ordered pairs will not be equivalent. So on the standard mathematical individuation by domain and range, **I, I', I''**, and **I'''** constitute four different functions. However, as cognitive scientists, we may find it interesting to individuate simply by the domain or the range. If we individuate by the domain we find that function **I** is indistinguishable from the function **I'''**, leaving us with three distinct functions. If we individuate by the range, we find that function **I''** is indistinguishable from **I'''**, again leaving us with three (but different) functions. So, for the purposes of cognitive science, there are several interesting ways of individuating functions.

Notion of Modularity: The second question is what notion(s) of modularity do functions allow for? That is, can we talk of modularity of cognitive functions? Insofar as functions are mathematical entities, it is rather odd to ask under what conditions they are modular. It is simply a category mistake. But one can get a corresponding notion in terms of (relative) disjointness of the sets of inputs, outputs, or input/output pairs, depending on your preferred way of individuation. On this account, functions that result in (largely) non-overlapping sets could be considered "modular" with respect to each other. Notice, however, that this is trivially true by definition of the function. It does not constitute a substantive claim about the world. The substantive issue in terms of individuating functions is not whether they are modular – they are modular by definition – but whether the subsystems which result when these functions are mapped onto computational procedures and/or mechanisms are modular.

Systems as Computational Procedures

Individuation of Procedures: A second aspect under which cognitive science is interested in cognitive systems is that of computational procedures. (It is not,

[3] What we are referring to as "different aspects" are generally referred to as "different levels." The term "aspects" is borrowed from the philosophy literature and is introduced to avoid confusion later on.

however, clear whether cognitive *neuro*science also needs to be interested in functions under the aspect of computational procedures. But this issue is beyond the scope of the present discussion.) That is, we are interested not only in functions, but also the computation of functions, thus we also find it useful to individuate functions by the elementary operations/computations involved in effecting the mapping. This criterion results in yet another set of categories.

Re-examining the above cognitive functions with this criterion we find that it is an open question whether we have one, two, three, or four distinct functions. It could be that some general induction operations/procedures effect all (or several) of the mappings, or it could be that each function requires specialized procedures.

Notions of Modularity: The notion of modularity does apply in a natural sense to computational procedures, though with a slight twist. We began the discussion by noting that modules (subsystems) emerge from varying degrees of interactions/relations among component parts of a system, and that the interactions/relations generally of interest are the ones that are causally efficacious in the behavior of the system. The twist is that, in computational procedures, what is causally efficacious is semantic content or the flow of information. So the notion of modularity in computational procedures has to do with restrictions on the flow of information among procedures as opposed to the flow of physical effect.

Systems as Mechanisms

Cognitive neuroscience is also interested in systems under a third aspect: as mechanisms that realize the procedures which compute the functions. Mechanical systems are individuated along the lines of brute physical causation (as opposed to semantic causation) and modules result whenever there are varying degrees of causal interaction among components.

So far we have three different aspects under which we are interested in cognitive systems, different ways of individuation of systems under these aspects, and several different notions of modularity. But we are not done yet. Things are more complicated still.

Two Dimensions of Modularity

Another complicating factor is that system decomposition (and integration) is generally hierarchical, and occurs along two dimensions. Different relations hold along the two dimensions and result in different notions of modularity. The relations that hold between elements at the same level in the hierarchy, i.e. between siblings or daughter nodes, might be referred to as dRd relations. The relation that holds between micro and macro levels, i.e. between mother and daughter nodes, might be designated mRd relations. Exactly what these relations are of, course, depends on the types of systems involved.

Both computational procedures and mechanisms can, in principle, be decomposed along the two dimensions and exhibit both micro/macro (mRd) relations and within-level (dRd) relations. The micro/macro (mRd) relation can be particularly varied and complex. For systems under the aspect of mechanisms, some candidates are emergence (e.g. the relation between water molecules and liquidity), part of (e.g. the relation between a wheel and an automobile), and composition (e.g. the relation between the mass of a micro and macro component). Candidates for the within-level (dRd) relation for mechanisms might be predicates like "near," "in contact with", etc. While the specific relations are very important in analysis of specific systems, the relevant point for our purposes is that the two sets are generally different.

For systems individuated under the aspect of computational procedures, some candidates for the mRd relations are subset/superset, set membership, and conceptual containment (Note that the relation of *implementation* is *not* a candidate for the mRd relation. It cuts across aspects, not levels, and the issue of levels is independent of the issue of aspects.) A candidate for the dRd relation might be the predicate "next element on the list". Again, the important point for our purposes is that the two sets are generally different from each other, and from the above sets of relations among mechanisms.

Given the different character of the mRd and dRd relations, we can specify different module types (M and M') as follows:

- (M) M (within-level) modularity occurs when the relevant mRd relations that a node has with daughter nodes are "stronger" ("denser," "richer") than the relevant dRd relations it has with sibling nodes. (We might denote this in the following manner: $\Sigma mRd\downarrow > \Sigma dRd$.)
- (M') M' (micro-macro) modularity occurs when the sum of the relevant dRd relations that a node has with sibling nodes are "stronger" ("denser," "richer") than the relevant mRd relation(s) it has with its mother node(s). (We can denote this as follows: $\Sigma dRd > \Sigma mRd\uparrow$). This is guaranteed to be the case when the decomposition is a tree structure because each node (except for the root) will have a single unique predecessor.[4]

Furthermore, as the sets of relations across computational procedures and mechanisms are different, we actually get four different types or notions of modules, which we might designate M-p, M-m, M'-p, and M'-m (see Table 1). M-p and M'-p are the modules which occur when the above definitions of M and M' are applied to procedures; M-m and M'-m modules occur when M and M' are applied to mechanisms. There is also the notion of "modularity" we get in the case of

[4]Talk of "summing up relations", here and throughout the paper is, of course, figurative.

functions (disjointness of sets), which we might – for the lack of a better term – refer to as set-theoretic (S-t). Each of these are noted in Table 1.

Table 1:
Possible notions of modularity in cognitive systems

	Types of Modularity		
	Within-Level	Micro/ Macro	Other
Systems as Functions	N/A	N/A	S-t
Systems as Computational Procedures	M-p	M'-p	N/A
Systems as Mechanisms	M-m	M'-m	N/A

It is worth noting that these are not mutually exclusive categories. The presence of any one does not preclude the presence of any other.

By way of an example of the M and M' categories, consider the (incomplete) functional decomposition of a computer in Figure 1. The relations between level L1 components CPU, memory, and Input/Output units, or the level L2 components ALU, control, and registers, are within-level or dRd relations. The claim that these components constitute an M type module or subsystem with respect to the other components within the level is the claim that the mRd relations that they engage in with daughter nodes at level L_{n+1} are stronger than the dRd relations that they engage in with sibling nodes at level L_n.

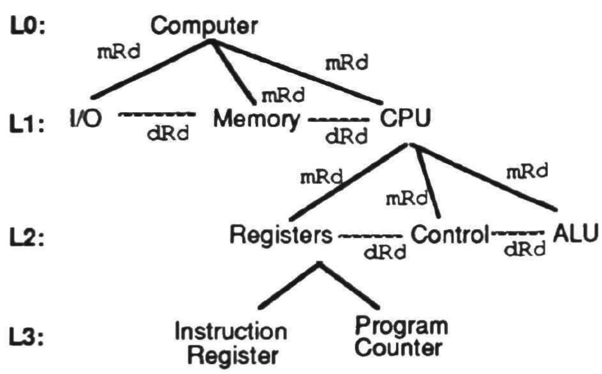

Figure 1. Incomplete functional decomposition of a computer

The relationships between the memory unit at L1 and the computer unit at L0, or the control unit at L2 and the CPU, at L1, or in fact the control unit at L2 and the computer unit at L0 are examples of micro-macro relations or mRd relations. To claim that these components constitute M' modules with respect to each other is to claim that the sum of the dRd relations that they engage in with sibling nodes is greater than the single mRd relation that they have with their mother node.

So what started out as a simple, intuitive idea has been complicated by the facts that we are interested in cognitive systems under, at least, three different aspects, have two different notions of causation, and must accommodate both within-level and micro/macro level notions of modularity.

Reviewing & Assessing Fodor's Discussion of Modularity

Having analyzed some of the relevant landscape, we now turn to Fodor (1983). We briefly review his analysis, map it onto ours, and then examine his argument that central systems are probably not modular.

Fodor suggests that talk of modules or faculties, divides up, at a first instance, into two categories, *horizontal* faculties and *vertical* faculties. The first thing to note is that the notion of "faculty" is very coarse. It cross-cuts both what we have been calling aspects of systems and modularity. There is something to be gained by slicing thinner.

Horizontal faculties are individuated functionally by their effects. Typical examples of such faculties would be memory and attention. Importantly, this individuation is independent of the content of the input to the faculty. So horizontal faculties are invariant across domains (e.g. the same memory faculty gets exercised in remembering your mother's face as remembering the 3rd movement of Brahm's *German Requiem)* and each horizontal faculty can, in principle, have access to every mental content at some point in time. Different cognitive processes are explained by varying degrees of interactions among horizontal faculties. It would be a mistake, on this account, to look for a language faculty or a music faculty. These are simply different manifestations/ interactions of the same horizontal faculties.

In our terminology, what Fodor is calling horizontal faculties are, in the first instance, a specification of systems as functions at the level of memory, learning functions, attention, etc., individuated by outputs or effects. The substantive claim associated with horizontal faculties is that this individuation survives a shift in aspect from functions to computational procedures and/or mechanisms.[5] The procedures and/or mechanisms which compute the basic cognitive functions of memory, learning, attention, etc. are modular because the sum of the mRd relations they have with daughter nodes is greater than the dRd relations that they have with each other. This constitutes within-level modu-

[5]It is not clear whether Fodor is interpreting the position to entail one, the other, or both. But it doesn't matter because the position allows for all three.

larity (M-p or M-m, depending on whether one is talking about procedures or mechanisms).

However, notice that this does not preclude the possibility of also individuating more complex functions such as perception and inductive inference and claiming some degree of modularity for them with respect to each other (within-level), and also with respect to the less complex procedures/mechanisms (micro-macro). Within-level modularity of these more complex procedures (M-m and M-p) would occur if the sum of the mRd relations they had with daughter nodes was greater than the dRd relations that they had with each other. Micro-macro modularity between the less complex cognitive procedures and/or mechanisms at level L_{n+1} and more complex procedures and/or mechanisms at level L_n, would occur if the sum of the dRd relations a node at L_{n+1} had with its siblings was greater than the mRd relations it had with nodes at L_n.

Vertical faculties, on the other hand, are individuated by reference to subject matter, or the domain of content of the input information (as opposed to the operations that are performed on this content). On a vertical individuation one would expect to find language and music faculties, because they are generally taken to constitute different domains. Whether one would also expect to find a memory or attention faculty depends on the brand of vertical faculties being evoked.

What Fodor is calling vertical faculties are again, in our terminology, a specification of systems under the aspect of functions, except this time the functions are more complex and individuated by the input or domain. The substantive claim of vertical faculties is that this individuation survives a shift in aspect to computational procedures and/or mechanisms.

Fodor distinguishes three different brands of vertical faculties: Gall's notion of *fundamental powers*, Chomsky's notion of *mental organs*, and his own notion of *cognitive modules*. All three individuate functions by domain, but beyond this they differ. Due to space considerations, we will only discuss Gall and Fodor here.

Gall, according to Fodor, seems committed to (i) individuation by domain, (ii) no sharing of common mechanisms between faculties, (iii) distinct mappings of faculties onto neural structures, and (iv) innateness of faculties. Criterion (i) merely makes Gall's *powers* a species of vertical faculties. Criterion (ii) denies both dRd and mRd relations between faculties and thus endorses M'-m and M-m modularity[6] (iii) is probably a commitment to spatial localization of mechanisms. Criterion (ii) requires that each faculty be self-contained with respect to mechanical resources and rules out the possibility of (horizontal) faculties, such as memory or attention, that are shared across domains. So one would postulate a "musical memory" in the musical faculty and a "linguistic memory" in the linguistic faculty, but not a faculty of memory per se. "Perception and memory are only attributes common to the fundamental psychological qualities, but not faculties in themselves; and consequently they can have no proper centers in the brain.... we have to discover the fundamental powers of the mind, for it is only these that can have separate organs in the brain." (Gall, quoted in Fodor 1983, p. 16).

Fodor's *cognitive modules* are similar to Gall's *powers*. The major difference is that while Gall's claim would be that the functional individuation by domain survives a shift in aspect from function to mechanism (i.e. does not require cross-classification), Fodor's claim is that it survives a shift in aspect from function to computational procedures. Thus Fodor's cognitive modules are (i) domain specific, (ii) informationally encapsulated, (iii) hard-wired into neuronal structure, (iv) computationally direct or primitive (i.e. there is a straight forward mapping between the virtual architecture and its realization in the neuronal mechanism), and (v) innately specified. Criterion (i) is interpreted as above for Gall. Criterion (ii) introduces the important idea of *informational encapsulation* which denies any significant dRd and mRd relations among the procedures which compute the domain individuated functions and indorses M'-p and M-p modularity between these procedures. Informational encapsulation is the procedural level counterpart to Gall's "no sharing of resources" (at the mechanical level).

On Fodor's account systems are modular not because they do not share mechanical resources, but because they do not share information. Informational encapsulation constrains the flow of information among cognitive procedures throughout the system. Cognitive modules compute their functions without requiring unlimited access to information computed by other modules.

Criteria (iii) & (iv) are a little more difficult to interpret. One way of understanding them is that they require a reasonably direct mapping between cognitive procedures and the mechanism in which they are realized. This would seem to imply that the physical causal constraints at the mechanical level will mirror the semantic causal constraints at the procedural level, which would result in M'-m and M-m modularity. (This, however, may not be Fodor's intended interpretation of the criteria because he considers himself noncommittal on the sharing of resources issue.)

So, Fodor's notion of cognitive module affirms M'-p and M-p, and perhaps also M'-m and M-m modularity. We now turn to his argument that central systems, such as those involved in belief fixation, will not be modular systems.

Fodor's argument for the nonmodularity of central, higher cognitive functions, such as belief fixation, is rather straight forward. It, however, does use some specialized vocabulary, and is illustrated with an example from the philosophy of science. In a nutshell, the

[6] Presumably, the M-p and M'-p categories do not apply to Gall.

argument is that belief fixation in science is (i) *isotropic* (i.e. everything that we know is relevant to determining what else we should believe) and (ii) *Quineian* (i.e. the degree to which we believe a proposition is a function of all our other beliefs). These are conceptual/logical claims about the structure of scientific reasoning. From these premises it is concluded that the central processes which mediate such reasoning cannot be encapsulated. The conclusion is a claim about human psychology. It does not follow as a point of logic from the premises[7], but given what we know of how the world works, it is certainly very plausible.

Response to Fodor

It is a good argument and has taxed researchers in the cognitive neuroscience community (Marshall, 1984; Shallice, 1984; Shallice, 1988). However, some researchers have failed to appreciate the limited scope of the argument. It is an argument only against *information encapsulation of domain individuated subsystems* engaged in non demonstrative inference. Of the various notions of modules reviewed, only Fodor's notion was domain individuated *and* dependent on information encapsulation.

Gall's fundamental powers are individuated by domain but the claim of modularity is made at the level of mechanisms, not procedures. They are untouched by the argument. The horizontal faculties, individuated by effect or output, can claim modularity of both computational procedures and mechanisms, without worrying about the argument.

So the upshot would seem to be something like the following: If you think that cognitive functions must be individuated by domain *and* that the only or most relevant notion of modularity for cognitive neuroscience needs to be based on information encapsulation of computational procedures, then Fodor's argument should give you reason to worry. If you are so worried, some of your options are to (i) deny the central role of non demonstrative reasoning in higher level cognitive functions, (ii) argue that one can place *a priori* limits on the propositions that may be relevant to a given non demonstrative inference, or (iii) argue that no conclusion about psychological mechanisms follows from the logical structure of non demonstrative argument forms. Good luck.

There is however, no reason to take this position. A more promising move would be to stop and question whether individuation by domain and modularity of computational procedures are necessary (or even sufficient) for cognitive neuroscience inferences. They seem to be neither necessary nor sufficient.

The logic of inference in cognitive neuroscience seems to require three things. First, it requires that one be able to individuate disjoint sets of functions. Whether this individuation be by domain or range, is in the first instance, irrelevant. Second, it requires a one to one mapping from functions to disjoint systems of computational procedures and/or[8] neurophysiological structures. Third, it requires that the cognitive system not spontaneously regenerate or reorganize itself after damage (i.e. that it be constant or nonplastic). These criteria would seem to be both necessary and perhaps sufficient to play the dissociation game, and neither individuation by domain, nor modularity of computational procedures is included among them. So, cognitive neuroscience has no methodological reasons to require both (or even one of) individuation by domain and modularity of computational processes. One may have other theoretical reasons for insisting on one or both, but nothing about the logic of the methodology requires one to do so.

References

Farah, M. (in press). Neuropsychological Inference with an Interactive Brain: A Critique of the Locality Assumption. The Behavioral and Brain Sciences.

Fodor, J. A. (1983). The Modularity of Mind: An Essay on Faculty Psychology. Cambridge, Mass.: The MIT Press.

Marshall, J. C. (1984). Multiple Perspectives on Modularity. Cognition, 17, 209-242.

McCarthy, R. A., & Warrington, E. K. (1990). Cognitive Neuropsychology: A Clinical Introduction. NY: Academic Press.

Shallice, T. (1984). More Functionally Isolable Subsystems but Fewer "Modules"? Cognition, 17, 243-252.

Shallice, T. (1988). From Neuropsychology to Mental Structure. Cambridge: Cambridge University Press.

Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, 106, 467-482.

Simon, H. A. (1973). The Organization of Complex Systems. In H. H. Pattee (Eds.), Hierarchy Theory N.Y.: G. Brazileer.

Vallar, G. (1991). Current Methodological Issues in Human Neuropsychology. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology, Vol. 5 Amsterdam: Elsevier.

[7] Fodor, of course, does not claim that it does.

[8] Whether this is a conjunction or disjunction depends on what one takes the scope of the discipline to be.