Modules and quasi-modules: language and theory of mind in a polyglot savant^{*}

NEIL SMITH & IANTHI-MARIA TSIMPLI

Abstract

In this paper we provide further evidence from the polyglot *savant*, Christopher, for the nature of Theory of Mind. In particular, we exploit a distinction between modules as classically defined (Fodor, 1983), and 'quasi-modules'. While the latter have the domain-specificity of modules, they are not informationally encapsulated and they exploit a non-perceptual vocabulary. We report the results of his performance on a variety of false belief tasks, showing that differences inherent in the various tasks allow for an explanation of his apparently inconsistent behaviour, which in turn provides evidence for the 'quasi-modularity' of his Theory of Mind.

In earlier work (see especially Smith & Tsimpli 1995, and references therein), we have given a detailed account of the skewed abilities of the polyglot savant, Christopher. He shows a marked contrast in his performance on verbal and non-verbal tasks, and within the linguistic domain he manifests a number of further asymmetries. His knowledge of his first language, English, is essentially normal, while his numerous 'second' languages show that he has a startling talent for lexical and morphological acquisition, but relatively reduced ability in the development of syntax. Moreover, even in English, there is a mismatch between the language faculty proper, that is, the computational system of language in the sense of Chomsky (1995), and those central cognitive processes which are mediated by or presuppose language. These include pragmatic processes of interpretation, such as disambiguation, reference assignment and enrichment, and the

^{*}Ianthi-Maria Tsimpli is at the University of Cambridge. Parts of the material in this paper have been presented at the SLE conference at the University of Leiden, the Topika workshop at the University of Thessaloniki, and the Linguistics Circle at Edinburgh. We are grateful to the audiences at these venues and to Peter Carruthers, Annabel Cormack, Kirsten Malmkjaer, Annette Karmiloff-Smith and John Williams for comments and criticisms. They are not to be taken to agree with anything we write.

operation of a putative Theory of Mind Module.

In addition to the asymmetric profile described above, Christopher displays some, but not all, of the characteristics of autism. Like many autists he has one obsessive preoccupation; uniquely, however, his obsession revolves around language and languages, whereas most autistic people are linguistically challenged or even mute. Like most autists, he is socially inept, tends to avoid eye-contact, and gives little external evidence of his emotional state; but when it is possible for him to demonstrate his linguistic prowess, he can become animated and (relatively) forthcoming. Most importantly, he fails some 'Theory of Mind' tasks though, unlike the typical autistic subject, his performance is inconsistent. The atypical pattern of his behaviour in ascriptions of false belief is the major focus of what follows.

Some of our earlier work involved Christopher in remembering and/or inferring the identity of different symbols on flash cards. The results suggested that he "has no difficulty in adopting a perspective different from his own; in recalling past, currently invalid, states of affairs; and in projecting hypothetical states of affairs" (Smith & Tsimpli, 1995:178). Moreover, like most autistic subjects (see for example Nichols *et al.*, 1996), he performed correctly in a version of the Zaitchik (1990) photograph task. In this test, a photograph was taken of a doll wearing a pink dress. While we waited for the photograph to develop, the doll's dress was changed for a blue one, and Christopher had to 'predict' the colour of the dress in the photograph. His response was confident and appropriate. Whereas tasks which require no ascription of belief are systematically within his capabilities, Christopher finds tasks which necessitate the imputation of false belief to others difficult or impossible. Here, however, his performance was interestingly inconsistent: he systematically failed the 'Sally-Anne' test but passed the 'Smarties' test. If autistic behaviour is partially a function of an impaired Theory of Mind, these results are problematic.

There appears to be a systematic relation between linguistic ability and the development of a Theory of Mind, and it has been suggested (see e.g. Frith & Happé, 1994; Happé, 1995) that success in Theory of Mind tasks is causally dependent on Verbal Mental Age (VMA). As indicated diagrammatically in (1), Happé predicts that autistic individuals whose VMA is below a certain point, Y, will fail Theory of Mind tasks, whereas autistic individuals whose VMA is at or above a point, X, will pass them. Predictions are suspended for those whose VMA falls between the two points. In other words, a VMA \ge X is **sufficient** to make passing the task possible and entails having a Theory of Mind, and a VMA \ge Y is **necessary** for having a Theory of Mind. (1)

This correlation seems to fit nicely both with the evidence from various groups of autistic individuals who differ in their performance on Theory of Mind tasks, and with the evidence from normals, where the age-boundary after which children are assumed to have a Theory of Mind matches the appropriate VMA point X. However, in Smith & Tsimpli (1995) Christopher's performance on these tasks was presented as a problem for Happé's correlation, as his VMA is within normal limits, but his performance on the Sally-Anne test was consistently unsuccessful. We therefore suggested that, although the correlation may be descriptively correct for autistic and normal behaviour, Christopher's exceptional performance indicates that it may not be a causal one. More precisely, we argued that Theory of Mind representations crucially involve language only in that language is the necessary medium for expressing them, and that it should therefore be possible to find cases where Theory of Mind and language ability dissociate. We then claimed that Christopher illustrated one such possibility, namely where VMA is normal (i.e. above X) but the Theory of Mind is deficient.¹

It is necessary to clarify what is meant by 'language' here. Determining VMA crucially involves measuring the mastery and use of **vocabulary**, rather than the **syntactic** knowledge which is characteristic of the language faculty in Chomsky's sense. We have argued previously (Tsimpli, 1992; Smith & Tsimpli, 1995:169) that the lexicon is not a single cognitive domain, but that one part (the mental lexicon) falls within the central system, whereas the linguistic lexicon falls within the language faculty, with the two meeting at the morphological level. If this is correct, it is not surprising that knowledge and use of vocabulary should correlate with Theory of Mind performance as both are functions of 'central' processes. In Christopher's case the link between the morphology and the mental lexicon is intact, so that retrieval of either type of information is unproblematic, and hence the Theory of Mind Module is helped.

Christopher's divergent performance in Smarties and Sally-Anne was unexpected given the symmetric behaviour of autists and normals. It would not be unexpected, however, if there were a major difference in the processes involved in the formation of the

¹The other possibility, where Theory of Mind representations are available but language ability falls below Y has recently been tested with Theory of Mind tasks that do not involve language, but rather visual cues available to the subject. (See Clements & Perner, 1994).

representations in the two tasks. Whereas performance in the Smarties task includes accessing the relevant encyclopaedic entry, the Sally-Anne task is context-bound, with each piece of the information computed being new at the time of the task. It follows that VMA is more likely to be directly relevant to Smarties than it is to Sally-Anne.

To evaluate this suggestion, we need to look more closely at the properties of the representations involved in the different tasks. Standard Theory of Mind tasks characteristically involve the subject in entertaining a thought such as "I think that X believes that P", where X is a person other than the subject, and P is some proposition. That is, such tasks crucially involve a second-order representation with distinct experiencer subjects in the two (or more) clauses, and a modal predicate introducing the embedded proposition, as in (2), from Smith & Tsimpli, 1995:185):

(2) [*m* thinks [*x* MODAL PREDICATE [that P]] where *m* is the subject, and *x* is some other person.

If any of these properties is changed, the representation no longer encodes a standard Theory of Mind task. Suppose that instead of a distinct pair of subjects, *m* and *x* are identical: that is, the subject thinks about his/her own thoughts and attitudes with regard to a particular proposition P. According to the format in (2), this would not qualify as a Theory of Mind representation, unless *m* believed, counterfactually, that *m* and *x* were distinct entities. In other words, Theory of Mind tasks necessarily pertain to *other* minds, not just one's own mind. Likewise, in a situation in which P is embedded under a verb of *saying* instead of a modal predicate, the resulting representation would not qualify as a Theory of Mind one.

Consider now the difference between Smarties and Sally-Anne. Assuming that the information that Smarties tubes contain Smarties is common (shared) knowledge, the subject can respond to the question: "what would x think is in the Smarties tube?" by replacing x with m and thereby arrive at the appropriate answer. In other words, one only needs to project shared encyclopaedic information and assume that the question refers to common beliefs about the contents of Smarties tubes, by-passing the formation of a second-order representation.² Such shared knowledge is not involved in the Sally-Anne

²This suggestion is reminiscent of a distinction drawn by de Roeck & Nuyts (1994). They describe the 'evidentiality' or 'evidential qualification' characteristic of epistemic expressions as being dependent on such factors as the source on which the belief is based (perceptual, hearsay, inferencing) and the degree of 'intersubjectivity', i.e. whether the belief is shared or expresses only the speaker's judgment.

case, so there is no comparable way of by-passing the formation of a second-order representation.³

The implication of this result is that the Smarties task is not necessarily a Theory of Mind test, as it can be successfully passed by someone with a Theory of Mind deficit. This predicts that high-functioning autists ought systematically to pass the Smarties test, even if they fail Sally-Anne, and it raises interesting questions about the nature of the Theory of Mind Module in general. If modular processes of the standard visual and linguistic type are encapsulated, any deficit associated with such a module should cause impaired performance in the relevant domain.

There is considerable evidence that, despite residing within the central system, Theory of Mind can be distinguished from other aspects of cognition (see Leslie, 1987; Frith, Morton & Leslie, 1991; Carruthers & Smith, 1996, among many others). Particularly striking is the dissociation between mental retardation and Theory of Mind as exhibited by Down's Syndrome and autistic children. If subjects from each group are matched for mental age, only the autistic group shows a theory-of-mind deficit (see Nichols *et al.*, 1996), suggesting that Theory of Mind representations are the function of a domain-specific cognitive component within the central processor. Domain-specificity does not entail modularity (although the entailment holds in the opposite direction), so whether Theory of Mind is a Fodorian module or not is an open question.

In Smith & Tsimpli (1995) we argued, *contra* Anderson (1992), that Theory of Mind is quasi-modular⁴ in its operations, but that it does not fulfill the necessary requirements to be a Fodorian module, as its operations are typically inferential.⁵ However, Theory of Mind processes are clearly domain-specific in that only a subset of second-order representations are relevant to them. This and the double-dissociation evidence mentioned above suggests that Theory of Mind is a domain-specific cognitive system which fails to be modular both because of the vocabulary in which its operations are

³We shall modify this overly simplistic characterisation below, when we turn to variations on the Sally-Anne test.

⁴We in fact used the term 'central module' (Smith & Tsimpli, 1995:188).

⁵The question of what vocabulary the various components of the mind exploit is not straightforward. Fodor's modules, other than language, use an exclusively perceptual proprietary vocabulary. However, as language is not only an input system but also an output system, characterisable as in Smith & Tsimpli (1995:165ff.), the language faculty, like the Theory of Mind Module, needs to analyse both perceptual and central cognitive representations.

couched and because of its apparent 'penetrability' by the central processor. The issue of penetrability is in fact somewhat opaque. The fact that Theory of Mind accesses a central data-base does not entail that its internal operations are transparent to central processes, hence does not necessarily impugn its encapsulation. We think there are two kinds of evidence that might nonetheless suggest the possibility of penetrability: the fact that Theory of Mind operations are partly accessible to consciousness, and the possibility, to which we turn below, that they may indeed vary on the basis of central information.

In the light of these considerations we tested Christopher on a number of variations of Sally-Anne and Smarties. First, in view of our explanation for the mismatch between his performance on Smarties and Sally-Anne, we conducted a Smarties-type test in which the help of encyclopaedic information was factored out. Second, in view of our characterisation of Theory of Mind representations in (2), we tested to see whether Christopher could recall and report on his earlier beliefs when these contradicted his current ones; whether he could systematically control the difference between the use of *say* and *think* in the test questions; and finally, whether there was a difference in his performance depending on the modality in which the material was presented.

In the standard Smarties test Christopher's performance had been normal, because, we argued, he was able to impute common encyclopaedic knowledge to others in giving his answers. Accordingly, we replaced the Smarties tube with a series of alternative containers: keys in a glasses case, an insect in a wooden box, an orange in an opaque white box, etc. Whereas his earlier responses had been consistently successful he now seemed to fluctuate between appropriate and inappropriate answers. We give typical examples in (3):⁶

(3)

- (a) NS Chris, see what's in here [Smarties tube]? We've got a pencil sharpenerC Yes
 - NS If we showed it to John like this [closed], what do you think he'd say was in it?
 - C Smarties
 - NS Why would he say that?
 - C [very fast] Because it's a Smarties box.

⁶These are versions of the 'tangerine test' mentioned in Smith & Tsimpli (1995:187).

- (b) NS Chris, see what I've put in here [opaque, white plastic box] you know what there are in there? Films.
 - C Films.
 - NS If we showed this white container to John [closed]
 - C [Interrupting] Films
 - NS What would he think was in here?
 - C Films [confirmatory, low fall intonation]
 - NS Why would he think that?
 - C Because there are films in it.
- (c) NS Now, this is a slide-container, but it's got that [a paper-clip] in ...
 - C [very fast] Paper-clip
 - NS A paper-clip, that's right. So, if you go and ask Clare what's in there [with the box closed] what will she say?
 - C A paper-clip
 - NS She wouldn't think there was a slide in it?
 - C No
 - NS No? Why not?
 - C Because, um, she would think there was a paper-clip
 - NS OK
- (d) NS I've got another [flat, rectangular, wooden] box; what do you think is in that box? You mustn't look underneath, because then you can see.
 - C Umm, dhen ksero $\{= I \text{ don't know}\}$
 - NS [showing that it has a dead, mounted scorpion in] It's got a scorpion in
 - C Yes. Scorpion of Tunisia {The box had 'Scorpion de Tunisie' written in French inside}
 - ... Summasina
 - NS Supposing now we take this and we ask John what's in it, what will he say?C A scorpion
 - NS YOU didn't know there was a scorpion, did you?
 - C No
 - NS So how will he know?
 - C [pause] Because it says 'scorpion' on the other side
 - NS If we show him just this side, then what will he say?
 - C Nothing.

There were three relevant dimensions of variation. First, Christopher was consistently able accurately to recall and repeat his own initial suggestions, whether correct or incorrect. Second, in the absence of clear-cut encyclopaedic clues, he was inconsistent in reporting the potential responses of a third person, answering apparently at random. Third, questions with *say* and *think* received identical answers when they were about Christopher himself, they received apparently random answers when they were about a different referent.

These results corroborate our idea that encyclopaedic knowledge in the standard Smarties test facilitated Christopher's performance. That is, an individual with a Theory of Mind deficit and high VMA can pass the original test because there is no need in this modified condition to utilise Theory of Mind. His appropriate responses about his previous beliefs is in conformity with Naito *et al's* (1994) results for autistic children, and show that he **can** project a representation which conflicts with the present one as long as it is attributed to himself. In other words, Christopher does not attribute his previous false beliefs to a different 'self'.

That Christopher's responses to the difference between say and think questions were consistent when the referent was himself shows that he could rely on his introspective abilities in deciding what he thought and said. His inconsistent responses when a referent other than himself was involved suggest an impaired ability to distinguish verbs of believing and saying, when these are predicated of others. We suspect that Christopher does not fully distinguish the semantics of the two types of predicates. The basic semantic difference is presumably one between verbalised and non-verbalised thoughts. With normals it is common to find a mismatch in the content of the proposition embedded under such verbs. Autistic subjects and Christopher by contrast typically have no ability to lie, so it is not surprising that it is difficult for them, and him, to conceptualize the distinction appropriately. Questions referring to his earlier thoughts, and statements which are later proved to have been false do not cause him any difficulty as there is a direct correlation between what he thought and his verbalisation of it. With a different referent, however, the possibility of there being a discrepancy between someone's thoughts and their verbalisations of them reflects, we think, a ToMM problem. There are basically two options: either the conceptual entries for say and think are not complete, or they are fully specified but the ToMM cannot exploit the difference because of the deficit in representation (2). Specifically, because m is not equal to x in (2), a modal predicate in the embedded clause makes the whole representation a ToMM one, hence the deviant performance. That is, the task is within Christopher's linguistic ability but beyond his ToMM. It is noteworthy that, even when he has not explicitly verbalised his original

thought, he is still able to recall what he at first thought. This is significant as it makes implausible the alternative explanation that he was simply remembering what he said with no real understanding of the notion belief at all.

The Sally-Anne variants in (4) - (6) involved three changes. First, in view of Christopher's obsession with language and his superior performance on verbal as opposed to performance IQ tests, we devised variants in which the scenario was read as a story instead of being acted out. The tentative hypothesis to be tested was that such 'verbalisation' would exploit his linguistic abilities to the maximum, and minimise the use of abilities in which he is putatively defective. Second, in view of the facilitating effect of encyclopaedic knowledge on the Smarties test, the stories were controlled for the familiarity of the characters depicted. The hypothesis to be tested here was that familiarity might reduce the cognitive complexity of the task and simultaneously render it easier for Christopher to identify with what was going on. Third, the content of the proposition for which the false-belief was supposed to be held alternated between the temporal and the spatial. Here the idea, prompted by Christopher's established difficulty with spatial orientation and hand-eye coordination, was that temporal variation, especially when linguistically encoded, might be less difficult for Christopher to manipulate than visuo-spatial variation.

The first of these variables is illustrated by the contrast between (4a) and (4b), with the reactions indicated:

(4)

(a) Three children, Alexia, Maureen and Jill are playing in the kitchen. While the others are watching, Alexia puts a chocolate under a teacup. Maureen then goes out of the room. While she is away, Jill removes the chocolate from under the cup and puts it in a saucepan. When Maureen comes back into the room, where will she look for the chocolate?

C In the saucepan.

- NS Why will she look in the saucepan?
- C Because Jill put it in the saucepan.

- (b) John, Clare and Chris were tidying up the house together. They agreed to put the stamps in the top drawer of the desk. When Clare had gone out, John and Chris changed their mind and put the stamps on the bookcase. Where do you think Clare will look for the stamps when she comes back?
 - C In the top drawer of the desk
 - IT Why do you think she will look there?
 - C Because they were in the top drawer of the desk at first.

Christopher's performance on earlier tests of Sally-Anne had been consistently poor, hence typically autistic. When tested with written, rather than acted out, variants of the test, his rate of success appeared to be at chance, as he failed half and passed half of the versions on which he was tested. In fact, even 'chance' performance is crucially different from the totally consistent 'failure' of the early versions, and on closer examination it becomes clear that his performance was not really random.

Examples like (4b) occasioned Christopher no difficulty at all; indeed, (4b) includes a justification of his correct answer (in terms of temporal priority) which clearly shows that his response was not random. It is also significant that he was successful on the example in which the three people whose names are mentioned (Christopher, John and Clare) are all familiar to him: they are the people he interacts with daily. Examples like (4a) never contained reference to more than one person whom Christopher knew, and even then the degree of familiarity was considerably less than in (4b), and the example was essentially an abstract exercise.

A possible reason for the apparent facilitatory effect of the choice of participant can be derived from the notion of mental models in the sense of Johnson Laird (1983). The success of any inferential process depends on the number and complexity of the premises involved, the mode of presentation of those premises, and the familiarity of the protagonists. Constructing a mental model is facilitated if the persons involved in the story are already present in the subject's memory, so that less computational space is required for retrieval and storage. In normals, a contrast in familiarity would affect inferencing only if the number and/or complexity of the premises involved were considerably increased.

Christopher's performance also improved in variants of the Sally-Anne test like those in (5), where the relevant proposition involved temporal rather than spatial variation.

- (5) John, Tony and Paddy agreed to meet in the House of Commons on Wednesday. Later, John and Tony changed the meeting till Thursday, but neither of them remembered to tell Paddy. Which day did Paddy arrive for the meeting?
 - C On Wednesday, but John and Tony didn't.

Christopher unhesitatingly provided the right answer, explicitly contrasting the third person's false belief with the correct one. Why should there be this discrepancy between the temporal and the spatial? There are at least two ways in which spatial and temporal representations differ: first, space is three-dimensional whereas time is uni-dimensional; second, temporal notions can only be conveyed **linguistically**, whereas spatial ones need not be. The combination of these two properties makes temporal representations easier to construct, as the linguistic form in which they are expressed is putatively isomorphic to the representation in the language of thought which is subjected to further computations. That is, we assume that this isomorphism makes unnecessary the projection and enrichment of a representation of the kind that the construction of a spatial mental model requires.

It is also relevant that in the 'temporal' case, there is no salient perceptual evidence contradicting the validity of the representation of the previous state of affairs as there is in the standard Sally-Anne test. The absence of such perceptual evidence leaves the previous representation 'unmasked': two linguistic forms can survive provided they do not contradict each other at the formal semantic level and are not too complex (e.g. in terms of dimensionality). The role of perceptual salience may also contribute to Christopher's success in written versions of the task: where there is only a written input (his favourite medium), there is no distracting perceptual element and the participants, when unfamiliar, may even be depersonalised, being simply ascribed properties independent of Theory of Mind.

At this point it is appropriate to spell out in a little more detail what we take to be the defining characteristics of both modules and quasi-modules, and to confront the modularity position with a modularisation position of the kind associated with Karmiloff-Smith (1992), and a pan-modular approach of the sort defended by Sperber (1994). We start from Fodor's defining characteristics for input systems, and juxtapose their instantiation in vision and language before comparing them with Theory of Mind.

Fodorian modules have the array of properties in (6):

(6) Domain-specific

 Fast
 Mandatory
 Informationally encapsulated
 Cognitively impenetrable
 Subserved by specific neural architecture
 Subject to idiosyncratic pathological breakdown
 Ontogenetically deterministic

Fodor restricts his criteria to input systems, comprising the sensorium and language.⁷ This has as corollary that the proprietary vocabulary of any module is perceptual or at least perceptually based. For instance, putative modules, such as face recognition, are dependent for their input on a transduced visual input. It is clear, however, that domain-specificity is not restricted to input systems, and the most obvious contrast between language and vision on the one hand, and Theory of Mind on the other is that the proprietary vocabulary of the latter is conceptual rather than perceptual.⁸ This immediately leads to the question whether Theory of Mind representations are a natural class, exclusive of other kinds of representation, or if they are simply an arbitrarily delimited subset of a more general class of (second-order) metarepresentations.

There are clear parallels between the second-order representations exploited in Theory of Mind processes and the second-order representations characteristic of metalinguistic tasks involving Sperber & Wilson's (1986:228-229) notion of 'interpretive use' (see Smith & Tsimpli, 1995:74). Christopher and autistic subjects have systematic difficulties with both, but poor performance in meta-linguistic tasks cannot be due exclusively to problems with the Theory of Mind.

In normal cognitive development children are successful at Theory of Mind tasks at the age of 3.6 - 4 years (see Wellman, 1993), but understanding irony, metaphor, and jokes is beyond them. It would appear then that the development of a ToMM may be a

⁷We accept Chomsky's point (1986:14) that language is unlike Fodor's other modules in being an output system as well as an input system, but we still consider the comparison to be fruitful (see footnote 5 above). It may be that even such archetypal modules as vision have output properties, depending on one's view of imagery.

⁸This observation can be maintained even if it is correct that "thinking is ... conducted in the medium of natural-language sentences" (Carruthers, 1996:1; cf. Smith, 1983), as the contrast would simply need to be recast in terms of the need for transduction.

necessary, but cannot be a sufficient, requirement for the ability to interpret metalinguistic phenomena.

We need to refer to at least three stages: subjects (e.g. very young children) who have not yet developed a Theory of Mind, but are clearly not autistic (cf. Boucher, 1996:237); subjects who have developed a (partial) Theory of Mind but are still unable to handle (some aspects of) interpretive use; subjects who can handle both. Progress through these stages involves the development of some metarepresentational ability, but this cannot be the whole story and needs to be supplemented. We think that two necessary components of an appropriate account can be found in Relevance Theory: the distinction between the Cognitive and the Communicative principles of relevance (Sperber & Wilson, 1995:260-1), and the notion of 'dissociation'. The cognitive principle of relevance accounts for the maximisation of the relevance of incoming stimuli, and the communicative principle optimises the relevance of ostensive communication in terms of a trade-off between the number of contextual effects acquired and the amount of processing effort required to gain them. Crucially, normal communication involves the intuiting of the intentions of one's interlocutor, presupposing that such intentions can be identified. In severe autism it may be the case that the communicative principle of relevance is entirely lacking. In less severe cases, such as that of Christopher, this cannot be the case, as he responds to questions appropriately and without delay, albeit laconically, and some forms of standard inferencing are intact. It follows that he has some notion of intention, but what appears to be missing is the ability to understand the speaker's dissociation from the proposition expressed. Crucially, his difficulty in such cases arises from the fact that the intended interpretation relies first on identifying premises embedded under another person's beliefs and desires, and second, realising that the speaker is not committed to them. In contrast, such attribution and dissociation are not required for the inferencing that does lie within his abilities: simple modus ponens, the use of implicated assumptions and conclusions, and so on. (Smith & Tsimpli, 1995:69-71).

The most important criteria for modularity are generally taken to be informational encapsulation and cognitive penetrability. Some people equate these two notions, but we prefer to keep them distinct. We use informational encapsulation to refer to the situation in which the workings of a module are insulated from any information, be it from the central system or other modules. We use cognitive impenetrability to refer to the insulation of the workings of some module from central processes only. That is, some system could be cognitively impenetrable but not informationally encapsulated, whilst the reverse relation is impossible. Even if the two notions contingently always cluster, we think it important to keep them distinct.

In this regard, we think, as indicated above, that there are significant differences between language and ToMM. Language is informationally encapsulated and *a fortiori* cognitively impenetrable, ToMM is probably cognitively penetrable and, depending on the precise relations between VMA and metarepresentation, perhaps informationally unencapsulated. In fact, this last possibility seems implausible if, as we suggested above, the relevance of VMA is precisely that it taps central rather than modular aspects of the language faculty.

The conclusion we are led to is that domain-specificity characterises both language and the ToMM. The major difference between them lies in the notion of informational encapsulation and/or cognitive penetrability. It is reasonably clear that the computational system of language is cognitively impenetrable: what would the central system be able to do with syntactic derivations of the Minmalist type anyway? However, assuming language to be an output system as well as an input system, neither of these notions fits the description of the language faculty appropriately. Accordingly, we reinterpret the Fodorian notions of informational encapsulation and cognitive impenetrability as epiphenomena that can be derived from either of the properties in (7) (or from a conspiracy between them, should it be the case that they always correlate):

- (7) (a) The possession of a distinct neural architecture
 - (b) The exploitation of a uniquely specified vocabulary

We believe (7a) to be straightforwardly characteristic of language but not of the ToMM, whereas the position with regard to (7b) is slightly more complicated. If (7b) refers to a vocabulary which can be employed and computed only by a particular cognitive system, it is indeed a property of the **computational system** of language: i.e. the derivational processes of the syntactic component, but not of the interface levels, nor of the interface lexicon. It is **not** a property of the ToMM. (7a) and (7b) are descriptions at two different levels: the neurological and the psychological. For a nativist who believes in modularity, (7a) would strongly suggest, perhaps entail, (7b). On Karmiloff-Smith's (1992) 'modularisation' view, on which there are at most skeletal representations with predispositions to attend to specific stimuli, (7a) would not entail (7b). Such representational predispositions do not dispose of an independent vocabulary that is opaque to other systems, and this view suggests strongly (and we believe incorrectly) that language could not differ from ToMM in terms of informational encapsulation.

Let us return to the other characteristics of (Fodorian) modules given in (6). There is no obvious distinction between modules and quasi-modules in terms of their speed of operation. Although the complexities of some versions of Sally-Anne, for instance, are such that they necessitate considerable time and effort to work out, this is no different from the complexity of some linguistic constructions. Typical examples are provided not only by garden-path sentences but also by such relatively straightforward sentences as Chomsky's (1986:9) "John is too clever to expect us to catch".

The mandatoriness of mindreading vis-à-vis that of sentence processing is similarly clear. Except in cases of pathology, we are no more capable of **not** attributing notions of agency and intention to automotive entities, animate or inanimate, than we are of failing to process sentences of our native language as such.

It is worth pointing out that a ToMM deficit also gives rise to fast and mandatory *wrong* answers. For instance, after giving a rapid (wrong) response to a question in one of the Sally-Anne tests, Christopher was questioned repeatedly until he came up with the right answer. Although his final response was correct, it is significant that he arrived at it neither fast nor mandatorily.

If autism is correctly characterised as necessarily involving a ToMM, then this condition constitutes an example of idiosyncratic pathological breakdown comparable to that of various language disorders. There are, however, two problems: first, as mentioned above, although the putative module may be involved in autism, it cannot be the sole basis for it, because of the obvious fact that children are diagnosable as autistic before the module develops. Second, there is a striking asymmetry between language and ToMM in that acquired (aphasic) disorders of obvious neurological aetiology are just as common as developmental ones in the case of language, but there is notoriously no traumatically caused example of 'acquired autism' (see Boucher, 1996:239).⁹

It is mainly, though not exclusively, the existence of idiosyncratic breakdown that motivates the claim that language is subserved by specific neural architecture. In the absence of traumatic autism, such motivation for a ToMM is largely excluded. Nonetheless, there is considerable evidence that autism is associated with specific neural deficits (see e.g. Courchesne, 1992), so the parallelism is apparently again close. It is, however, significant that it is precisely in the case of semantic loss and in degradation of

⁹It is worth investigating the possibility that aphasia is typically not so much a problem with the language faculty as such, as with language processing. If it turns out that acquired aphasia can be adequately analysed as a processing problem, where there is a breakdown in the ability to access available linguistic categories, then there is no acquired deficit affecting the language faculty any more than there is acquired autism. The issue has not been resolved, and it is plausible that genetic or developmental disorders constitute clearer evidence for modularity or domain-specificity than do acquired ones. In this respect Language and ToMM pattern together.

central functions such as Theory of Mind that neural localisation is least clearly in evidence (see Shallice, 1988, ch.11).

The deterministic nature of ontological development in the two domains is similarly comparable. In both cases there is a regular, maturationally determined unfolding of the respective abilities, apparently independent of cultural idiosyncrasies.¹⁰

We turn finally to the relation between our notion of modularity, the modularisation of Karmiloff-Smith (1992) and Sperber's (1994) pan-modular epidemiology. We have argued elsewhere (Smith, 1994) that Karmiloff-Smith's elegant discussion of modularisation is fatally flawed when it comes to language. The early, deterministic, and unconscious development of syntactic knowledge, together with the informational encapsulation we have argued for above, combine to suggest that the language faculty involves a module in the Fodorian and Chomskyan senses. We would emphasize, however, that our picture of language crucially involves aspects of the central system as well, so the bald alternative of 'modular/non-modular' is simplistic, indeed false.

When it comes to the development of the ToMM, we find Karmiloff-Smith's ideas more congenial. We believe that the case of Christopher provides evidence for the modularity (and domain-specificity) of language, but only for the domain-specificity of the ToMM. For instance, some second-order representations, such as simple embedding under verbs of saying, do not need recourse to the central system, as their property of recursion can be accommodated entirely within the language faculty. It is only when we come to examples like metaphor, metalinguistic negation and jokes, which violate linguistic semantics (specifically, in terms of giving rise to a contradiction) and which require the exploitation of interpretive use and an attitude of dissociation, that the ToMM is essential. For these representations it is entirely plausible that a modularisation account is adequate.

If the ToMM is not modular, it is unsurprising that we disagree with Sperber's (1994) 'everything is a module' approach. In ongoing work Sperber has suggested that modules are ubiquitous: not only does he accept Fodorian modularity for input systems, but claims, *pace* Fodor, that the central system is pervasively modular too. This central modularity has two facets: on the one hand, every concept is a micro-module, so there is a module devoted to 'goldfish', for instance; on the other hand the whole of metarepresentation constitutes a single module. We have trouble with both these claims. Our unease is partly a function of our suspicion that Sperber's suggestions are merely terminological, but we also think that there are serious problems with both claims. For the micromodules the

¹⁰We find the view that "the theory of mind, perhaps even mind itself, is a cultural invention" (cited in Astington, 1996:188) implausible.

problem is that they have none of the properties of (Fodorian) modules; for the metarepresentational module, it is that there is no place for inferential integration, and the 'module' reduces to a terminological replacement for Fodor's 'central system'.

We are sympathetic to modifications of Fodor's proposals: this paper is an example. Our position, however, is that while the basic insight about input systems is correct, it is necessary both to refine the claims about language and to temper the pessimism about the central system. Language is only partially modular. It also belongs in the central system. This is not just vague anarchic agnosticism, we have made explicit suggestions about which parts of language belong in which domain. Further, the central system is not an unstructured, unknowable mess, but itself has structure of a quasi-modular kind. This too goes - if only slightly - beyond hand-waving, in that we have suggested some of the properties which inhere in such quasi-modules.

References

Anderson, M. (1992) Intelligence and Development: A Cognitive Theory. Blackwell.

- Astington, J. (1996) "What is theoretical about the child's theory of mind?: a Vygotskian view of its development". In P. Carruthers & P.K. Smith (eds.) *Theories of Theories of Mind*. CUP, pp. 184-199.
- Baron-Cohen, S., H. Tager-Flusberg & D.J. Cohen (eds) (1993) Understanding Other Minds: Perspectives from Autism. OUP.
- Boucher, J. (1996) "What could possibly explain autism?". In P. Carruthers & P.K. Smith (eds.) *Theories of Theories of Mind*. CUP, pp. 223-241.
- Carruthers, P. (1996) Language, Thought and Consciousness: An Essay in Philosophical Psychology. CUP.
- Carruthers, P. & P.K. Smith (eds.) (1996) Theories of Theories of Mind. CUP.
- Chomsky, N. (1986) Knowledge of Language: Its Nature, Origin and Use. New York, Praeger.
- Chomsky, N. (1995) The Minimalist Program. Cambridge, MA; MIT Press.
- Clements, W.A. & J. Perner (1994) "Implicit understanding of belief". Cognitive Development 9:377-395.
- Courchesne, E. (1992) "A neurophysiological view of autism". In E. Schopler & G. Mesibov (eds) *Neurobiological Issues in Autism*. Plenum.
- Fodor, J. (1983) The Modularity of Mind. MIT Press.
- Frith, U. & F. Happé (1994) "Language and communication in autistic disorders". *Philosophical Transactions of the Royal Society of London* B 346:97-104.
- Frith, U., J. Morton & A. Leslie (1991) "The cognitive basis of a biological disorder". *Trends in Neuroscience* 14:433-438.
- Happé, F. (1995) "The role of age and verbal ability in the Theory of Mind task performance of subjects with autism". *Child Development* 66:843-855.
- Johnson Laird, P. (1983) Mental Models. CUP.

- Karmiloff-Smith, A. (1992) *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press.
- Leslie, A. (1987) "Pretence and representation: the origins of 'Theory of Mind'". *Psychological Review* 94:412-426.
- Naito, M., S. Komatsu & T. Fuke (1994) "Normal and autistic children's understanding of their own and others' false belief: a study from Japan". *British Journal of Developmental Psychology* 12:403-416.
- Nichols, S., S Stich, A. Leslie & D. Klein (1996) "Varieties of off-line simulation". In P. Carruthers & P.K. Smith (eds.) *Theories of Theories of Mind*. CUP. pp.39-74.
- de Roeck, A. & J. Nuyts (1994) "Epistemic modal expressions by high-functioning autistic adults: A test case for the 'theory of mind' hypothesis". *Antwerp Papers in Linguistics* 82.
- Shallice, T. (1988) From Neuropsychology to Mental Structure. CUP.
- Smith, N. (1983) *Speculative Linguistics*. An inaugural lecture delivered at University College London. Published by the College.
- Smith, N. (1994) Review of Karmiloff-Smith (1992). *European Journal of Disorders of Communication* 29:95-105.
- Smith, N. & I.-M. Tsimpli (1995) The Mind of a Savant. Blackwell.
- Sperber, D. (1994) "The modularity of thought and the epidemiology of representations". In L. Hirschfeld & S. Gelman (eds) *Mapping the Mind: Domain Specificity in Cognition and Culture*. CUP, pp.39-67.
- Sperber, D. & D. Wilson (1986/1995²) *Relevance: Communication and Cognition*. Oxford, Blackwell.
- Tsimpli, I.-M. (1992) Functional Categories and Maturation: The Prefunctional Stage of Language Acquisition. PhD Thesis, University College London. {Forthcoming with Garland}.
- Wellman, H.M. (1993) "Early understanding of mind: the normal case. In Baron-Cohen et al., pp.10-39.
- Zaitchik, D. (1990) "When representation conflicts with reality: The preschooler's problem with false beliefs and 'false' photographs". *Cognition*, 35:41-68.