# Modularity and cognition

Max Coltheart

**Modularity is a concept central to cognitive science, and Fodor's analysis of cognitive modularity in his book *The Modularity Of Mind* has been widely influential – but also widely misunderstood. It is often claimed that the possession of some or other system-property is a necessary condition for that system to be modular in Fodor's sense, but Fodor made it clear that he was not proposing a definition of modularity, nor proposing any necessary conditions for the applicability of the term. He was simply suggesting a number of system properties that are *typical* of modular systems. I argue that it is nevertheless possible to derive a useful definition of modularity from the kinds of arguments put forward by Fodor: A cognitive system is modular when and only when it is domain-specific. Given any such proposed module, the other features of modularity discussed by Fodor should be dealt with as empirical issues: for each feature (innateness, for example), it is an empirical question whether or not the proposed module has that feature.**

The concept of modularity has been central in cognitive science since the birth of the discipline. For example, David Marr[1] considered the concept so important that he 'was moved to elevate it to a principle, the principle of modular design' (Ref. 1, p. 102) according to which 'any large computation should be split up into a collection of small, nearly independent, specialized subprocesses' (p. 325), and cognitive scientists commonly refer to their models of cognition as being modular in nature.

The concept of modularity looks simple enough in, say, computer programming or engineering, but is likely to be more complex in cognitive science. Marr did not discuss exactly what might be meant by modularity in relation to cognition; but this was done soon afterwards, by Fodor in his book *The Modularity of Mind*[2]. In this book, Fodor proposed a completely general theory of perception and cognition and listed certain features that he took to be characteristic of modules (see Box 1). I will begin by considering Fodor's view of the logical relationship between these features of cognitive systems and the property of modularity.

**Necessary conditions for modularity?**

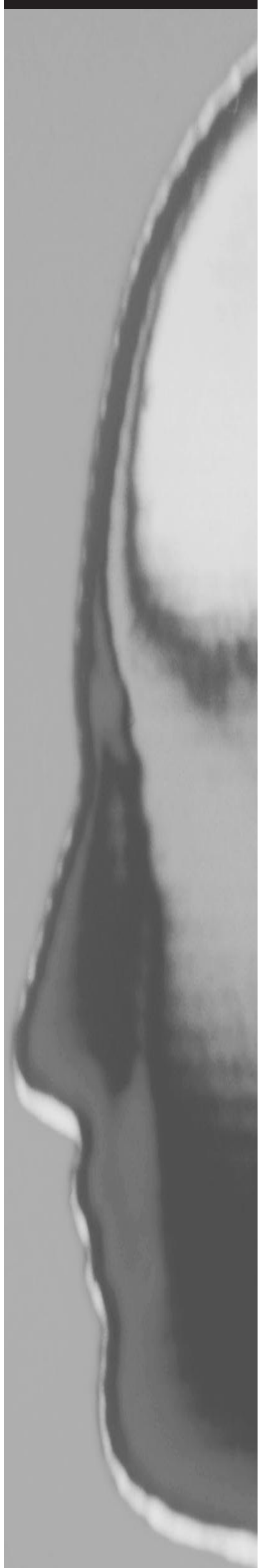
It is clear that many authors have understood Fodor to be proposing that the features of modularity listed in Box 1 are necessary conditions for the applicability of the term – that is, a cognitive system is not a module unless it possesses all of these features.

For example, Wojciulik, Kanwisher and Driver[3] found in an fMRI study that neural activity in the fusiform gyrus in response to faces was modulated by attention, and took this result to be surprising because that region is supposed to be modular in nature, and therefore its response must be mandatory (according to these authors' reading of Fodor) and so not modifiable by attention.

A commentary on these findings[4] observed 'This result has implications for a strictly modular view of the face recognition system. Because it shows that face-specific fusiform gyrus activations are not purely automatic and unmodulated, this study throws doubt on such a strict modular view of face recognition'. Again, the reasoning here is that one of Fodor's features of modularity is automaticity, and if this is a necessary condition for a system to be called modular, then the face recognition system of the fusiform gyrus is not a module.

Similar reasoning is used by Bishop[5]. She argued that the cognitive-neuropsychological approach is not well-suited to the study of developmental disorders of language, because that approach assumes that much of our cognitive processing is modular in Fodor's sense. She asserts that this assumption does not hold for children who are developing language: 'Overall, Fodor's notion of a modular system that operates entirely in bottom-up mode might be a reasonable characterization of the stable state that is achieved in the adult once language is fully learned…; however, it provides an unrealistic model of processing in the child who is developing language, where there is ample evidence of top-down processing… A similar point has been made by Hulme and Snowling (1992) who concluded that "interaction between systems is probably the norm in development, and it may only be after a very extensive period of development that the relative modularity or autonomy of different systems in the adult is achieved".' (Ref. 5, p. 906).

Here both Bishop, and Hulme and Snowling[6], are arguing that, because (a) autonomy or purely bottom-up

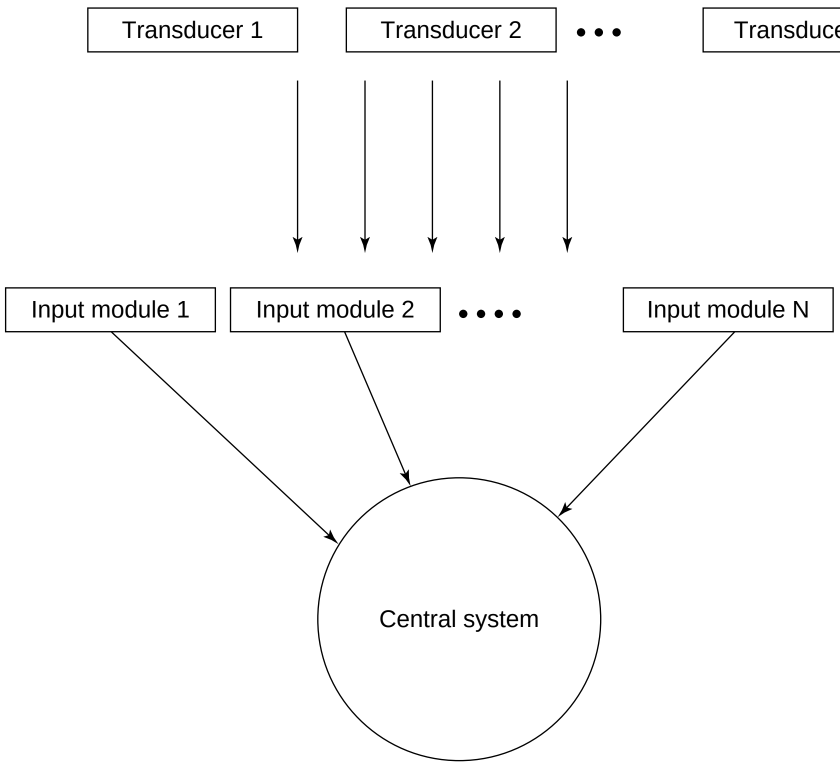


*M. Coltheart is at the Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia.*

**tel: +61 2 9850 8086**
**fax: +61 2 9850 6059**
**e-mail: max@currawong.bhs.mq.edu.au**

1364-6613/99/$ – see front matter  PII: S1364-6613(99)01289-9

## Box 1. Fodor's model of perception and cognition




In this model three levels are distinguished:

(a) The transducers, whose function is to convert physical stimulation into neural signals.

(b) The input systems, whose function is to interpret transduced information. They are responsible for basic cognitive activities such as language and vision, and they are modular.

(c) The central system, which is responsible for more complex cognitive activities such as analogical reasoning, and it is not modular.

Most of Fodor's book (Ref. a) is devoted to the input systems and what it means to say is that these are modular. Fodor (pp. 37, 101) does this by providing a list of features that characterize modular systems.

Modules are:
• domain specific
• innately specified
• informationally encapsulated
• fast
• hardwired (neurally specific)
• autonomous
• not assembled

**Reference**

**a** Fodor, J.A. (1983) *The Modularity of Mind*, MIT Press/Bradford Books

On-line, see Fig. I

processing is, according to Fodor, a necessary condition for the application of the term 'modular', and (b) cognitive neuropsychology is committed to the Fodorian concept of modularity, and (c) the language performance by children shows evidence of top-down processing or non-autonomy, therefore, (d) the cognitive-neuropsychological approach is intrinsically unsuitable for studying developmental language disorders.

My final illustration of a line of argument in which Fodor's features of modularity are construed as necessary conditions for a system to be considered as modular comes from Farah[7]. She refers to 'Fodor's main defining criterion for modularity, information encapsulation' (Ref. 7, p. 90). This idea that a system can only be considered as modular if it possesses the property of information encapsulation is crucial to her paper, because she wants to argue, from evidence that she takes to demonstrate violations of information encapsulation, that many cognitive processing systems are not modular – any such violation 'counts as evidence against modularity' (p. 59). That would only be true, of course, for an account of modularity according to which information encapsulation is a necessary condition for modularity.

We have seen, then, that both Bishop and Farah construe Fodor as proposing to *define* modularity. But Fodor[2] explicitly disavowed any such intention:

> 'I am not, in any strict sense, in the business of defining my terms… So what I propose to do instead of defining 'modular' is to associate the notion of modularity with a pattern of answers to such questions as 1–5'. (Fodor, p. 37; the five questions were about possible features of modularity.)

So it is a mistake, and it turns out to be an important mistake, to think that Fodor was offering a definition of modularity. It is also a mistake to think that he was proposing any necessary conditions for the applicability of this term. Just as he explicitly indicated that he was not offering a definition of modularity, he also explicitly indicated that, on his view, none of the features of modularity he discussed was necessary for the term to be applied:

> 'One would thus expect – what anyhow seems to be desirable – that the notion of modularity ought to admit of degrees. The notion of modularity that I have in mind certainly does. When I speak of a cognitive system as modular, I shall therefore always mean "to some interesting extent"'. (Fodor, p. 37.)

and, even more clearly:

> 'One interpretation might be this: given that a system has any of the properties in question, then the likelihood is considerable that it has all the rest… However, I doubt that a claim that strong could be empirically sustained, since it is reasonably easy to think of psychological processes that are fast but not encapsulated, or involuntary but not innate, and so forth. The present contention, in any event, is relatively modest: it's that if a psychological system has most of the modularity properties, then it is very likely to have all of them'. (Fodor, p. 137.)

Clearly, then, on Fodor's view of modularity, evidence for non-autonomy, or for violation of information encapsulation, is not evidence for non-modularity. That is not to say that he gave equal importance to all of the features of modularity; for him, 'The informational encapsulation of the input systems is… the essence of their modularity'. (Fodor, p. 71.)

**What is information encapsulation?**

Bishop[5] and Hulme and Snowling[6] both argue that Fodorian modularity is inconsistent with top-down processing, because such processing constitutes a violation of information encapsulation. They then go on to argue that, because top-down processing is characteristic in children when they are dealing with language, the concept of modularity is inappropriate in a developmental context. Both concede, however, that this is not a problem as far as the skilled language user is concerned because skilled language may well be purely bottom-up.

Farah[7] also claims that the occurrence of top-down processing is inconsistent with Fodorian modularity, but goes even farther than Bishop, and Hulme and Snowling: the basis of her article is the claim that even the skilled cognizer makes much use of top-down processing, and hence that Fodorian modularity does not hold even in the case of skilled cognition.

However Fodor himself not only discussed top-down processing but also took the view that it was perfectly compatible with the view of modularity he was proposing:

> '*The claim that input systems are informationally encapsulated must be very carefully distinguished from the claim that there is top-down information flow within these systems…* phoneme restoration provides considerable prima-facie evidence that phoneme identification has access to what the subject knows about the lexical inventory of his language. If this interpretation is correct, then phoneme restoration illustrates top-down information flow in speech perception. It does not, however, illustrate the cognitive penetrability of the language input system. To show that that system is penetrable [hence informationally unencapsulated], you would have to show that its processes have access to information that is not specified at any of the levels of representation that the language input system computes… If… the 'background information' deployed in phoneme restoration is simply the hearer's knowledge of the words in his language, *then that counts as top-down flow within the language module*'. (Fodor, p. 76; my emphases.)

The mistake here has been a failure to understand what Fodor meant by information encapsulation. The basic virtue of information encapsulation, according to Fodor, is that it enables input systems to do their jobs quickly. This would not happen if a person's input systems had access to, and sought to use, everything that the person knew. But the kinds of information from which modules are encapsulated, according to Fodor, are of very particular kinds – a person's general beliefs, desires and utilities:

> 'There are some prima-facie reasons for doubting that the computations that input systems perform could have anything like unlimited access to high-level expectations and beliefs'. (Fodor, p. 66.)

and:

> 'Pylyshyn (1980) speaks of the 'cognitive impenetrability' of perception, meaning that the output of the perceptual systems is largely insensitive to what the perceiver presumes or desires'. (Fodor, p. 68.)

It simply doesn't follow from this that top-down feedback from, say, lexical to phonetic levels in the language processor counts as a violation of information encapsulation, as Bishop (Ref. 5, pp. 905–907) claimed. On the contrary, it is precisely this lexical-to-phonetic feedback that Fodor proposes as the explanation of the phoneme restoration effect.

All of the examples which Farah[7] offers as examples of violation of information encapsulation fail for the same reason. None represents penetration of a module by a person's expectations, beliefs, presumptions or desires; and that is what violation of information encapsulation is, according to Fodor's account of modularity.

However, I think Fodor is partly to blame for the misunderstanding here, for failing to develop one aspect of his views that would have helped avoid such misunderstandings. This has to do with his idea that modules are 'not assembled'.

**Could modules be 'assembled'?**

In the preliminary discussion in his book, Fodor poses the following question. If a computational system is a module, could it be

> '…"assembled" (in the sense of having been put together from some stock of more elementary subprocesses) or does its virtual architecture map relatively directly onto its neural implementation?' (Fodor, p. 37.)

His answer was that modules are not assembled. It is interesting to note here not only that he gives no reason for making this choice, but also that, when he provides a more detailed list of the features of modules later in the book, 'not assembled' does not appear as one of these. I infer from this that his commitment to this particular feature of modularity was less than fervent.

This might be just as well, because, in my view, there is an inconsistency between the property 'not assembled' and Fodor's account of top-down processing within modules, and, more generally, between this feature and his general idea of interlevels of representation within modules. Figure 1 gives three examples of this general idea.

Fodor is explicit about how the existence within a module of distinct processing levels, which communicate with each other, does not constitute a violation of the principle of information encapsulation. For example, discussing the model of visual object recognition set out in Fig. 1C, he says:

> 'To develop such a model would be to show, in detail, how an informationally encapsulated visual processor could perform object identification at the basic category level. That would make modularity theorists very happy'. (Fodor, p. 137.)

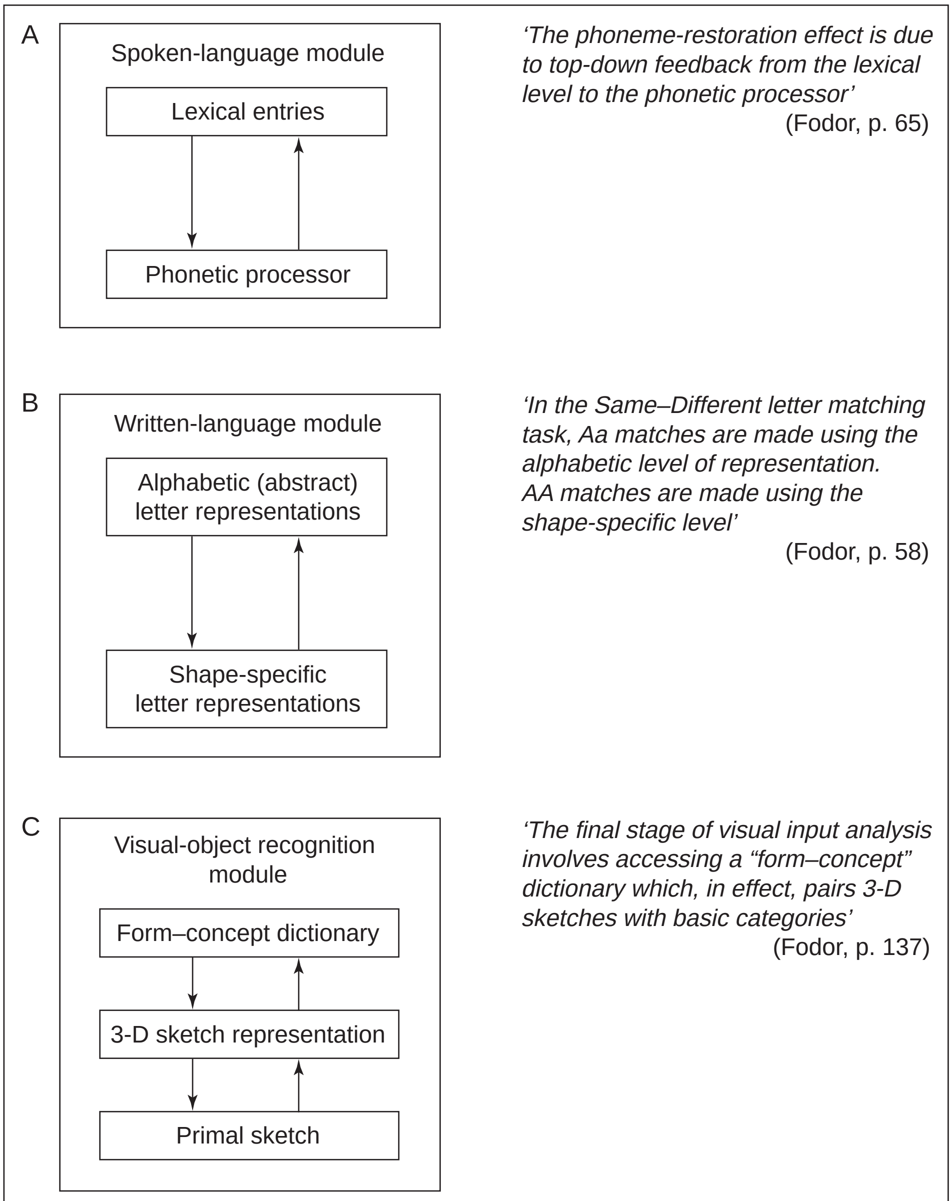



**Fig. 1. Three examples of within-module interlevels of representation in Fodor's account of cognition.**

In all of the examples in Fig. 1, each interlevel within a module has a particular domain to which it is specific: the phonetic versus the lexical, the visual versus the abstract letter representation, or the 3-D sketch versus the form concept. Inside the module, then, there are distinct domain-specific subsystems. Why don't such subsystems deserve to be called modules? Why not propose that inside the language module there is a phonetic-analysis module and a lexical-form module?

Exactly this has been proposed by Block[8]; for him, a cognitive module may be decomposed into a set of smaller modules, and concerning this decomposition Block says: 'Decomposition stops when all the components are primitive processors – because the operation of a primitive processor cannot be further decomposed into suboperations. For example, an AND-gate: it is just defined in terms of its input–output function, and that function is not decomposable. Primitive processors are the only computational devices for which behaviorism is true – the largest components of the system whose operation must be explained not in terms of cognitive science, but in terms of a realization science such as electronics or physiology… If the mind is the software of the brain, then we must take seriously the idea that the functional analysis of human intelligence will bottom out in primitive processors in the brain'.

To refuse to countenance the idea that a module itself can be composed of modules is to adopt the view that a cognitive system can be domain-specific yet not be a cognitive module. Requiring that modules not be assembled is the only obstacle here. Abandoning this particular requirement, one in any case for which Fodor offered no rationale, avoids difficulties for Fodor's account of interlevels of representation. An even greater advantage is that it allows one to do something which Fodor did not do – to offer a definition of modularity.

**A neo-Fodorian account of modularity**

First, it is helpful here to distinguish between 'knowledge module' and 'processing module'. A knowledge module is a body of knowledge that is autonomous: independent of other bodies of knowledge; for example, a linguist might say 'syntax is a module'. A processing module is a mental information-processing system; for example, a psycholinguist might say 'during language comprehension, sentences are parsed by a syntactic processing module'. I am concerned here with processing modules, not knowledge modules. Of course, processing modules will generally incorporate knowledge modules – the syntactic processor will have, as part of its internal structure, a body of knowledge about syntax, just as the language processor in Fig. 1 has, as part of its internal structure, a lexicon.

Fodor was not concerned with defining modularity, nor with specifying any properties that a cognitive system must necessarily possess for it to be considered modular. I am suggesting that one can be more ambitious than this, by defining 'module' as 'a cognitive system whose application is domain specific'; here domain-specificity is a necessary condition for the applicability of the term 'modular'.

Now it is necessary to say something about what might be meant by 'domain-specific'. I mean that a cognitive system is domain-specific if it only responds to stimuli of a particular class: thus, to say that there is a domain-specific face-recognition module is to say that there is a cognitive system that responds when its input is a face, but does not respond when its input is, say, a written word, or a visually-presented object, or someone's voice. There is no circularity here, because the claim that there is a face-recognition module does not derive merely from the existence of faces as a stimulus class that is conceptually distinguishable from other stimulus classes such as written words, objects or voices. The claim is derived from empirical observations, as follows.

Suppose that we entertained the idea that there was a single module for visual recognition which accepts as inputs faces, objects and printed words. Then we noticed that in the neuropsychological literature there were reports of patients with impaired visual word recognition but who retained face recognition[9] and of patients with impaired visual word recognition but who retained visual object recognition[10]. We also noticed reports of patients with impaired visual object recognition but who retained face recognition or visual word recognition[11]. Finally, we also noticed reports of patients with impaired face recognition but who retained visual object recognition[12]. This collection of results refutes our original idea that there is a cognitive module whose domain is the recognition of all forms of visual stimuli. Instead, it suggests that there are three separate modules: a face-recognition module, a visual-object-recognition module, and a visual-word-recognition module (a claim that I believe is in fact correct). The domain specificities are obvious here: visual objects, faces, and printed words, respectively.

'Informationally encapsulated' and 'domain-specific' are different concepts. I follow Fodor in using 'informationally encapsulated' to mean 'not having complete access to a person's expectations, beliefs, presumptions or desires'. That is quite a different idea from 'not responding to inputs except those of a particular class', which is what I mean by 'domain-specific'. Hence if one claims that a certain cognitive processing system is domain-specific, and therefore is a module, it is perfectly sensible then to go on to consider whether this system is or is not informationally encapsulated. When we recognize a face, do we draw upon information represented by our expectations, beliefs, presumptions or desires, or is face recognition performed with no reference at all to these? At bottom, this question about the face-recognition module is an empirical one, though one can make the kinds of arguments that Fodor does about the adaptive advantage of having object recognition systems that are fast, pointing out that an object-recognition system that did consult all of the perceiver's knowledge is unlikely to be fast enough to generate appropriately swift recognition of a predator.

More generally, having inferred, on the basis of empirical evidence, the existence of some cognitive system that is a module according to the definition of modularity as domain-specificity, one can then consider all of the other features of modularity described by Fodor in a purely empirical way. That is, none of these other features is a *necessary* feature of modules in general, so whether the particular module proposed has any one of these features is simply an empirical question – and usually a very interesting one.

It is likely that modules as defined in this way will be fast-acting, and that their operations will be mandatory (automatic), not only for the evolutionary reasons discussed by Fodor, but also because a system that need respond only to one specific class of stimulus can be tuned to respond with special efficiency, provided that its operation does not have to be controlled by some slower system such as Fodor's central system. To the degree to which such control is absent, the operation of the module is mandatory. Consider the Stroop task: a word printed in a particular colour is shown, and an observer is required to name the ink colour. When the stimulus word is a colour-related word, naming the colour of the ink is faster when ink colour and word agree (e.g. the colour-naming response is faster for 'sky' printed in blue than for 'coal' printed in blue). This indicates that observers cannot prevent print colour from accessing meaning, even when such access is to their disadvantage, which is evidence that visual-word recognition followed by access to the meaning of the word is a mandatory (automatic) process. Similarly, colour naming is slower when the word shares no phonology with the colour name than when it shares some phonology (e.g. when the ink colour is red, colour-naming is slower to the word 'cat' than to the word 'rat'[13]). This indicates that observers cannot prevent print colour from accessing phonology, even when such access is to their disadvantage – evidence that access to the phonology of a printed word is a mandatory process.

Notice that what is meant here by 'mandatory' is 'cannot be completely turned off' (rather than 'completely impervious to any control from higher centres'). Because subjects can successfully name the ink colour or read the word, according to the instructions they are given, they must have some control over the processes that access phonology from colour and phonology from text. The existence of the Stroop effect is empirical evidence that such control is not complete, and that is what is meant here by referring to processes as mandatory.

A further point is that, as Fodor says, there is no necessity that a cognitive module be associated with localized fixed neural architecture:

> 'Is faculty psychology literally incompatible with, say, an equipotential brain? Remember that faculties are, in the first instance, functionally rather than physiologically defined'. (Fodor, p. 98.)

Once again, it is an empirical question whether any hypothesized functional module has this kind of neural representation. Fodor (pp. 98–99) provides some *a priori* reasons why it might be advantageous for modules to be localized in restricted brain regions, and there is much empirical evidence demonstrating such localization, from cognitive-neuropsychological studies of selective effects of localized brain damage on cognition.

If, however, it had turned out that in people with selective cognitive deficits one could never attribute the deficit to damage in some specific and localized region of the brain, that would just be an empirical fact about the neural representation of cognitive modules, not evidence against the view that the mind has modules. It would be relevant to some concept of neuroanatomical modularity, but not to the concept of cognitive modularity.

It cannot be the case that all modules are innate, but some might be. Modules for reading and writing, for example, cannot be innate, because these abilities are too new in evolutionary terms, and not ubiquitous among current members of the human species. But there is no reason why a theory-of-mind module, of the kind whose absence or impairment could be responsible for autism, could not have an innate foundation, as indeed some have suggested[14,15]. However, if there is a theory-of-mind module that is associated with autism in this way, whether this module is innate or acquired is an empirical question. And, of course, if one took the view that modules are domain-specific in the sense of responding only to a particular class of inputs, one would need to explain what this class is in the case of a theory-of-mind module: it might, for example, be the class of bodily movements that convey information about the mover's mental states.

If we allow that modules can be assembled – that is, that a module itself can have an internal modular structure – then another empirical question is whether information flow within this modularly organized system is purely feed-forward (bottom-up) or whether feedback (top-down processing, recurrence, interactive activation) also occurs. If there is such interactive activation, it would not represent any violation of informational encapsulation in the Fodorian sense of that term, and is in no way incompatible with the concept of modularity. Nor would there be anything odd about claiming that some modules use top-down processing and others do not. For each module, that is a matter to be decided empirically.

**Acknowledgements**
I thank Robyn Langdon, Anna Woollams, Greg Currie and Hugh Clapin for their comments on earlier drafts of this article, and participants at the ARC Workshop on Cognitive Simulation, Theory of Mind and Modularity, particularly Martin Davies, Jay Garfield, Philip Gerrans, Philip Pettit, Kim Sterelny and Daniel Stoljar, for valuable discussion and criticism.

**References**

1 Marr, D. (1982) *Vision*, W.H. Freeman
2 Fodor, J.A. (1983) *The Modularity of Mind*, MIT Press/Bradford Books
3 Wojciulik E., Kanwisher, N. and Driver, J. (1998) Covert visual attention modulates face-specific activity in the human fusiform gyrus: an fMRI study *J. Neurophysiol.* 79, 1574–1578
4 Monitor contribution (1998) Attention to faces in the fusiform gyrus *Trends Cognit. Sci.* 2, 205
5 Bishop, D.V.M. (1997) Cognitive neuropsychology and developmental disorders: uncomfortable bedfellows *Q. J. Exp. Psychol.* 50A, 899–923
6 Hulme, C. and Snowling, M. (1992) Deficits in output phonology: an explanation of reading failure? *Cognit. Neuropsychol.* 9, 47–72
7 Farah, M.J. (1994) Neuropsychological inference with an interactive brain: a critique of the locality assumption *Behav. Brain Sci.* 17, 43–104
8 Block, N. (1995) The mind as the software of the brain, in *Thinking: An Invitation to Cognitive Science* (Smith, E.E. and Osherson, D.N., eds), MIT Press
9 Chialant, D. and Caramazza, A. (1998) Varieties of pure alexia: the case of failure to access graphemic representations *Cognit. Neuropsychol.* 15, 167–201
10 De Renzi, E. and Di Pellegrino, G. (1998) Prosopagnosia and alexia without object agnosia *Cortex* 34, 403–415
11 Rumiati, R.I. and Humphreys, G.W. (1997) Visual object agnosia without alexia or prosopagnosia *Vis. Cognit.* 4, 207–217
12 Buxbaum, L.J. *et al.* (1996) Relative sparing of object recognition in alexia–prosopagnosia *Brain Cognit.* 32, 202–205
13 Coltheart, M. *et al.* A position-sensitive Stroop effect: further evidence for a left-to-right component in print-to-speech conversion *Psychonomic Bull. Rev.* (in press)
14 Leslie, A.M. (1992) Autism and the 'Theory of Mind' module *Curr. Dir. Psychol. Sci.* 1, 18–21
15 Roth, D. and Leslie, A.M. (1998) Solving belief problems: Towards a task analysis *Cognition* 66, 1–31

## *TICS* Editorial Policy

*Trends in Cognitive Sciences* is indispensable reading for all interested in this diverse field. The journal carries a core of authoritative **Reviews**, written in an easily accessible style by leading authors, summarizing the exciting developments of the subjects you need to know about. Accompanying the Reviews are a range of other types of article. The **Comment** section offers short updates and discussions of one or two ground-breaking and important primary papers. The authors of the primary papers have the opportunity of replying to the Comments, thus extending the dialogue on their work. **Monitor** pieces summarize, in 100–200 words, recently published papers. **Meeting reports** convey the excitement of recent conferences or seminars, focussing on the latest developments and the discussions surrounding them. **Opinion** articles are reviews of subjects written from a personal slant, and therefore highlight some of the controversies in cognition. **Books etcetera** features stimulating essay-reviews of recent publications, whether books, software, CD-ROMs, films, etcetera. And a list of books received for review will be published regularly. **Letters** stimulated by any of the articles published in *TICS* are welcome. Authors will be offered the chance to reply to any points raised.