GÖTEBORGS UNIVERSITET

The Modularity Theses

The main versions, the problems, and what might be done

(Modularitetsteserna: De huvudsakliga versionerna, problemen och vad som skulle kunna göras)

Niklas Rydberg

Kandidatuppsats i kognitionsvetenskap

Rapport nr. 2013:009
ISSN: 1651-4769

Göteborgs universitet
Institutionen för tillämpad informationsteknologi
Göteborg, Sverige, februari 2013

TABLE OF CONTENTS

This paper is about mental architecture. Its main purpose is to examine claims that the internal organisation of the mind is modular either to a modest extent or to a massive extent, respectively. It also includes a summary of the pointed critique of modularity, according to which "mental modules", if such there be, are most likely few and far between. According the original version of the modularity thesis, put forward by Jerry A. Fodor in 1983, modules are responsible for the link between mind and world: they all have to do with various aspects of perception and sensation, but do not figure in higher cognition. Fodor's modules all exhibit "all or most" features found on a list of nine. This list is very controversial. Peter Carruthers, who arguably is the main proponent of a massive version of the thesis, wishes to include higher cognition in his theory of modular architecture and therefore modifies the list by interpreting some of the features differently, as well as by shortening the list as such. However, while massive versions are popular amongst evolutionary psychologists and AI researchers, controversy remains – not only concerning the (new) list, but also the range of the thesis. Much of the critique reflects the old divide between rationalist and empiricist philosophy, or more pointedly, the question of whether the capacities of the human mind are mostly innate or mostly learned. Jesse J. Prinz belongs to the latter camp and rejects both modularity theses, arguing that they are largely mistaken about the workings of the mind, and in some respects are completely unfounded. The paper ends with my own proposal for a middle-ground theory attempting to reconcile these positions, since I believe each has something important to bring to the table.

What Are the Modularity Theses?

This paper is about idea that the mind can be analysed in terms of functionally distinct, principally dissociable units of greater or lesser size. If it just meant that the hemispheres are subdivided into somewhat distinct regions and areas associated with certain types of processes, this thesis may seem to amount to little more than basic neuroanatomy in so far as the *brain* can be described as "modular", or to the uncontroversial thesis of functional decomposition (see Jesse J. Prinz, 2006). The thesis would indeed be quite uninteresting if that were the case. Such "brain modularity" hardly raises any eyebrows even outside the physicalist or naturalist camp – as, of course, it should not. However, the idea at stake here is about a similar organisation of the *mind*, and that certainly makes for a different story.

While any strong identification of operations of the mind with the workings of the brain remains controversial in itself,[1] one might easily construe a version of modularity thesis that does not involve such a claim. It is true that Jerry A. Fodor, whose seminal *The Modularity of Mind* introduced the theme of modularity in 1983, in effect makes that identity claim by suggesting that each module should be innately channelled and/or have its own specific neural architecture. But at the same time he maintains that (1) central systems are not in fact modular, and (2) that not every

1 For instance, consider the infamous, australian materialist identification of a phenomenal pain-state with "c-fiber firing" (type-type identity), and the admittedly more versatile functionalist type-token identities, whereby a pain-state is the function of effectively any sort of fibers (or *any* sort of substrate) reacting to a certain sort of stimuli in a certain sort of way. My point here is that both direct, type-type identification and indirect, type-token identification – as well as computationalist identification of mental states with what in the end amounts to binary processing – are or at least seem to be unable to account for either phenomenological observations about perception or for intentionality.

module must exhibit every feature on his list.[2] Taken together, his identity claim is weak and seems to be conditional upon whether the central systems are in fact essential or not. If not, a strong identity obtains only for peripheral input systems. Peter Carruthers, a fellow but rather different sort of modularist, is on the other hand much more lax about mind-brain identity (while being much more ambitious about the *extent* of modularity). For example, he does not find it reasonable to hold that modular systems generally have dedicated neural structures and therefore strikes this feature from his list. This move is not without merit as it makes the thesis more plausible as such. But how much of the list can be struck out without ending up with an empty, central concept? According to these authors, how many properties on their respective lists are essential for candidate systems? Suppose that only a subset of these features suffices for a system to be modular. If so, how big must this subset be to retain interest? Critics of the modularity thesis have often and quite rightly wondered just how many modules there are supposed to be. Too few would make the thesis rather meaningless. Too many would be equally disastrous; if a given function ends up with a set of modules, all of which can perform it independently, they better have some unique job to do as well. Questions like these are certainly important. Surprisingly, however, I have not come across any investigations into how many (at least logically) possible *types of module*s there might be. This is odd given its probably quite decisive nature: one would expect something like a bijection between types of function and types of modules, or at least that certain types of modules rank higher than others because of their greater versatility (say). If all modules were of the same type it is hardly their modularity that is of interest, but rather the set of instructions governing them. Have too many types, on the other hand, and one would probably benefit from assigning them unique names and to regard 'module' as a sort of umbrella term. In that case, saying that a system within the mind is modular becomes somewhat trivial – even if regarding 'module' as an umbrella term happens to be perfectly in keeping with the sort of untraditional faculty psychology that inspired the thesis in the first place. Accordingly, the question of types is taken up in section IV.2 below. As it turns out, different interpretations of 'most' yield very different and quite startling results. Since answering the question is not just a matter of simple arithmetic, but actually requires somewhat advanced mathematics, an appendix has been added to justify the candidate numbers I came up with.

In what follows, Fodor's and Carruthers' respective theses will represent the span of the modularism spectrum, from modest to massive. Together with Prinz' (2006) empiricist critique of both (and, in effect, of using the term 'module' about systems of the mind at all), they mark the three extreme positions of the debate. Anything falling in between will, or so I shall assume, be

2 See for example pages 37 and 98-99 in Fodor (1983). The rubric on p. 98 is "Input systems are associated with fixed neural architecture".

subject to the same sorts of arguments in favour of or against either position. Before going into this treatment of the debate, the main motivation for the modularity thesis seems to be this, regardless of which sense of 'module' is employed: If the mind is modular to some extent, explaining how such an odd phenomenon as mentality could have evolved turns out to be a quite straightforward matter after all. The mind becomes less "mysterious", as it were, since the general principle for its formation is part of the thesis about its organisation: the arrangement of domain- or function specific units into a whole either identical to or greater than the sum of its constitutive parts forms a perfect analogue to the evolution of the eye.

The basic principles needed to understand the spontaneous formation of a complex system, thought by creationists and others to be "inexplicable" other than in terms of a pre-formed and intelligently designed unit, is not very difficult at all once one stops thinking in terms of piecemeal processes and starts thinking in terms of division of labour: everybody does not have to do everything at once in order for everything to get done. It is not *a* linear process, but rather several. In his (2006), Peter Carruthers cites Herbert A. Simon's 1962 analogy of the two watchmakers: one of them assembles his watches component piece by component piece, whereas the other first assembles component sub-systems and then assembles these into complete watches. Since the latter strategy allows for a better overview and fewer steps per operation, it is both much more efficient, less prone to error, more beneficial the more complex the system is, and last but not least epistemically tractable even if the number of component parts should be very large. This presumably explains why modularity remains a popular theme amongst evolutionary psychologists and others despite its having sustained scathing empiricist critique.

The purpose of this paper is primarily to sum up and comment on the modularity debate using the above-mentioned three poles, and then seek out the middle of the resultant triangle. If this strategy is sound, and if the debate itself held genuine philosophical interest, a reasonable alternative to the thesis in question should be found more or less right in between these extremes. The next task thus becomes to describe and explain it. In order to do this in as clear a fashion as possible, the paper falls into four parts whose themes are the following: (1) The proposed modularity of low-level, peripheral systems *only*, (2) the proposed modularity of *all* or *most* systems of the mind, (3) the categorical rejection of both propositions, and (4), thoughts about what may be taken from the debate in the form of a positive theory of mental architecture. Part I is by far the longest, and part IV includes a modest note on method.

One final note before we begin: any interesting sense of 'modularity' should involve a number of non-trivial features, and the explanatory power (or at any rate, the reasonable promise) of

a modular theory of mind should exceed that of consistent, non-modular alternatives. Broadly speaking; if such turn out to be available, the modularity thesis may have played out its role.[3]

I. Fodor's Original Version

In his (1983), Fodor is far from claiming that *all* or even *most* of the mind's operations are modular, even if the title of his book may suggest as much.[4] As Prinz (2006) points out, the original version of the modularity thesis is restricted to low-level peripheral systems and sub-systems responsible for the mind's link to the world (and vice-versa), and does not have anything to do with higher-order processes such as thinking. Modules are always located on the very edge, as it were, making sensory systems and parts thereof primary candidates: in endnote 14 on page 132 (note to page 47), Fodor remarks that "[g]enerally speaking, the more peripheral a mechanism is in the process of perceptual analysis – the earlier it operates, for example – the better candidate for modularity it is likely to be." Primary candidates for modular subsystems of visual perception, then, would plausibly include but perhaps not be limited to "mechanisms for color, for the analysis of shape, and for the analysis of three-dimensional spatial relations" and more fine-grained varieties of these; according to a 1980 study, such features are in fact registered slightly *before* any perception of the object exhibiting them takes place. In a while, we shall return to that strange observation.

It is important to note that Fodor's modularity thesis does not imply that these mechanisms operate according to the same, fundamental principles. Such a common basis is instead typical of the way traditional, "horizontal" faculty psychology would have it: according to Fodor, however, there is no such thing as a general-purpose faculty of perception. His theory stems from a very close reading of Franz Joseph Gall, the notorious father of phrenology. Fodor revises and reintroduces the latter's theory of *vertical* faculties (but not the pseudo-science, mind), thus departing from mainstream psychology and philosophy of mind. The move is highly interesting. While Gall does not himself use the term 'vertical' to describe his doctrine, Fodor introduces it to mark the great difference between the two versions of faculty psychology (see page 14): where the traditional, horizontal version has *one* general faculty of memory, *one* general faculty of perception, and *one* of vision as a subset of the latter, and so on, the Gallian, vertical counterpart has dedicated *faculties* of memory for different types of information, *faculties* of reasoning, and *faculties* of perception for different types of perceptual input. Nested within the latter we find *faculties* of vision for different types of visual stimuli, such as those mentioned above. The next step is to label this great number of faculties "modules". While this step may be motivated by a desire to further differentiate the

3 Throughout this paper, all quotes are given with their original spelling. Full references to all works cited are found in the bibliography.

4 Unless otherwise noted, all references in this section are to Fodor (1983).

position from its traditional counterpart, it may also simply stem from the fact that 'module' as a concept, at least in the usual sense, has a much less mysterious air about it: Where 'faculty' reeks of nineteenth century idealism, 'module' is much better suited to a computationalist view of the mind grounded in physical structures.

Back to the various aspects of vision and the delay between the mind's registering of features such as colour, shape, three-dimensional spatial relations, and its registering the object instantiating them. This does not imply that they "build up" the object one by one: the features are all registered at the same time, and we have a sort of division of labour here as well. To quote Fodor's illuminating quote from Treisman and Gelade (1980, page 98): "[...] features are registered early, automatically, and in parallel across the visual field, while objects are identified separately only at a later stage, which requires focused attention". It is found in endnote 14 (note to page 47) on page 132 of *Modularity*. In the same endnote, Fodor goes on to say that "[t]here is analogous evidence for the modularity of phonetic feature detectors that operates in speech perception [...], though its interpretation is less than univocal [...]".[5] Extending this analysis to the entire perceptual apparatus is warranted within this theoretical framework not because it should operate according to the same basic principles (which is a horizontal standpoint), but because 'perceptual apparatus' as a concept extends to a number of systems that are all peripheral.

Thinking and fixation of belief are, as we touched on above, *not* modular on Fodor's theory: as we shall see, this sort of operations belong instead to the larger class of non-modular central systems. According to Fodor, these systems do not lend themselves to straightforward analysis precisely because of their non-modularity. He therefore has little hope that cognitive science (or psychology, or studies of mind in general) will ever be able to gauge the principles of their operation completely. (See for example page 38.) Genuinely modular systems, on the other hand, are comparatively easy to understand seeing as their overall nature is, by definition, functionally distinct. To summarise the above, one might well say that Fodor's thesis is *weak* in so far as it only concerns peripheral systems, but that the concept at its centre is *strong* as its description is comparatively detailed.

I.1 The nine features of a Fodor-module

Since Fodor's use of the term 'module' is rather idiosyncratic, the sort of systems that satisfy his comparatively extensive list of characteristic features have become known in the debate as 'Fodor-modules'. A rather clumsy term, it will here be shortened to 'F-module' when simply putting 'module' might be confusing. Note, however, that Fodor does not require that each and every

5 Full references to the studies Fodor mentions here are included in the bibliography.

candidate system exhibit every one of the following nine features: it suffices that they satisfy "most or all of them" (page 47). What 'most' might mean will be made clear in a while. Until then, let us have a look at the list as such. Robbins (2009) points out that their original order or presentation is somewhat misleading, but even so there is a point in giving it anyway. Jesse Prinz' concise descriptions (found in his 2006) give the rubrics, which are then followed by lengthier discussions based on Fodor's account of them:[6]

1. Domain specificity: "modules cope with a restricted class of inputs"

According to Fodor (pages 47 to 52), a domain is something having a distinct sort of content. A module is domain specific in Fodor's sense if, and only if, it processes only the content of precisely that domain and *nothing else*. Moreover, the feature of domain specificity in a module requires that the stimulus domain whose content is up for modular processing is *eccentric*, meaning "one in which perceptual analysis requires a body of information whose character and content is specific to that domain" (page 49). A *ceteris paribus*-clause applies here: "All things being equal", Fodor goes on, "the plausibility of this speculation [i.e., about special-purpose computational systems in pre-modular cognitive science] is about proportional to the eccentricity of the domain" (ibidem).

If an F-module is to be domain specific, it must be the case that it always processes information from precisely *this* domain and never *that*. In other words, this feature not only applies to the module *m*, it also applies to the information processed by *m*. Call the latter *m*-specific. Therefore, only content that belongs exclusively to domain *D* (e.g., visual stimuli) may be handled by module *m* (e.g., vision), and conversely: *m* will only handle content belonging to *D*. Note, however, that it is not enough to define a domain in order to make a sound inference of a corresponding module; while *m* implies *D*, just in case *m* has the feature now discussed, the possibility of consistently defining a domain *D'* does not imply the existence of a module *m'* specific to it, nor that the content of *D'* should be *m'*-specific.

2. Mandatoriness: "modules operate in an automatic way"

Fodor's discussion of this feature covers pages 52 to 55. He opens with an important remark: "You can't help hearing an utterance of a sentence (in a language you know) *as* an utterance of a sentence, and you can't help seeing a visual array *as* consisting of objects distributed in three-dimensional space". (My emphases). The same principle supposedly holds for other modules and domains, be they specific or not. However, Fodor also notes that this feature might be manipulated and, as it were, overridden: for example, concentrating on something else than what is being uttered

6 All page numbers in this section refer to Fodor (1983) unless otherwise specified.

– or covering one's ears so as not to hear it – shifts focus from the utterance to that other something. Moreover, some painters and phoneticians seem to be able to gain access to "raw transducer output", as it were, which is to say their extensive training should in some cases allows access to the input a peripheral module receives from its transducer (page 54). While that feat might seem to contradict the notion that we only ever access or make use of *processed* information, Fodor is of the opinion that these "highly sophisticated phenomenological reductions" are in fact "*super*sophisticated perceptual achievements" possible only for subjects with extensive training (ibidem). It is so rare that it may well be treated as a harmless exception from the rule.

Speaking of phenomenology, however, it should be noted that what Edmund Husserl meant by "phenomenological reduction" does not apply here at all. The "things themselves" are not identical to putative acts of "seeing the visual field" or "hearing the speech stream", as Fodor mistakenly seems to think. The things themselves, or the phenomena in Husserl's sense, are rather *Gestalten*, which here means "figures [*Figuren*] within the visual field" or "the speech stream *qua* language-figure". These figures stand out as more or less well-organised wholes against a (back)ground, and they are not identical to the sum of their constituent parts. A phenomenological reduction, then, is not the reduction of an object of perception into its individual features. It is instead the reduction of possible descriptions of a given phenomenon (i,e., prejudices about its nature) to an optimal description thereof – the phenomenon as a sort of pure figure – by way of actively suspending or "bracketing" one's judgement. Doing so is indeed super-sophisticated and the feat as such is irrelevant here. While we might shift our attention from the things themselves in that sense, or misinterpret them (as we usually do) or wish not to acknowledge them, a mere act of will cannot disable our *perception* of them (which, again, does not usually coincide with how they are in themselves). Fodor draws this conclusion too, but does so based on a somewhat grave misunderstanding the correction of which may have come across as unnecessary. But it is worth pointing out because phenomenological reductions – if such are indeed possible in the way Husserl thought they were – is something that might rather be undertaken by non-modular, central systems themselves than by the "intrusion" of higher cognition into modular systems. Thus, we might disregard the possibility of such intrusion altogether, at least in so far as it implies *conscious* access.

In sum: whether or not one's attention to a particular stimulus domain – or rather to stimuli belonging to the proper domain of the relevant module – is being manipulated or impaired, mandatory operation means that *if* the module is online and offered an application possible for it, it has to take it; it has no choice in the matter. As Fodor puts it on page 55, "perceptual processes apparently apply willy-nilly in disregard of one's immediate concerns". Perforce, we normally have no choice in these matters either.

3. Inaccessibility: “higher levels [...] have limited access to the representations within a module” Both low-level, peripheral systems and central systems have few or no means of using information while it is being processed within a module, Fodor argues on pages 55 to 60; instead, they must rely on it to give them the processed information as input. Also, as noted above, features are registered *before* the object. The idea here is that modular systems of perception make stage-wise, bottom-up approximations of the perceived object (which does not mean that they somehow build up *the object*, but rather its representation):

> The present point is that the subject doesn't have equal access to all of these ascending levels of representation – not at least if we take the criterion of accessibility to be the availability for explicit report of the information that these representations encode. Indeed [...], the lowest levels (the ones that correspond most closely to transducer outputs [e.g., nervous transmissions]) appear to be completely *in*accessible for all intents and purposes. The rule seems to be that, even if perceptual processing goes from 'bottom to top' (each level of representation of a stimulus computed being more abstractly related to transducer outputs than the one that immediately preceded), still *access* goes from top to bottom (the further you get from transducer outputs, the more accessible the representations recovered are to the central cognitive systems that presumably mediate conscious report).

Earlier in the book, Fodor utilises the structure of a syllogism as a metaphor to describe this characteristic behaviour of input systems: its associated transducers provide sensory information as premises, and the role of the module is to infer the (more or less) correct conclusion. Central systems, then, under what might – with some degree of accuracy – be called “normal circumstances” have access only to these conclusions. A couple of interesting examples that Fodor cites on page 57 are (a) the fact that when one checks the time on an analog watch, only the time indicated but not the shape of the numerals is remembered, and analogously that (b) the gist of a sentence is usually recalled with some degree of accuracy whereas details of syntax – and sometimes even the language used, provided one knows it[7] – are quickly forgotten. This is presumably due to these details being irrelevant for a specific task and therefore discarded in process of registration of the stimulus. However, not all circumstances are normal in that sense: Sometimes stimuli are presented in such a way that instead of simply being able to recall not just the general character or the salient features of an object, but structural details which are otherwise forgotten. The explanation seems to lie in the *mode* of presentation:

> Witness the fact that in tasks which minimize memory demands by requiring comparison of *simultaneously* presented stimuli, responses that are sensitive to

7 This is my addition. Fodor does not mention it, but it is nonetheless both true and consistent with his example.

> stimulus properties specified at relatively low levels of representation are frequently *faster* than responses of the sort that high-level representations mark. Here, then, the ordering of relative accessibility reverses the top-to-bottom proposed above.

This was brought out by a study in which subjects were asked to state whether pairs of letters were identical either with respect to their *font* – or here: *case* – (t;t; T;T) as well as their *place in the alphabet*, or alphabetically identical but not case identical (t;T; T;t). When the stimuli were presented *simultaneously*, the response times were faster for pairs that were identical in both case and alphabetical position, than for pairs that were only alphabetically identical. However, when the stimuli were presented *sequentially* (in broken pairs with a slight interstimulus delay, that is) response times for alphabetical identity were faster both in comparison to case identity, as well as faster/slower in comparison with the simultaneous presentation of the stimuli.[8] Here Fodor takes for granted that shape (here: case) is always registered earlier in the process than (alphabetical) position, and accordingly draws the conclusion that it is "a matter of the way in which the subsystems of the input processors interface with memory systems" (page 59). Though while it certainly is the case that with respect to some of their properties, simultaneously presented stimuli are recalled more efficiently than sequentially presented counterparts, this fact does not necessarily have to do with system interface even within a computationalist framework. Interface relations make for a comparatively complicated explanation while for familiar reasons, theoretical simplicity is not only more elegant and therefore more preferable, it is also usually much more successful. I submit, therefore, that this is rather a matter of *task relevance* and *property weighting*: if one is asked to look for certain properties, they will be sorted out from the noise because they are weighted accordingly. Besides, it is most likely simpler for evolutionary reasons to become aware of qualitative identities (here: case) than of quantitative identities (here: position in the alphabet): it is a familiar fact that our brains have not undergone any dramatic changes for the last 10,000 or 20,000 years or so. So, if for prehistoric humans it mattered more whether an in some sense identical pair of objects consisted of dangerous sabre-toothed cats or benevolent clan members – that is whether the members of the pair were, in a quite pragmatic sense, similar to the observers –, than whether the members of another pair were in fact similar *to each other* in a comparatively useless sense, there is no reason why the same principle should not hold for us today. This plausibly explains why identities of case and position were registered faster when presented at the same time than one after the other. It also provides a way for testing the modularity thesis as such: does a particular interpretation of "the evidence" hold up against evolutionary scrutiny? Given the interest in it amongst evolutionary psychologists, that test should certainly be made on a regular basis. But

8 Posner (1978). The details are given by Fodor on the page quoted.

yet again, even if Fodor's reasons for drawing a particular conclusion are flawed, the conclusion does seem to hold as such (compare footnote 10). He is, however, right on the mark when he says that it is "less a matter of information being unconscious than of its being unrecalled" (page 59).

4. Swiftness: "modules generate outputs quickly"
In comparison to central systems (notably conscious thinking), intra-modular processes are supposedly very fast, according to Fodor's description on pages 61 to 64. If the auditory system is modular, then its having this feature means that one cannot help but *hear* unless one is either asleep, unconscious, or permanently or temporarily deaf. Such situations are examples of when the auditory system is offline for some reason or other, meaning that swiftness of processing (as well as mandatory operation above) only applies when the module in question is online.

5. Informational encapsulation: "modules cannot be guided by information at higher levels [...]"
(Pages 64 to 86.) Modules are not normally able to use information stored or used centrally (for instance by conscious thought) in its operations, and in case they are they cannot use very much of it nor use it to a great extent.

6. Shallowness: "modules have relatively simple outputs (e.g., not judgements)"
(Pages 86 to 97.) Since modules are input systems, they do not analyse information as much as they prepare it for central analysis. Thus, their output will be comparatively superficial.

7. Localisability: "modules are realised in dedicated neural architecture"
(Pages 98 and 99.) Good examples of this are the tympanic membrane and the cochlea in the auditory system, and the rods, cones and retina in the visual system: these structures are dedicated to the handling of certain types of stimuli, auditory and visual, respectively, and have no other application. Likewise, modules with this feature do not have parts that are also parts of other systems, modular or not. The key word here is "dedicated", which Fodor uses in a strong sense: *x* is dedicated to *y* if, and only if, *x* is not functionally associated with anything but *y*. This is not the only possible interpretation, of course, but it is the one Fodor seems to take for granted. A module *m*, then, can *only* be realised in structure *S*, and *S* is such that it can only realise *m*. The structure, that is, could not be anything but the substrate of a modular system, and moreover of a particular one. So far this is basically a rehash of domain specificity. Here is the interesting part, however: the neural location of *S* should be all but invariant between subjects able to perform the sort of cognitive functions associated with the having of a specific sort of module: a neurologist might then

confidently say that *if* a subject is unable to perform a given function, her brain will have a lesion in precisely this or that area. Note, therefore, how this feature may unequivocally seem to imply both domain specificity and the sort of characteristic breakdowns that are related above. One might also argue that it clearly indicates ontogenetic determination. Fodor's system is indeed very intricate and carefully thought out, but as we shall see it does not quite work the way he wanted.

8. Characteristic breakdown: "modules can be selectively impaired"
(Pages 99 and 100.) This is to say that if a module were to be injured or destroyed by, say, a lesion in the brain, the only consequence should be that that particular function is either impaired or altogether wiped out. The feature, as we shall see, is very problematic. Its inclusion, however, is nonetheless reminiscent of the everyday sense of 'module': if a system is thoroughly modular, its functionality as such does not depend on the functionality of any one of its components.

9. Ontogenetic determination: "modules develop in a characteristic pace and sequence"
(Pages 100 and 101.) Basically, we may predict the development of any given modular system in an infant subject by previous empirical study of other infants. If this feature applies, the structural description of a module is roughly identical between subjects, but as far as this feature goes the exact, neuronal realisation of that particular module may vary.

I.2 How many types of modules are there?
Below we see the list of features again, this time with Philip Robbins' (2009) shorthand and with bracketed numbers describing their relative importance (or weight). An assigned weight of [9], then, alludes to the high probability that the feature in question is instantiated by a possible candidate system for modularity:

1. Domain specificity [2]
2. Mandatory operation [7]
3. Limited central accessibility [8]
4. Fast processing [6]
5. Informational encapsulation [9]
6. 'Shallow' outputs [5]
7. Fixed neural architecture [3]
8. Characteristic and specific breakdown patterns [4]
9. Characteristic ontogenetic pace and sequencing [1]

This new list enables a somewhat nit-picking argument designed to illustrate the problem of not having a single and unequivocal definition of 'module' in a context where the extension of a central term is of great consequence. Seeing as some features are held to be more important than others, their respective weights are given within the parentheses; '[9]' means 'most important' and '[1]' means 'least important'.[9] Providing these numbers allows us to interpret the somewhat fuzzy phrase "[*x* has] most or all [of the features listed]" as follows: The weights add up to [45], meaning that if 'most' means "at least five of the nine available features", the lightest combined weight of an individual F-module ranges between [11] and [35]. If one were worried about vagueness from the start, this result is not exactly reassuring. It therefore seems that if excessively "light" modules are to be avoided, we therefore need some sort of constraints governing which features may be present in conjunction with which.

That last remark still holds if 'most' actually means "the total weight of the main features exhibited by *x* must amount to at least half the total weight of all listed features", which in turn means that every genuine F-module should have a combined weight of at least [23]. This at least fixes the minimum weight, which might seem to resolve the issue of vagueness. However, at the same time it allows possible candidates to exhibit as few as three of the features associated with F-modularity; moreover, they must be counted as equally good or promising as those exhibiting at least seven. In the first case the top three most important features are present, whereas only one of them – the least important of them, to boot – needs to be present in the latter. This alone makes the list somewhat arbitrary. For why should the most important feature of all, informational encapsulation, not be required for *every* good candidate? And if it is, why does Fodor even bother with the least important features? It certainly would appear to be something amiss about his whole project. At this point it should be noted that the threshold values suggested here are my own invention; that they are completely arbitrary, and that they only serve to illustrate the problem of having many possible types of moduless. Also note that some features do seem to imply each other: the issue here is that Fodor neither specifies something akin to a threshold value – he effectively avoids doing so by using the vague phrase 'most or all' –, nor argues that certain features should always cluster together with certain others.

At this stage, the objection might be raised that the above line of reasoning has been an exercise in futility because neither the respective weights of the features, nor the principles of their combination, have any bearing on the description of a module. Provided that the point of the original version of the modularity thesis is not so much to explain the mind once and for all as it is to give us a fighting chance for *having* a fighting chance of explaining at least some of its

9 Robbins's entry on modularity in the Stanford Encyclopedia of Philosophy is very clear on this hierarchy.

operations, it is perfectly fine if the sense of 'module' is a bit vague. After all, those sympathetic to the original version (apparently precious few, these days) might claim that Fodor's project is preparatory in character. However, the simple answer is that *if* the undeniably influential *Modularity* is indeed meant to be a preparatory work, it is also deeply flawed: unless the meaning of 'module' (with or without the F-prefix) can be fixed even by the philosopher introducing it as a technical term, the thesis as such remains very slippery indeed. As an umbrella term it turns out not to do much good at all: it does not work as intended, and Fodor's "preparation of the field" turns out not to help anyone very much – at least not in any direct way. This is unfortunate since, despite appearances (made abundantly clear by Prinz (2006)), he might be onto something important.

The more complicated answer is the following: Any interpretation of 'most' will be arbitrary, and the number of possible types of F-modules will therefore vary greatly. Thus, if 'most or all' means "at least five out of nine features", there might be as many as 258 different types. Quite a few, in other words. So make it "seven out of nine" for a much more attractive maximum of 46 types, and it seems we are getting somewhere. To really narrow it down without assuming that 'most' in this context actually means "all", arguing that at least eight of the features must be exhibited by an eligible candidate yields a maximum of merely nine types. But, as was hinted above, some of these types will have mostly relatively unimportant or light-weight features. Why should those types be included at all, that is, why should systems that only exhibit comparatively inconsequential features count as modular? So, in an attempt to sort out this problem, I suggested that one might establish a minimum total importance – or weight – to keep the list as a whole while at the same time ruling out such types that, though logically possible, fall below the threshold. (This may or may not entail that such systems are not metaphysically possible). If, then, 'most' means "a combined value of [23] or more" (which was just described as seemingly non-vague, due to the threshold), there might be about 116 different types of F-module.[10] This is without doubt excessive, so we might try combining these constraints: Stipulating that number of features exhibited must be, say, at least seven out of nine, and their total weight must be, say, [35] or more – thus excluding certain combinations – might seem to be the answer to this conundrum. However, for two reasons it is not a very good one: Firstly, this strategy is likely to yield about 15 types.[11] 'Module' is thus an umbrella term rather than concept with a well-defined intension. It is true that Fodor thinks input systems, which for him means "modular systems", form a natural kind. It is also perfectly all right if natural kinds contain several sub-kinds. However, it does have some interesting consequences for his thesis which shall be looked into in a minute. Secondly, it goes without saying that these

10 The calculation can be found in the appendix.

11 I say 'likely' because establishing a reasonable threshold is a complicated matter, whether philosophically or mathematically. See appendix.

constraints are completely arbitrary; in the end it matters very little which of these numbers we choose, since all of them tend to make the modularity thesis as rather unhelpful approach to mental architecture. Here is why:

Whether we have (at most) 258, 116, or 15 different types of F-module, proponents of the original version of the modularity thesis must able to say which of these types are (1) consistent – that is, which features form logical clusters and thus might imply each other, meaning that it may not be allowed to combine features from two clusters without also combining the clusters –, (2) which are genuine candidates (or metaphysically possible candidates for mental or at least mind-like operations), and (3) which actually warrant the name of 'module' rather than 'token operation' or some such term. Especially, they must provide a reasonable (or at least non-arbitrary) threshold clause. But for the sake of argument, let us assume that the first two tasks can in fact be solved successfully, for instance by answering "all of them" in both cases, and that a successful solution to the third rules out the use of any other term than 'module'. If so, we would have an inkling about how large the sub-class of (purportedly) modular systems are as well as a paradigm for theoretical studies of mind. But this alone makes the thesis come across as more than a little desperate.

Above, the importance of finding out whether a proposed candidate for an F-module is logically consistent was mentioned. Admittedly, the use of consistency here is a bit awkward: what it means is that if two or more sets of features form logical clusters, it would be inconsistent to break them apart and form a new set of features from the spoils – for one thing, it would seem to fly right in the face of evolution (especially Simon's "architecture of complexity"). For another, if some features do not form a natural cluster, why should they (and no others) form a cluster at all?[12] Both Prinz and Robbins analyse the list by way of logical clusters of its members, and it is a good strategy that both authors employ with great skill.[13] I therefore refer the reader to their respective texts. For now, suffice it to note that these clusters are pairs or trios (such as "dissociability and localizability" or "mandatoriness, speed and superficiality" (see Robbins, 2009)). However, while Fodor frequently notes that one feature, e.g., informational encapsulation, implies another or a couple of others, such as speed and superficiality, there is nothing to say (a) how strong these connections are, or (b) which *clusters* imply each other. Since a lack of strong connections both

12 For instance, would an F-module *m1* that has characteristic ontogenetic pace and sequencing, fixed neural architecture, is fast in its processing and yields shallow outputs, and finally follows characteristic and specific breakdown patterns, while lacking every other feature on the list, be either consistent, possible, or genuine? To me, the above description sounds like a sputtering engine running idly and without purpose. Of course, this alone does not rule out the metaphysical possibility of there being such an F-module as *m1*, but I cannot think of any mental operation that would fit the description. Finally, unless *m1* had a well-defined task at one point (or just evolved by accident and was never selected away because it does not impede reproduction), does it qualify as a genuine instantiation of for the concept of 'module'? Well, perhaps in the trivial sense of the term. In this context, however, it does not seem to do at all.

13 Carruthers (2006) might be said to do so too, but to a much less obvious extent.

within and between clusters results in the more or less numerous types of F-module discussed above, a strong adherence to this version of the modularity thesis may well turn out to twist facts about the organisation of the mind – and to do so whether or not there are systems that in fact fit the description of a certain type of module, i.e., that exhibits "most or all" of the features in Fodor's list and have no other interesting characteristics. At most we might say that the putative subclass consisting of such systems is little more than a theoretical construct – and a vague one at that – designed to fit well within a computationalist framework.

II. Carruther's Massive Version

In "The case for massively modular models of mind", Peter Carruthers presents an outline for what he thinks amounts to a reasonable thesis of massive modularity. His account is positive in the sense that it does not consider counter-arguments and possible ways for the massive modularist to defuse them; for this, we are referred to other texts by his pen. "The case ..." is the first chapter of the 2006 anthology *Contemporary Debates in Cognitive Science* (edited by Robert J. Stainton) and is probably best described as an exposé of a line of reasoning.

"The case ..." begins with an attempt to define the problematic and rather vague concept of 'module' by using everyday technology as an analogue: a component part *C* of a system *S* is modular if, and only if, the general functionality of *S* can be retained in the absence of *C*. Carruthers exemplifies this by a hi-fi system: the system as a whole retains its functionality even if one of its parts, a speaker or the optional tape deck, say, breaks down. The component can then be replaced, repaired, or simply removed without affecting the functionality of the system as such. Arguably, this is the weakest sense of 'module' and its cognate terms: a functionally distinct, easily identifiable component part of a system that, in turn, is modular if it consist entirely or to some interesting extent of such components. However, this sense of 'module' is arguably too strong (and, ironically, at the same time too weak) for a reasonable theory of mental architecture and operational organisation. We shall come back to this later on.

As Carruthers' purpose in "The case ..." is to outline a thesis of massive modularity, some of the features in Fodor's list will have to be struck out. It is, after all, meant to apply only to non-central systems (or as Prinz puts it in his introduction (see below); low-level and peripheral systems). Suffice it to say that if a reasonable theory of mental architecture can be construed, in which the mind is held to consist entirely or mainly of functionally distinct components in something like the sense given above, 'module' cannot mean 'Fodor-module'.

So, then, if the mind is massively modular (or if it consists of "a *great many* modular

components"[14]), what sense of 'module' and its cognates could possibly apply? This question lies at the core of "The case ...", and its answer is found in the list of features Carruthers ends up with. In order to get there, he employs three main arguments that shall be briefly examined in what follows. These are (1) The argument from biology, (2) the argument from task specificity, and (3) the argument from computational tractability. Already in our introduction we familiarised ourselves with the first and most important of these, namely Simon's 1962 analogy of the blind watch-makers. You will recall that its explanatory power is great, seeing as it shows in a clear and simple way how the construction of complex structures can be easily accounted for by evolutionary processes.

The evolutionary process of assembling vastly complex structures is actually deviously simple: divide and conquer. If a watch-maker first assembles 'atomic' parts into functionally unified, 'molecular' components – which ideally consist of few enough 'atoms' to easily keep them and their order of assembly in mind – and arrange them side-by-side, she may fail to assemble a particular 'molecule' without having to start over from scratch. Starting instead at the point of the mistake, she can simply re-assemble the component and continue as before on the others. When the molecular components are finished, they can easily be assembled into 'super-molecules', and these in turn to the complete watch. At no point is a plan of the complete watch necessary. Now, blurring the line between metaphor and actual, evolutionary processes, the fact that all of the components should successfully form a unified whole can be explained in terms of convergence upon a function favouring the survival of the genome (here: the design of the watch). A rough description might be that molecule A increases the efficiency operation of molecule B, and that these two together (super-molecule AB) increases the efficiency of operation of super-molecule CD. The complete system ABCD formed in part due to the proximity of its components, and in part due to their compatibility and/or adaptability to each other.[15]

While Carruthers is certainly right about the great mileage that can be gotten from Simon's argument, he is quite wrong about its proper application. It concerns architecture, or better yet: the construction of complex *structures*, and only indirectly has to do with the function or use of a *system*. Specifically, while biological structures such as the eye and the brain doubtlessly evolved in the above manner, it does not follow that the operations of the brain – i.e., that which gives rise to

14 Carruthers' alternative, slightly weaker (and yet more problematic) thesis. See his introduction to "The case ..." in Stainton, 2006. His emphasis.

15 This might sound as if evolution is driven by flukes and happy coincidences. But given natural selection, such systems that are neither functionally compatible nor succeed to form unified super-systems will simply be tossed of the equation. This process is not coincidental at all, nor are the "flukes" completely random – they are instead probabilistic results of a non-linear process. That some systems are compatible and will converge upon what turns out to be an essential function for a given organism is bound to happen. The process, after all, takes generations. Despite its blindness, the pragmatic nature of the survival of the fittest it is actually quite efficient over time.

the mind[16] – can all be directly linked either to specific systems or neuro-anatomical regions, which is what Carruthers implicitly tends to do. For instance, many systems co-operate to enable 'synergic' functions, and so form functional units that transcend anatomical regions. (See Prinz, 2006 for some examples.) Carruthers is well aware of this, of course. He even bases part of his other arguments on such functional or operational nesting. The issue here, however, is that one must distinguish between the *architecture* of a structure, and its actual *use(s)* or *function(s)*. (Note that 'use' and 'function' may differ: the function of a bottle-opener is to open bottles, but in case of emergency it might be used as an improvised hammer.) Sometimes the one cannot easily be inferred from the other; in fact, in some cases such inferences are all but impossible or even destined to fail. Much more can be said about this, but the above suffices for present purposes.

Carruthers ends his discussion of the first argument by claiming that "[r]oughly speaking, [...] we should expect there to be one distinct sub-system for each reliably recurring function that [...] minds are called upon to perform". This, while being straightforward enough, is taking it all too far. Not only does he conflate architecture with use, and moreover seem to uphold domain specificity after all, he insists that these reliably recurring functions are *myriad.* Perforce, he implies that there should be myriad sub-systems in order to account for them. To illustrate his point, he lists a number of aspects – or functions – within the social domain; typical behaviours of our species (and, indeed, of many more) such as identification of "degrees of relatedness to kin", incest avoidance, relationship building, and so on and so forth. Again, these are all *behaviours*, many of which are perfectly conscious and volitional. When Carruthers remarks that taken together, the above social functions – if the term is allowed – "is just the tip of a huge iceberg, even in this one domain" he, alongside many if not most evolutionary psychologist, seems to conflate reliably recurring *functions or operations of the mind* with reliably recurring *patterns of behaviour exhibited by social animals*. Instances of these domains – and here I use the word in its usual sense – do not correspond exactly to one another. The strong conclusion that each of these behaviours has a dedicated and modular sub-system responsible for bringing it about is completely unwarranted. So is the weaker conclusion that each definable domain, or sub-domain if you will, has brought about one or several dedicated systems for handling their brand of information. All we can say with some confidence is that these sorts of behaviours seem to be typical for a number of species, and that their mental architectures may have features that help explain them.

Interestingly, Carruthers line of thought indicates a rather sharp break with Fodor. "It would

16 I point this out because while the overall activity of the brain does not necessarily exhaust the character of the mind, Carruthers sometimes seems to hold that cerebral and mental systems are conceptually interchangeable to a great extent. But even if he in fact does not think so, there are good reasons to uphold the distinction. A case in point would be the controversy that strong identification of the two brings.

be a bad idea (not to say an incoherent one [...])", the latter says on page 26 of his 1983, "to postulate a faculty corresponding to each prima facie distinguishable behavior and let it go at that". Instead of relatively few modules that are all (or mostly) innate or innately channeled, we end up with a plethora of assembled or activated *token* modules. Here, 'token' pertains to the implementation of individual modules, whereas types may comprise several individuals. (Compare section 3.2 of "The case ...").

Against this background, Carruthers' employment of the second argument (from task specificity) may seem to reek of tragic irony. That is somewhat unfortunate since the argument has a certain beauty to it. The gist of it is this: a task *x* is such that it can only be performed *as x* by a process or operation *y*, and *y* is such that it can perform at least *x* (weak version) or *x* only (strong version); therefore all processes or operations are task specific.

In short, different tasks are thought to "require quite different learning mechanisms to succeed".[17] As with the previous argument, Carruthers is all to eager to draw far-reaching conclusions when such are poorly motivated by "the evidence". Since it has little to be said for it, other than that it is elegant at first glance, we shall not dwell on it very long. (See the section on Prinz' critique.) A couple of quotes along with brief comments will do.

When Carruthers, whose desire to map distinct tasks to dedicated modules is now abundantly clear, writes that "[i]t is very hard to believe that there could be any sort of *general* learning mechanism that could perform all of these different roles", a rather pointed question comes to the fore: has Occam's razor gone blunt from disuse? Positing distinct mechanisms of the mind on the sole basis of the distinctness of the tasks performed by the organism as a whole surely does not comply either with the principle of parsimony, nor with the prudent sort of methodology outlined below. So, while the 1:1-mapping of tasks and dedicated (sub-)systems is reassuringly straightforward, it is also quite uneconomical and in this context cannot help but appear naïve. Imagining something along the lines of a general learning mechanism capable of "performing all of the different roles" indicated above turns out not to be very hard at all. Since Carruthers' painstaking and careful work on modularity over the years is admirable in its own right, it was a bit awkward to find that on the contrary, imagining such a system is almost embarrassingly easy. The only thing one has to do is to distinguish between a system and its applications: If a system *S* has more than one task to satisfy, it does not have to fall into a corresponding number of distinct sub-

17 The attentive reader should at this point recall the vertical faculty psychology underlying Fodor's original thesis, i.e., the organisation of the mind put forward by Gall according to which each propensity or aptness of the subject (such as playing music or doing philosophy) is associated with, as it were, domain specific mental faculties of memory, perception, reasoning, etc. According to that picture, the subject's capacity for remembering pieces of music is distinct from her capacity for remembering pieces of philosophy: they do not refer to a general-purpose memory faculty handling both kinds of input in the way the traditional, horizontal counterpart to Gall's theory of mental organisation would have it.

systems $s_1 \ldots s_n$, as Carruthers would have it. Instead, if *S* were a multi-purpose system *S(mp)*, such that its number of applications $a_1 \ldots a_n$ corresponds to the number of tasks it is called upon to perform, it would "perform all of the different roles" while still being essentially general-purpose. (Of course, one might well take 'application' to mean something along the lines of "sub-system", but doing so would also be to use the word in a somewhat unusual sense.) If you thought of a mind or a brain while reading the above description, that was precisely the point.

Carruthers third argument, that from computational tractability, is rather interesting and has a lot to be said for it. Given the vast and dizzying amount of information in the world, how can the mind be so efficient in its operations? Even if, as in this case, the efficiency of operation is"quick and dirty" rather than "optimised", the mind's ability to take short-cuts through droves of input is nothing short of amazing. Are we then to think that the principles governing this ability have some sort of similarity to those governing, say, a Google search? Yes, and no. Carruthers suggestion of *frugal principles* is rather akin to the A*-algorithm and best-first searches, where information is sorted through according to the weight (or cost) of individual nodes and their proximity to each other. The idea of frugal computation, however, is best explained as a rehash of Kahneman and Tversky's 1974 study of heuristics used by human subjects for, among other things, rough estimation of probabilities. For an explanation of how these heuristics work, I refer to their article. Suffice it to say that Carruthers' principle of frugality (involving brief instructions for scanning represented information for, say, a word or phrase, as well as stopping rules for terminating the search within a reasonable amount of time) is quite elegant but does not by itself suggest anything like a modular architecture of the mind.

II.1 Fewer features for a stronger thesis

Carruthers first version of his list comprises five items, and is derived from a critical inventory of Fodor's. These are the remaining features: (1) isolable, function-specific processing, (2) mandatory operation (3) specific (or distinct) neural realisation, (4) informational encapsulation, and (5) inaccessibility. Concerning (1), it is interesting to note that while Carruthers rejects "domain specificity" (where 'domain' is interpreted as "subject-matter" or "type of information", which is Fodor's original intention) for being too restrictive for a massive version of the modularity thesis, his alternative can hardly be thought of as anything but a substitution for what evolutionary psychologists mean by 'domain' – namely "function". Thus, he retains the feature as such but changes its extension from subject-matter to task.

By the end of "The case ...", however, the list has gone through some changes. It now states that a candidate system will exhibit at least three of the following features: (1) (for all) a distinct

function or set of functions, (2) (for all) a distinct neural realisation, (3) (for many) significant innateness or genetic channelling, (4) (for many) processing algorithms unique to the system, and finally (5), (for all) frugal operation. The Fodorian counterparts to features (1), (2), and (3) are easily seen – domain specificity, localisability, and ontogenetic determination, respectively –, and (5) is an interpretation of what informational encapsulation ought to mean. Note the difference between Fodor's "dedicated neural architecture" and Carruthers "distinct neural realisation": above, we saw that "dedicated" in Fodor's text implies a strong mind-brain idendification. Carruthers, however, was said in the introduction to be much more lax about such things. And he is: all Carruther's choice of words imply is that it would show up clearly in an *f*MRI-scan. Such distinctness does not entail that the location of a module in one subject should be generalisable to the whole population, nor even that the region in which it was found should be invariant between subjects. That move, I think, is both subtle and well-motivated. Now, feature (4) seems to be a new addition, unless of course it is a variant of mandatoriness: the processing of a certain class of information by a unique algorithm would indeed be mandatory, since there would be no alternative systems to complete the task. But be that as it may. Carruthers' list contains five or six elements, out of which three are necessary, which makes the absolute minimum number of types 10 or 11.

Whether these types of system at the end of the day actually deserve the name 'modular' is not something Carruthers worries about very much: at the end of "The case ..." he casually remarks that similar sorts of systems are commonly called 'modules' in AI, but that this sense is more unusual in philosophy and psychology. Despite his confidence, however, that discrepancy is indeed a source of worry: Successful models for *artificial* intelligence (or designed minds) need not in any way resemble or to an interesting degree describe *sponaneously evolved* intelligence (or non-designed minds), as the latter may well turn out to be organised in a very different way: the designer of an artificial neural network or a piece of software has a plan, whereas evolution does not. Nonetheless, Carruthers' shorter list makes the massive version of the modularity thesis stronger than the original. Firstly because the number of types is smaller, and secondly because it concerns *most or all* of the mind. At the same time, the resulting sense of 'module' is very much weaker – not as well-defined – due to the reduced number of features in the list.

III. Prinz' Empiricist Critique

"When Fodor titled his (1983) book *The Modularity of Mind*, he overstated his position. His actual view is that the mind divides into systems some of which are modular and others of which are not. The book would have been more aptly, if less provocatively, called *The Modularity of Low-Level Peripheral Systems.*" These are the opening words of Prinz sceptical review "Is the mind really

modular?". He begins its first proper section by calling attention to the important difference between it and the uncontroversial assumption that the mind is functionally decomposed. He describes it with a single sentence as the simple idea that " the mind contains systems that can be distinguished by the functions they carry out."

Let us have a closer look at that point of view before moving on: As the mind performs a great many and shifting functions in a non-linear and simultaneous manner, a sensible explanation of how such decompositon works might be that the mind allocates systems apt to perform these particular functions. Then, the difference between functional decomposition and (massive) modularity is that in the first case, we are dealing with a multi-purpose system (recall the example of such a system, *S(mp)*, having a set of applications $<a_1 \ldots a_n>$, from section II.1 above), wheras in the second we are dealing with a set of distinct sub-systems (i.e, $<s_1 \ldots s_n>$). A more elaborate interpretation of functional decomposition would add that these systems (or applications of a system) – which might be permanent or temporary – may well be dynamic both regarding their structure and the (number of) functions they are capable of performing. These systems need not be exclusively dedicated nor specialised at all – strictly speaking, they need not even exhibit a single one of the features on either Fodor's or Carruther's lists. It suffices that they are, for some reason or other, good candidates for performing the function. The set of neurons included therein may well vary depending on the character of the function that needs to be performed: individual neurons in all likelihood partake in the performance of many different tasks, due to being included in a range of operational constellations. We shall revisit this line of thought below.

Back to Prinz. "Fodor's criteria can be interpreted in different ways" he remarks, suggesting first that "[p]erhaps a system is modular to the extent that it exhibits properties on the list." I take it he had Fodor's 'most' (in the numeral sense) in mind here. Another possibility, according to which "some of the properties may be essential, while others are merely diagnostic" alludes to the sort of conspicuously missing threshold-clause that we attempted to define in section I.1 above, using Robbins' hierarchic list to determine the relative weights of the elements in the set of features. Since Fodor in a later text (*The Mind Doesn't Work that Way*, 2000) "treated informational encapsulation as a *sine qua non* for modularity" (Prinz' italics), it seems that he, too, had in mind a something like a hierachic version of his original list. So, as we noted above, encapsulation has the greatest weight both according to Fodor and to his critics. Here we might note parenthetically that with the exception of Peter Carruthers "[d]efenders of massive modularity focus on domain specificity in ontogenetic determination" and so seem to have in mind a different sort of hierarchy.

Prinz goes on to discuss the controversial features in accordance with Fodor's implication that some of them "cluster together", laconically remarking that he (Prinz) is sceptical about them

whether they are thought to occur jointly or individually. According to him, neither alternative provides a reasonable way to "circumscribe an interesting class of systems".

So far we have not yet moved beyond the first section of Prinz' insightful and engaging critique, but already we seem to have found support for our above discussion of the consequences of having a number of distinct types of modules, as well as of possible strategies – notably thresholds pertaining to both the size of the subset (of features of a viable candidate for modularity) and the weight of its members – for determining that number. As said, the consequences of treating 'module' as an umbrella term are enough to make one more than a little uneasy about the utility of the theses, as well as about any form of "modularity of mind" deserving its name. Clearly, Prinz shares in this uneasiness, and has very good reasons indeed for his qualms. In what follows, therefore, his analyses of the pairs and trios of features that would seem to "cluster together" (as well as of individual features) will be related only in brief, focusing on his main reasons for rejecting each of them. If the character of uneasiness in Prinz' opening section was theoretical, however, that of his analyses of the "clusters" is markedly practical and empirical in nature.

III.1 Rejection: Nine features, nine failures

In sum, Prinz' reasons for rejecting Fodor's (and Carruther's) proposals are largely empirical. There is not any conclusive evidence to be had in support for any one of the features on their lists, but merely what might (with a bit of a squint) be interpreted as such. The analysis of localisation and characteristic breakdown makes a great example of Prinz' critique: "The claim that mental faculties [i.e., in this case vertical faculties, or modules] are localized is supported by the fact that focal brain lesions cause selective deficits." So far all seems well. Neuroimaging studies provide "further evidence" in support, as they "purport to pinpoint the brain areas that are active when healthy individuals perform mental tasks." These areas are often thought to be well-defined, at least by non-neuroscientists, when they really are not. For instance, Broca's area is not even precisely located (i.e., researchers are not in agreement about exactly where it is), and it may well turn out that it *cannot* be precisely located other than in single individuals. The region in which it is found is probably fairly universal, but narrowing it down more than that could easily be a hopeless task.

Likewise, studies of brain lesions are not easily generalised to the entire population, seeing as they are performed on individual patients under very specific circumstances that commonly are not replicable. Neither are the effects of these lesions easily predictable (not that mere difficulty is a sound argument against anything), as "[w]ell-known deficits, such as visual neglect, are associated with lesions in entirely different parts of the brain (e.g., frontal eye-fields and inferior parietal cortex)" and may therefore have more to do with the specific neuroanatomy of the patient than with

general neuroanatomy. Additionally, "[s]ometimes lesions in the same area have different effects in different people, and all too often neuropsychologists draw general conclusions from individual case studies. *This assumes localization rather than provid[es] evidence for it*." (My emphasis.)

It seems to me, in light of these remarks, that systems in the brain are not only somewhat dynamic as to their specific architecture – meaning that they are highly susceptible to effects of cerebral plasticity, say – and their spatial relations to each other; they also seem to be highly integrated in a patently non-modular way. However, Prinz does not wish to rule anything out of court: "I do not want to exaggerate the implications of these considerations. There is probably a fair amount of localization in the brain". I would like to underline that that amount most likely has more to do with regions than with areas, not to mention individual systems.

Decades of research have clearly borne out that neurons differentiate over time, but this does not support Fodor's and Carruthers respective theses any more than it supports a rejection of them. However, (many) defenders of modularity, especially those who are prone to think of domain specificity in conjunction with localisation, seem to think neuronal differentiation and cerebral decomposition into regions unequivocally rules in favour of their school of thought. But as Prinz remarks, "[i]f, in reality, mental functions are located in large-scale overlapping networks, then it would be misleading to talk about anatomical regions as modules". Doing so would not only entail that 'module' is little more than an umbrella term, it would also mean that that particular umbrella were held all too high to be of any use at all.

Let us now have a look at an interesting discovery obtained by using connectionist models: "a massively distributed artificial neural network can exhibit a selective deficit after a few nodes are removed (simulating a focal lesion), even though those nodes were not the locus of the capacity that is lost [...]". Now, while modularity is perfectly compatible with the dispersion of a specific system across neurons that do not connect directly with each other, or that at least are located relatively far from one another (and therefore might accommodate a discrepancy between the locus of a lesion and the locus of the resulting loss of function), the fact that such dispersal is allowed at all within that framework seems to deprive the technical concept 'module' of an important aspect that one really would expect to be there: Modules in the technical sense, that is, are not really modular in the non-technical sense. As Prinz remarks at the end of that section; "[o]ne could escape the localization criterion by defining modules as motley assortments of abilities (e.g., syntax plus orofacial control; social exchange plus *faux pas*), but this would trivialize the modularity hypothesis". Do we really need to go on? In principle, yes, but for present purposes it suffices to note that all of the remaining features are rejected by the same means. Mandatoriness, swiftness, and shallowness; ontogenetic determination; domain specificity; inaccessibility and encapsulation, all suffer the same fate: they

do not stand up either to theoretical or to empirical scrutiny. Notwithstanding, an illuminating quote regarding the final and, according to Fodor and Carruthers, most important feature is due:

> Computational systems can sort through stupendously large databases at breakneck speed. The trick is to use "frugal" search rules. Frugal rules are ones that radically reduce processing load by exploiting simple procedures for selecting relevant items in the database. Once the most relevant items are selected, more thorough processing of those items can begin. Psychologists call such simple rules "heuristics" (Kahneman et al., 1982). There is overwhelming evidence that we make regular use of heuristics in performing cognitive tasks. For example, suppose you want to guess which of two cities is larger, Hamburg or Mainz. You could try to collect some population statistics (which would take a long time), or you could just pick the city name that is most familiar. This Take the Best strategy is extremely easy and very effective; it is even a good way to choose stocks that will perform well in the market (Gigenzer et al., 1999). With heuristics, we can avoid exhaustive database searches even when a complete database is at our disposal. There are also ways to search though a colossal database without much cost. Internet search engines provide an existence proof (Clark, 2002). Consider Google. A Google search on the word "heuristic" sorts through over a billion web pages in 0.18 seconds, and the most useful result appears in the first few hits. Search engines look for keywords and for webpages that have been frequently linked or accessed. If we perform the mental equivalent of a Google search on our mental files, we should be able to call up relevant information relatively quickly. *The upshot is that encapsulation is not needed for computational tractability.*

My emphasis.

IV. Final Discussion

"There is no such thing as evidence, unless there is evidence that something is evidence." That is not the exact catchphrase with which Nancy Cartwright summarises her theory of evidence, but at the moment I cannot recall *verbatim* how she put it. At the first of her Pufendorf lectures at Lund University (27 May, 2012), however, she presented a theory according to which 'evidence' is to be regarded as meaningless until a significant collection of related 'evidence' has been organised and made sense of by a sound and valid argument. Only then, when it has been ordered and given a precise direction, can it be regarded as evidence. Before that has been accomplished, anything might be regarded as 'evidence' for anything else. This is reminiscent of what Sherlock Holmes said to Dr Watson in "A Scandal in Bohemia" (Conan Doyle, 1891): "It is a capital mistake to theorize before you have all the data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts." My only point with this note is that philosophy, especially the sub-fields thereof that

concern "the real world" (e.g., philosophy of mind), would benefit greatly from thinking along the lines indicated above in so far it is at all possible (this field of inquiry is, after all, seldom evidence-driven nor particularly concerned with the empirical world). Fodor and Carruthers did not, and even if their proposals are impressive and *prima facie* interesting in their own right, they turn out not to be true as such – their theories fail to order and direct the 'evidence' they use as support.[18]

IV.1 Outlines of a possible middle-ground theory

In light of the above, it must be noted that while Fodor's and Carruthers' vertical sensibilities do have their perks – such as a diminished workload for individual systems as well as a simple model for simultaneous processing of, say, memory bases (vaguely reminiscent of the blind watch-maker's side-by-side assembly of molecular components), suggesting greater operational efficiency –, the traditional, horizontal view probably yields a sounder picture overall, namely one of functional decomposition. To me, it seems as if some sort of a two-dimensional model would be preferable: a model with both vertical and horizontal features would accommodate task-specificity (or rather, say, the adaptability of a neuronal structure to suit a given task) within the confines of a general purpose processing mechanism. All in all, a viable theory of mental architecture (or organisation) would most likely draw almost equally from all of the extreme positions discussed above. Here is a list of what I take to be the salient features of such a middle-ground theory:

1. Input systems are generally *assembled* rather than *innate*, which stresses the need for configuration by way of exposure to relevant stimulus contexts. For instance, depth perception in infants (of whatever species) is not fully developed until a certain stage of their overall development, and in general, the senses of an infant are not nearly as differentiated as those of an adult (see Prinz, 2006). The same goes for self-perception and self-directed behaviour in those species for which this is a real possibility. (Prior et al., 2008; Plotnik et al., 2006) For human infants, this occurs between about 18 months and three years of age.

2. The neuronal substrate of systems of the mind is *not* generally fixed.

3. Carruther's "cycles of reliably recurring [types of] operations" or processes are probably restricted to systems and sub-systems handling input or the mind-world link for reasons of

18 It should be said, however, that the bibliography in Fodor (1983) is impressive, and that his handle on the numerous studies he cites in support for his theory is remarkable. But this might be explained by his eagerness to fit them into his framework. At some points, too, the conclusions drawn by the authors of a given study are all too ambitious and groundlessly "proto-modularist".

efficiency and the non-conceptuality of basic sensation and perception. For instance, the sort of heuristics discovered by Amos Tversky and Daniel Kahneman in the 70's apply to many different and generally non-eccentric "domains" (see for example their (1974)).

4. There is an important distinction between *assembly* and *learning* which must be taken into account; in general, the former applies to input systems and the latter to central systems.

5. *Frugal computation* and *limited central access* should be thought of as flexible and dynamic properties of a given system, where the former enables quick processing and the latter is a pragmatic consequence: there simply is no need for most processes to be consciously accessible.

6. The mind and its architecture/organisation is highly adaptive, within reasonable bounds such as specific (neuro-)anatomy and environmental demands, suggesting a sort of 'mental plasticity' akin to that of the brain. Mental plasticity, however, is not necessarily parallel to that of the brain. Ways of processing certain types of information/(eccentric) data can be re-learned/re-configured given proper training (here: characterised by a high degree of structure and planning) or exposure (here: characterised by a high degree of spontaneity or randomness, or a distinct *lack* of pre-ordained structure) to new or different settings and conditions over a significant amount of time. Such training/exposure involves a significant number of repetitions (indeterminate, perhaps), and the process of re-learning or re-configuration may be either continuous or interspersed with periods of "rest". (Compare, for example, with Kuhnian paradigm shifts and the "rites of passage" required to enter into a new "world" of scientific research. (Kuhn, 1996)).

A semi-strong interpretation of the above points would be preferable, since the resulting theory must both be interesting enough in its own right as well as admit of revision and, indeed, its own eventual replacement by a better candidate.

References/Bibliography

Arthur Conan Doyle, *A Study in Scarlet*, 1887

_____, "A Scandal in Bohemia", in *The Adventures of Sherlock Holmes*, 1892

Jerry Fodor, *The Modularity of Mind*, The MIT Press, 1983

_____, *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*, The MIT Press, 2000

Robert J. Stainton (ed.), *Contemporary Debates in Cognitive Science*, Blackwell 2006

Peter Carruthers, "The Case for Massive Modular Models of Mind", in Stainton (2006)

Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed., The University of Chicago Press, 1996 (1962, 1970)

Jesse J. Prinz, "Is the Mind Really Modular?", in Stainton (2006)

Michael Posner, *Chronometric Explorations of Mind*, Lawrence Erlbaum Associates, 1978

Prior et al. 'Mirror-Induced Behavior in the Magpie (*Pica pica*): Evidence of Self-Recognition' in *PLoS Biol* 6(8) 2008: e202.

Plotnik et al., 'Self-Recognition in an Asian Elephant' in *PNAS*, vol. 104 no. 45, 2006, pp. 17053-57

Herbert A. Simon, "The Architecture of Complexity", in *Proceedings of the American Philosophical Society*, vol. 106, no. 6, 1962, pp. 467-82

A. Treisman & G. Gelade, "A Feature-Integration Theory of Attention", in *Cognitive Psychology* 12, 1980, pp. 97-136.

Amos Tversky & Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases", in *Science* vol. 185, no. 4157, 1974, pp. 1124-31.

Online resources:

Stanford Encyclopedia of Philosophy, entry: "Modularity of Mind" (ed. by Philip Robbins (2009))
URL: http://plato.stanford.edu/entries/modularity-mind/, first published on April 1, 2009. Retrieved on December 28, 2012, 02:02 a.m. CET+1.

URL for Prior et al, 2009: http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0060202

URL for Plotnik et al., 2006: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636577/

Appendix[19]

An improvised sort of weighted combinatorics was used to estimate how many types of modules there might be under certain obligatory conditions. To be precise, the question is how many times different subsets with a size or cardinality of at least |B| can be taken from an original, well-ordered set with cardinality or size |A|. Each member of A $\{k_1 \ldots k_n\}$ is weighted according to its place in the set, and the total weight of A is found by summation:

$$1 + 2 \ldots + n$$

The total weight of A is given as W(A). A threshold clause to the effect that at least a part *x* of W(A) is necessary for eligible candidate subsets; the value of *x* must be positive and cannot exceed W(A), but is otherwise arbitrary. W(A) is bolded in the tables below.

Nota bene: Since such a threshold clause can be satisfied by some (but not all) subsets B *whose minimum cardinality is significantly less than* |A|*, the number of subsets of that size that* in fact *do so most likely differs greatly from the below indication of the absolute minimum. The threshold, that is, does* not *strictly translate to a percentage of all subsets available; those clearing the threshold are, in their turn, elements of a subset* C. *Finding out the exact size of* C *is a knapsack problem whose solution requires quite advanced mathematics. I have therefore only indicated whether* one*,* at least two *or* all *subsets* B *of a given size clear the threshold to give a (very) rough idea of how many types result from implementing such a clause.*

I. Fodor-modules

Note: Fodor does not give any indication as to the minimum number of features an F-module must exhibit.

Available (B)	*Min. W(B)*	*Max. W(B)*	*Threshold = 23*	*Threshold = 35*
C(9,5) = 126	1 + 2 ... + 5 = 15	5 + 6 ... + 9 = 35	≥ 2	1
C(9,6) = 84	1 + 2 ... + 6 = 21	4 + 5 ... + 9 = 39	≥ 2	≥ 2
C(9,7) = 36	1 + 2 ... + 7 = 28	3 + 4 ... + 9 = 42	36	≥ 2
C(9,8) = 9	1 + 2 ... + 8 = 36	2 + 3 ... + 9 = 44	9	9
C(9,9) = 1	1 + 2 ... + 9 = **45**	1 + 2 ... + 9 = **45**	1	1

(I.1) If |B| = 5, there are 126 + 84 + 36 + 9 + 1 = 258 subsets.

(I.1') With a threshold at 23, the number of subsets above it is *at the very least 50*.

(I.2) If |B| = 7, there are 36 + 9 + 1 = 46 subsets.

(I.2')With a threshold at 35, the number of subsets above it is *at the very least 15*.

II. Carruthers-modules

Note: On page 23 of his (2006), Carruthers indicates that three features are mandatory for all C-modules.

II.1 First list (six elements)

19 I am grateful to Ralf Northman for providing valuable criticism on this section. Any remaining errors are, of course, entirely my own.

Available (B)	*Min. W(B)*	*Max. W(B)*	*Threshold = 11*	*Threshold = 15*
C(6, 3) = 20	1 + 2 + 3 = 6	4 + 5 + 6 = 15	≥ 2	≥ 2
C(6, 4) = 15	1 + 2 ... + 4 = 10	3 + 4 ... + 6 = 18	≥ 2	≥ 2
C(6, 5) = 6	1 + 2 ... + 5 = 15	2 + 3 ... + 6 = 20	6	6
C(6, 6) = 1	1 + 2 ... + 6 = **21**	1 + 2 ... + 6 = **21**	1	1

(II.1.1) If |B| = 3, there are 20 + 15 + 6 +1 = 42 subsets.

(II.1.1') With a threshold at 11, the number of subsets above it is *at the very least 11*.

(II.1.2) If |B| = 4, there are 15 + 6 + 1 = 22 subsets.

(II.1.2') With a threshold at 15, the number of subsets above it is *at the very least 11*.

II.2 Second list (five elements)

Available (B)	*Min. W(B)*	*Max. W(B)*	*Threshold = 8*	*Threshold = 11*
C(5, 3) = 11	1 + 2 + 3 = 6	3 + 4 + 5 = 12	≥ 2	≥ 2
C(5, 4) = 11	1 + 2 ... + 4 = 10	2 + 3 ... + 5 = 14	≥ 2	≥ 2
C(5, 5) = 1	1 + 2 ... + 5 = **15**	1 + 2 ... + 5 = **15**	6	6

(II.2.1) If |B| = 3, there are 11+11+1 = 23 subsets.

(II.2.1') With a threshold at 8, the number of subsets above it is *at the very least 10*.

(II.2.2) If |B| = 4, there are at least 11+1 = 12 subsets.

(II.2.2') With a threshold at 11, the number of subsets above it is *at the very least 10*.