On Fodor's Problem

P. CARRUTHERS

Abstract: This paper sketches a solution to a problem which has been emphasized by Fodor. This is the problem of how to explain distinctively-human flexible cognition in modular terms. There are three aspects to the proposed account. First, it is suggested that natural language sentences might serve to integrate the outputs of a number of conceptual modules. Second, a creative sentence-generator, or supposer, is postulated. And third, it is argued that a set of principles of inference to the best explanation can be constructed from already-extant aspects of linguistic testimony and discourse interpretation. Most importantly, it is suggested that the resulting architecture should be implementable in ways that are computationally tractable.

1. Introduction

Let *Fodor's Problem* be the problem of how to build distinctively-human cognition out of modular components. We need to be able to solve this problem if human cognitive processes are to be computationally realized. For there is good reason to think that computations have to be modularly organized if they are to be tractable, as we shall see shortly. Notoriously, Fodor himself thinks that this problem cannot be solved, and hence believes that central cognition (as opposed to the peripheral input/output systems of perception, language processing, and motor-control) will have to remain mysterious for the foreseeable future (Fodor, 1983, 2000).

Fodor makes his own problem look rather more difficult than it really is, however, by construing human cognitive processes by analogy with science. For science is public; science is social; science is externally supported by books, written records, diagrams and so forth; and science only changes on a timescale of months, years, and sometimes centuries—none of which is true of individual cognition. Our thought processes are private, often operate without external support, and occur within a time-frame of seconds and minutes. Recognizing these differences doesn't make Fodor's Problem go away. But it does make it a bit more tractable. For there is no reason to think that individual unaided cognition can do what science can do (Carruthers, 2003).

Thanks to Richard Samuels for a long telephone conversation which prompted me to write this paper; and thanks to John Horty, Georges Rey, Richard Samuels, Dan Sperber, Stephen Stich, John Tooby, and two anonymous referees for comments and suggestions.

Address for correspondence: Department of Philosophy, University of Maryland, College Park, MD 20742, USA.
E-mail: pcarruth@umd.edu

Mind & Language, Vol. 18 No. 5 November 2003, pp. 502–523.

There are a number of features of human cognitive processes that are genuinely distinctive and problematic, however. As we shall see, they can be divided into three different categories: *flexibility* of content; *creativity* of content; and *abductive inferences* performed upon such contents.[1] Each of these can be addressed somewhat differently, I shall suggest. We can therefore hope to adopt a policy of 'divide and conquer'. Of course, actually *solving* Fodor's Problem in detail would require nothing less that a worked-out computational model of the human mind. And needless to say, such a solution lies well beyond the scope of this paper. My aim here is just to show that there is no reason of principle why the problem *shouldn't* admit of a solution. Put differently, I shall aim to sketch out the directions in which scientific enquiry should proceed if Fodor's Problem is ultimately to be solved. First, though, I need to provide some background on modular models of cognition.

2. Mental Modularity

A number of cognitive scientists and evolutionary psychologists have defended a thesis of *massive modularity* in recent decades, arguing that the human mind consists either entirely or largely of mental modules (Gallistel, 1990; Tooby and Cosmides, 1992; Sperber, 1996; Pinker, 1997). A variety of converging lines of argument are supposed to support some thesis of more-or-less massive modularity. There is the evidence of domain-specificity in development (Hirschfeld and Gelman, 1994), and the evidence of precocious development in some of these domains (Sperber *et al.*, 1996). There is the evidence of dissociable damage (either in development, or through brain damage) to many individual systems, which can leave all else intact (Shallice, 1988; Tager-Flusberg, 1999). And there is the argument from comparative psychology (and evolutionary biology more generally), that we should *expect* the mind to contain many distinct learning mechanisms, each of which was designed to solve some adaptive problem or other (Barkow *et al.*, 1992; Gallistel, 2000). Moreover (and most importantly), there is the methodological argument that only if minds are wholly or largely made up of modules will the operations of minds be computationally tractable (Fodor, 1983, 2000; Tooby and Cosmides, 1992). (I shall return to this last argument briefly in section 3 below.)

I shan't make any serious attempt to argue in support of a thesis of massive modularity here (although see section 3). But then nor do I need to. For Fodor's Problem can be cast in the form of a conditional question, thus:

> *If* we suppose that the mind has a massively modular organization, *then* can we explain the existence of distinctively-human creative and flexible cognition?

[1] Strictly speaking, we should also add a fourth category: our distinctively flexible *practical* reasoning. But I shall follow Fodor's example by focusing entirely on processes of belief-fixation in this paper; leaving any discussion of practical inference from a modularist perspective to another occasion. (See Carruthers, 2002d, forthcoming.)

Whether or not the antecedent of this conditional question is true isn't now at issue. A negative answer to our question would provide good reason for thinking that the antecedent is false, however, just as Fodor (2000) supposes. And conversely, a convincing positive response to our question would remove one of the main obstacles standing in the way of acceptance of massive modularity. For many of those who oppose such a thesis base their opposition on the alleged inability of modular approaches to explain distinctively-human cognitive processes.

I do need to say something more about what modules *are*, in this context, however. For many in philosophy—having read Fodor's classic 1983 book, perhaps, but not much more recently in the modularity genre—have ingrained false assumptions about the nature of modules. It is obvious that many of the systems involved in a thesis of massive modularity will be *conceptual* modules, designed to process conceptual inputs and to deliver conceptual outputs. Therefore the original Fodorian constraints that modules should have proprietary transducers and should deliver shallow (i.e. non-conceptual or simple-conceptual) outputs should plainly have no place. In fact, in order to see what notion of *module* is appropriate to a thesis of massive modularity, we need to look at the arguments in support of massive modularity. Roughly, a module is then whatever those arguments would warrant.

On reflection, it can be seen that not all of the arguments mentioned above really do support a thesis of *massive* modularity. For example, the argument from precocious development in some domains only suggests (at best) that the mind will contain *some* mental modules, not that it will be wholly composed of such systems. In fact, the only argument supporting massive modularity as such, derives from the claim that all mental processes need to be computationally tractable, and therefore realized in encapsulated modular mechanisms (see section 3 below). *Some* of these modules will be domain-specific in their input conditions, no doubt, and *some* will be evolved adaptations; *some* will be genetically channeled (or 'canalized') in development, and *some* might run on unique algorithms. But from the perspective of a claim of massive modularity, *all* modules will be isolable functional sub-components of the mind whose processes are significantly encapsulated. So this is how the notion of *modularity* should be understood in the present context.

A little more needs to be said about the notion of computational tractability which is at work here, however, and the kind of *encapsulation* which it supports, before we can get down to addressing Fodor's Problem itself.

3. Computational Tractability and Encapsulation

The claim that cognition is realized in *computational* processes of some sort is the guiding assumption lying behind all work in computational psychology, hence gaining a degree of inductive support from the successes of the computationalist research program. And many of us believe that this form of psychology represents

easily our best hope (perhaps our *only* hope) for understanding how mental processes can be realized in the physical processes of the brain (Rey, 1997).

Just about the only people who disagree, are those who endorse an extreme form of distributed connectionism, believing that the brain (or significant portions of it, dedicated to central processes) forms a vast collection of connectionist networks, in which there are no local representations. The successes of the distributed connectionist program have been limited, however, mostly being confined to various forms of pattern-recognition; and there are principled reasons for thinking that such models cannot explain the kinds of structured thinking and one-shot learning of which humans and other animals are manifestly capable (Fodor and Pylyshyn, 1988; Fodor and McLaughlin, 1990; Horgan and Tienson, 1996; Gallistel, 2000; Marcus, 2001). Indeed, even the alleged neurological plausibility of connectionist models is now pretty thoroughly undermined, as more and more discoveries are made concerning localist representation in the brain (e.g. Rizzolatti *et al.*, 2001).

The computational processes that realize human cognition will need to be *tractable* ones, of course; for they need to operate in real time with limited computational resources (Cherniak, 1986). And it is this that dictates that cognition should be realized in a system of encapsulated modules. You only have to *begin* thinking in engineering terms about how to realize cognitive processes in computational ones to see that the tasks will need to be broken down into separate problems which can be processed separately (and preferably in parallel). And any processor that had to access the full set of the agent's background beliefs (or even a significant sub-set thereof) would be faced with an unmanageable combinatorial explosion.[2] We should therefore expect the mind to consist of a set of processing systems which are isolable from one another, and which operate in isolation from most of the information that is available elsewhere in the mind.

Modularism is now routinely assumed by just about everyone working in artificial intelligence, in fact (Bryson, 2000; McDermott, 2001). So anyone wishing to *deny* the thesis of massive modularity is forced to assume a heavy burden (as of course Fodor, 2000, fully recognizes). It must be claimed, either that minds *aren't* computationally realized, or that we haven't the faintest idea how they can be. And either way, it becomes quite mysterious how minds can exist in a physical universe.

Now, *one* way in which one might try to make the operations of modular systems computationally tractable would be to place constraints on their input conditions. This would be to make them informationally encapsulated in the sense of *input*-encapsulated. But it is far from clear that this is the only or best way. Some form of *process* encapsulation would work better. (Some modules might be input-encapsulated as well as process-encapsulated, of course; but *all* must be the latter,

[2] As Fodor (2000) puts the point, the demand for modularized processing derives from the fact that tractable computations have to be *local* ones. Computational processes have to be sensitive to the syntactic structures of the representations that they govern; and any such process which attempted to handle more than just a few pre-defined kinds of representation at once would rapidly fall subject to combinatorial explosion.

I think, for reasons that will shortly become apparent.) However, it is important to distinguish between two forms of process-encapsulation, too, at this point.

One kind of process-encapsulation is a matter of the algorithms being run by the module being impervious to change. A process-*un*encapsulated system, in this sense, would be one whose algorithms can be altered through some manner of learning, or through changes in belief elsewhere in the mind. It does seem likely that many modules will possess this kind of process-encapsulation. For after all, in the case of those modules that have been selected for and are genetically channeled in development, at least, one might expect that their algorithms should be relatively fixed and unalterable. It is doubtful, however, whether the above sense of encapsulation does anything for the problem of computational tractability by itself. If we designed some system to run a fancy Bayesian algorithm, for example (albeit a fixed and unalterable one), which required it to access all of the subject's beliefs simultaneously, then this looks as if it would be the very opposite of computational tractability.[3]

The other form of process-encapsulation has to do with the *processing data-base* of a module (Sperber, 2002). A fully encapsulated module, in this sense, is one that can't draw on any acquired information (i.e. information that isn't built into the structure of its algorithms) in the course of processing its input. A partially encapsulated module would be one that can only draw on the contents of some domain-specific data-base, not the total set of the subject's beliefs. This sort of process-encapsulation is surely what is required to deal with the problem of computational tractability. And it looks as if it ought to be independent of input-encapsulation, too. One can imagine systems that can take any propositional content as input, but which can only draw on a very limited range of attitudes in processing that input (Carruthers, 2002d).

In conclusion of this section, then, the thesis of massive modularity seems best supported by the claim that cognitive processes must be tractably computationally realized. And what that thesis then says is that cognition is built largely or entirely out of isolable functional sub-components, whose internal processing is significantly encapsulated, operating in isolation from much of the information that is held elsewhere in the mind. And then Fodor's Problem is the problem of seeing how our minds could possibly have such an architecture, given what we already know about the mind's powers and capacities.

4. First Problem: Flexibility of Content

As I remarked at the outset, there are three distinct and distinctive features of human cognition which collectively constitute Fodor's Problem. The first is that

[3] Nor does it seem likely that this sort of process-encapsulation is even *necessary* to ensure computational tractability. A module could surely be process-encapsulated in the (other) sense that matters, even if its algorithms were alterable through some form of learning.

humans are capable of freely combining together concepts and propositions across modular boundaries. This is manifest to ordinary introspection. I can be thinking about thoughts one moment, horses the next, and then a landslide the next; and I can then wonder what led me to think about thoughts, horses and falling stones—thereby combining into a single thought concepts or propositions drawn from the domains of folk-psychology, folk-biology and folk-physics. How is this possible, unless there is some a-modular central arena in which modular contents can be received and recombined, further inferences drawn from the results, and so forth?

This problem might well be solved by postulating a role for natural language sentences in cognition, as I have argued at length elsewhere (Carruthers, 1998a, 2002c, 2002d). For, first, there is good reason to think that the natural language faculty is a module, with distinct input and output elements, together with a dedicated knowledge data-base. Second, there is good reason to think that this module would have been set up within the architecture of a modular mind in such a way as to take inputs from all of the various conceptual modules, so that their contents should be reportable in speech. And third, there is reason to think that the abstractness and re-combinatorial powers of natural language syntax would make it possible for the language faculty to combine together sentences encoding the outputs of different modules into a single natural language representation. If such sentences can then be displayed in auditory or motor imagination, and can adopt some of the causal roles distinctive of thought, then we shall have explained how thought can acquire some of its flexibility of content within a wholly modular cognitive architecture. Since I have pursued these points in some detail elsewhere, here just let me elaborate very briefly on each of them.

The modularity of the language faculty isn't *wholly* uncontroversial, of course; but in the present context it can be regarded as such. For language forms one of the archetypal input and output modules defended at length by Fodor (1983). Admittedly, the language faculty is probably unique in having both input *and* output functions, since it handles the main elements of both comprehension and production of speech. But there is some reason to think that these distinct functions are underpinned by distinct modular *components* of the language faculty, and that each is subserved by a common linguistic-knowledge data-base (Chomsky, 2000). At any rate, this is what I shall take for granted in what follows.

In the context of a modular model of mind, the language faculty would plainly need to receive inputs from the various central-process conceptual modules (for folk-psychology, folk-biology, and so on), in such a way that the outputs of those modules should be reportable in speech. Here classical models of speech-production can be assumed (Levelt, 1989). The job of the conceptual modules is the production of domain-specific beliefs or thoughts. So the communicative process begins with a thought-to-be-communicated, just as classical models require. And then lexical items, syntactic structures, and phonological properties are recruited and assembled to subserve the expression of that thought in speech.

The language faculty would also need to feed inputs *into* the various central modules, so that those modules can get to work on domain-specific contents

received through the testimony of other people. Here there is some reason to think that the language faculty adopts the general strategy of building a *mental model* (a quasi-perceptual analog representation) from the input sentence, which is then in the right format to be received by the conceptual modules. (The latter would already have been set up to process perceptual input, of course.) At any rate, there is a significant body of evidence pointing to the role of mental models in discourse comprehension (see Harris, 2000, for reviews). And this strategy would also have the virtue of enabling the language faculty to make the contents of *non*-domain-specific sentences available to all the various conceptual modules simultaneously, without first having to parse those sentences into their respective domain-specific elements.[4]

As for how distinct domain-specific sentences might be combined into a single domain-general one on the output side, two points are suggestive. One is that natural language syntax allows for multiple embedding of adjectives and phrases. Thus one can have, 'The toy is in the corner with the *long* wall on the left', 'The toy is in the corner with the *long straight* wall on the left', and so on. So there are already 'slots' into which additional adjectives—such as 'blue'—can be inserted. The second point is that the references of terms like 'the wall', 'the toy', and so on will need to be secured by some sort of indexing to the contents of current perception or recent memory. In which case it looks like it wouldn't be too complex a matter for the language production system to take two sentences sharing a number of references like this, and combine them into one sentence by inserting adjectives from one into open adjective-slots in the other. The language faculty just has to take the two sentences, 'The toy is in the corner with the long wall on the left' and, 'The toy is by the blue wall' and use them to generate the sentence, 'The toy is in the corner with the long *blue* wall on the left', or the sentence, 'The toy is in the corner with the long wall on the left *by the blue wall*'.[5]

As for how a natural language content-integrating sentence can acquire some of the causal roles distinctive of thought, our story can go like this. If the syntactic structure in question is used to generate a phonological representation of that sentence, in 'inner speech', then this might normally co-opt the resources of the input sub-system of the language faculty in such a way as to generate a 'heard' sentence in auditory imagination. By virtue of being 'heard', then, the sentence would also be taken as *input* by the conceptual modules which are down-stream of

---

[4] I believe that the contents of perception are already integrated, carrying information about a wide range of domains (Carruthers, 2000, ch.11). The problem of integrating contents across domains arises *only* at the level of output from the various conceptual modules, each of which processes and draws inferences from a different aspect of the perceptual input (Carruthers, 2002d).

[5] This example is purposefully chosen from the one case where there is robust experimental evidence of the role of language in integrating module-specific contents. See Hermer-Vazquez *et al*. (1999), who show that neither rats nor young children can integrate geometric with object-property information when disorientated in a small rectangular space, whereas human adults *can* integrate such information, but only when the resources of the language faculty are not tied up with other tasks (such as shadowing speech). See Carruthers (2002c, 2002d) for extensive discussion of the significance of these data.

the comprehension sub-system of the language faculty, receiving the latter's output. So the cycle would go: thoughts created by central modules are used to generate domain-specific natural language sentences, which are then combined to frame a content-integrating natural language representation; the latter is then used to generate a sentence in auditory imagination, which is then taken as input by the central modules once again. One can suppose that cycles of processing of this sort might sometimes issue in usefully-novel information.[6]

A comparison with visual imagination may be of some help here. According to Kosslyn (1994), visual imagination exploits the top-down neural pathways which are deployed in normal vision to direct visual search and to enhance object recognition, in order to generate visual stimuli in the occipital cortex. These are then processed by the visual system in the normal way, just as if they were visual percepts. A conceptual or other non-visual representation (of the letter 'A', as it might be) is projected back through the visual system in such a way as to generate activity in the occipital cortex, just as if a letter 'A' were being perceived. This activity is then processed by the visual system to yield a quasi-visual percept.

Something very similar to this presumably takes place in auditory (and other forms of) imagination. Back-projecting neural pathways which are normally exploited in the processing of heard speech will be recruited to generate a quasi-auditory input, yielding the phenomenon of 'inner speech'. In this way the *outputs* of the various conceptual modules, united into a natural language sentence by the production sub-system of the language faculty, can become *inputs* to those same modules by recruiting the resources of the comprehension sub-system of the language faculty, in inner speech. (And indeed, there is evidence that *both* the language production area *and* the language comprehension area of the cortex are active when subjects engage in inner speech. See Paulescu *et al.*, 1993; Shergill *et al.*, 2002.) And the distinctively-flexible, non-domain-specific, character of human thought processes would be the result.

5. Interlude: A Comparison with some Related Views

How do the ideas sketched here relate to other similar-sounding proposals which may be familiar to many readers, specifically the *global workspace* model of Baars (1988, 1997) and the *Joycean machine* of Dennett (1991)? Each of these other

[6] Does this mean that deaf people will be incapable of distinctively human flexible thinking? Of course not. Sign language is fully language; and the kinds of cycles of processing envisaged above can be subserved by visual or motor forms of imagination equally well. Does it mean that aphasic people will be incapable of thinking flexibly? In the sense of thinking in a domain-integrating way across central modules, yes. (The proposal remains to be tested.) But note that there are other *kinds* of flexibility which can be generated by the sequential use of a set of powerful conceptual modules. (On this, see Carruthers, 2002d.) So my prediction (born out by the data) is that there will be lots of ways in which aphasic people can still be pretty smart (Varley, 1998, 2002).

proposals is made in the context of an account of consciousness, of course, which isn't my concern here (but see Carruthers, 1998b). Yet each does also propose a significant role for natural language sentences in cognition more generally.

Baars (1997) thinks that for perceptual states to become conscious is for them to be broadcast to a wide array of other cognitive systems. And he thinks that inner speech, too, will often be conscious, piggy-backing on the same broadcasting capacities as auditory perception. So far there is agreement, *modulo* the commitment to massive modularity in my own proposal. For I, too, think that the outputs of the perceptual modules will be made widely available to a range of conceptual modules; and I, too, think that natural language sentences displayed in inner speech will be consumed by the comprehension sub-system of the language faculty and made available to the same range of conceptual modules. However, there is nothing in Baars' work to suggest that language will play the role of *integrating* the outputs of other conceptual systems, which is the main component of the proposal sketched in section 4 above.

In contrast to Baars, Dennett (1991) does propose a role for language in integrating the outputs of other systems (although this is not at the forefront of the theory). For he thinks that underlying the language-involving stream of consciousness which is the *Joycean machine* will be the operations of a whole host of implicit processing systems. These both compete with one another and form shifting alliances for control of the language production system, generating the stream of inner speech. However, two things mark off Dennett's theory from the proposals being sketched here. First, Dennett thinks that non-linguistic thought is unlikely to involve discrete, structured, content-bearing states. He therefore believes that language production cannot be classically conceived as the encoding of propositional thought, but must rather be understood on a 'pandemonium' model. Second, Dennett thinks that the mind becomes radically re-programmed by the Joycean machine, with learned associations, habits of thought, and 'memes' absorbed from the surrounding culture leading to a whole new *kind* of cognition.

My view, in contrast, is that the conceptual modules will already deal in discrete structured thoughts, and that their outputs will frequently be in a propositional format appropriate for encoding into language, classically conceived (Levelt, 1989). And while I do think that language makes a great deal of difference to our cognitive powers, I think that this is to be understood computationally (broadly understood), rather than in terms of learned associations and habits. This will become more evident as our discussion proceeds.

6. Second Problem: Creativity of Content

Humans are capable of generating new ideas and new hypotheses which aren't directly tied to perceivable properties of the environment—in fantasy, in play, in science, and in ordinary problem-solving. And many of these ideas, too, can have cross-modular contents. (See, for example, the stone gargoyles which can talk like

people in Disney's version of *The Hunchback of Notre Dame*.) Where would these new ideas come from in a modular cognitive architecture? They cannot be the outputs of particular modules, if the latter are designed to process information and issue in new domain-specific beliefs and desires. And they cannot result directly from the combinatorial powers of language, if all that language does is combine together and integrate the outputs of the various modular systems.

This is where we need to postulate a further module-integrating element of the mind—but one that ought to be computationally tractable, nevertheless (and so which might itself count as modular, in the sense employed here). We need to introduce a *supposition-generator* or *supposer*, in fact (Nichols and Stich, 2000); and for my purposes, it needs to be attached to the language faculty. The supposer would exploit the combinatorial powers of the language faculty to generate novel sentences, either initially at random, or (more likely) cued by similarities, analogies, and past associations. There is good reason to think that in childhood pretend play we see the first manifestations of this species-specific ability; and that the evolutionary function of pretend play is to develop and hone our capacity for creative thinking, as I have argued at some length elsewhere (Carruthers, 2002a). Here, as in section 4, let me just provide a few brisk points by way of elaboration.

Both childhood pretend play and adult creativity would appear to be species-specific capacities. So far as I know, the young of no other species on earth engages in pretend (as opposed to rough-and-tumble) play. And no other species displays anything like the creativity of thought and behavior that we do. Plausibly, both adult creative thinking and childhood pretend play involve essentially the same cognitive underpinnings. This is a capacity to generate, and to reason with, novel suppositions or imaginary scenarios. When pretending, what a child has to do is to *suppose* that something is the case (that the banana is a telephone; that the doll is alive), and then think and act within the scope of that supposition (Perner, 1991; Jarrold *et al.*, 1994; Harris, 2000; Nichols and Stich, 2000).[7] Similarly, when adults are engaged in the construction of a new theory, or are seeking a novel solution to a practical problem, or are composing a tune, they have to think: 'Suppose it were the case that P', or 'Suppose I did it like *this*', or 'Suppose it sounded like *so*'. Given these commonalties, it is then very plausible that the young of our species should engage in supposition-for-fun in childhood in order that they may be better able to suppose-for-real when they reach adulthood. (Note that this is a hypothesis about evolutionary function, not about children's aims and intentions.)

To see the kinds of principles on which a suppositional faculty might operate, consider the case of a young child pretending that a banana is a telephone. The overall similarity in shape between the banana and a telephone handset might be sufficient to activate the representation TELEPHONE, albeit weakly. If the child

[7] Leslie (1987) argues, in contrast, that what children need is the capacity to *meta-represent* their own representational states, hence 'de-coupling' them from their normal connections with belief and action. See Jarrold *et al.* (1994) and Nichols and Stich (2000) for critiques of this view.

has an initial disposition to generate an appropriate sentence from such activations, then she might construct and entertain the sentence, 'The banana is a telephone'. This is then comprehended and processed, accessing the knowledge that telephones can be used to call people, and that grandma is someone who has been called in the past. If the child *likes* talking to grandma, then this may be sufficient to initiate an episode of pretend play. By representing herself *as* making a phone-call to grandma (using the banana), the child can gain some of the motivational rewards of a real conversation. The whole sequence (including the initial generation of a supposition) is then reinforced, making it more likely that the child will think creatively again in the future.

From such simple beginnings one can imagine that children gradually build up a set of heuristics for generating fruitful suppositions—relying on perceptual and other similarities, analogies which have proved profitable in the past, and so on—leading eventually to the capacity for creative thinking and problem solving which is distinctive of human adults.

There doesn't seem any reason to think that the process of supposition-generation should prove computationally intractable, then. But notice that neither pretend play nor adult creative thinking involve *just* the generation of new ideas. Both children and adults also have to think, reason, and act within the scope of their initial supposition. For example, the child pretends (or supposes) that the banana is a telephone, then makes some dialing movements and begins an imagined conversation with her grandma. How does this get to happen? What needs to be put in place in order for us to reason from and act on our suppositions?

Arguably, nothing very much. There does need to be some adaptation of the working-memory system, so that we can keep track of the scope of our suppositions. And there has to be an accompanying disposition to bracket our suppositions and their consequences from issuing in belief. But there is nothing here that need strike one as computationally intractable, surely.[8] And as for drawing inferences from our suppositions, this can arguably be done by formulating the relevant sentences in 'inner speech' where they can be taken as input by the comprehension sub-system of the language faculty, and fed forward through the various modular conceptual systems. Most of the inferences can thus be module-specific ones; and those that aren't, we can suppose to come for free with the semantics of natural language.

7. Third Problem: How New Contents Get Accepted

Integrating beliefs from distinct domains and generating new suppositions is by no means the full extent of what is distinctive of human cognition, of course. We also come to *believe* some of our suppositions. This happens in problem solving, where

[8] Intuitions of computational tractability are notoriously unreliable, however. Here, as elsewhere in this paper, there is a promissory note drawn on future work in AI.

we generate a number of proposed solutions to some task, and then settle upon the one that we judge the best. It also happens in ordinary abductive inference, as when hunters consider a number of hypotheses concerning the most likely animal to have made a given scuff-mark in the dust, and then agree upon one of them as the most plausible (Liebenberg, 1990; Carruthers, 2002b). And it happens in science. In short, humans can do 'inference to the best explanation'. How is this possible within a modular framework? Does inference to the best explanation require us to postulate a radically a-modular central processing arena, just as Fodor has always suspected?

These questions will now occupy us for the remainder of this paper. But to begin with let us ask: what are the main elements of inference to the best explanation? While no one any longer thinks that it is possible to codify the principles involved, it is generally agreed that the good-making features of an explanation include such things as: *accuracy* (predicting all or most of the data to be explained, and explaining away the rest); *consistency* (internal to the theory or model); *coherence* (with surrounding beliefs and theories, meshing together with those surroundings, or at least being consistent with them); *simplicity* (being expressible as economically as possible, with the fewest commitments to distinct kinds of fact and process); *fruitfulness* (making new predictions and suggesting new lines of inquiry); and *explanatory scope* (unifying together a diverse range of data). Our problem now reduces to: how do these elements get built into our procedures for believing new sentences?

It may prove helpful to note, at this point, that the problem of deciding whether or not to believe new sentences would have been around from the first evolution of the language faculty, probably from significantly before the first appearance of our capacity for creative thinking, if my own proposals for the timing of the emergence of that ability are correct (Carruthers, 2002a). For one of the main things that language has always been for, presumably, is *testimony*. People tell other people things, and those others then have to decide whether or not to believe what they are told.

I propose, then, to look at what is involved in the acceptance of testimony, to see if we can understand it in modularist terms—using that as some sort of prototype or basis for explaining how inference to the best explanation might come to operate in language-involving thought. What I shall suggest, in fact, is that the principles involved in linguistic testimony and discourse interpretation might *become* a set of principles of inference to the best explanation once self-generated sentences start to be processed internally, in inner speech. And I shall suggest, too, that those principles could very well operate in ways that are computationally tractable. This will then be the final plank in my sketch of a proposed solution to Fodor's Problem.

8. Testimony

There are two broad models of testimony in the epistemological literature. One is the sort of reductive approach dating back at least to Hume, which claims that you

should only believe an item of testimony if you have independent evidence of the reliability of testimony in general, and of this individual source in particular. This view has come in for sustained criticism in recent years, since it is psychologically extremely implausible. Children start believing what they are told from the start. Yet it is hard to accept that they are irrational in so doing. For if they didn't, then they would never learn the language of their parents, and they probably wouldn't live for very long, either.

This critique motivates the contrary model, which claims that testimony is epistemologically basic, and that the default is just: believe what you are told (Coady, 1992; Burge, 1993; Owens, 2000). Extending this view just a little and turning it into a historical hypothesis, someone might propose that throughout the period of the evolution of language this default setting was the only rule in operation; and that it is only with the arrival of our capacity for creative thinking and inference to the best explanation that we acquired the power to be a bit more selective in the testimony we accept. If this proposal were on the right lines, then it would undercut the suggestion that we might appeal to pre-existing features of testimony in explaining how our capacity for inference to the best explanation first arose, and in explaining how that capacity is now realized in a modular cognitive system. On the contrary, the order of explanation should be the other way round.

Luckily, this extended view isn't remotely plausible. On just about all accounts of the evolution of language, our theory of mind abilities would need to have been highly developed and pretty firmly in place before language could make its appearance (see, e.g., Origgi and Sperber, 2000). And on just about all views of human ancestry, many of the pressures on our evolution were social/competitive ones (Byrne and Whiten, 1988; Mithen, 1996). So for sure, from the outset of language-use, language would have been used to manipulate and deceive as well as to inform. In which case consumers of testimony would have needed, from the start, to be discriminating about what testimony to accept. It might be correct to say that the default setting is acceptance. But it is still the case that utterances and circumstances would need to be monitored for clues indicating that a given item of testimony should be rejected.

This thought receives some confirmation from recent studies of the role of testimony in child development, which have found that even quite young children can be discriminating about sources of testimony and the likely reliability of testimony (Harris, 2002, forthcoming). In particular, even very young children will reject items of testimony that conflict with what they already believe ('Fish live in trees', 'Cats bark'). Yet these same children will happily accept and reason with such statements if they are introduced with a 'let's pretend' voice-intonation, or in the context of story-telling. Moreover, it should be noted that these are pre-four-year-old children, who as yet lack any explicit conception of false belief, and who cannot recall their own previous false beliefs (Gopnik, 1993). So the processes in question are likely to be automatic and unreflective.

It seems plausible, then, that the principles of testimony-acceptance are historically and developmentally prior to the principles of inference to the best

explanation. Two questions arise. First, could the principles of testimony-acceptance be realized in ways that are both computationally tractable and consistent with a modularist framework? And second, might our principles of inference to the best explanation have been constructed out of those of testimony-acceptance, once internally-generated sentences started to be produced by the language system and processed in something like the way that items of testimony are?

9. The Cognitive Components of Testimony Evaluation

Although accuracy, consistency and coherence of theories aren't quite the same thing, for our purposes they can be treated together. Subjects are less likely to accept a piece of testimony if what they are being told doesn't fit in with what they already believe and/or is internally inconsistent. How could one test for these properties within a modularist framework in computationally tractable ways?

How do you check a statement for consistency with what you already believe? Do you need to access every single one of your beliefs, simultaneously, for comparison? If so, then checking testimony for accuracy with one's beliefs would be the very epitome of computational intractability! But actually, it looks as if one can get a fair approximation to what is required in computationally tractable form. Assuming that memory systems are organized along content-addressable lines, one can do a search on the conceptual elements of the statement being evaluated. If the statement is, 'Buffaloes are dangerous' then one should do a search of one's buffalo-beliefs and a search of one's harmless-thing-beliefs to see if one can find a direct conflict. At the same time the sentence can be fed as input to the various conceptual modules, which will draw inferences from it, and those further conclusions can also provide the subject with materials for further searches of memory.

When human beings first started producing new sentences creatively through the supposition generator, then, and began displaying those sentences to themselves in 'inner speech', there would already have been in place some of the processes necessary to evaluate those sentences for acceptability, treating them as if they were items of testimony. And those processes look like they should be computationally tractable ones.[9]

But what of simplicity, fecundity, and explanatory scope? These are crucial elements of inference to the best explanation. But would they play a role in testimony-evaluation prior to the beginnings of creative thinking and hypothesis-generation? Not directly, perhaps. It is hard to see a role for such principles when evaluating a statement like, 'Buffaloes are dangerous'. But I find it highly suggestive that on some influential accounts of discourse interpretation, principles of *relevance*

[9] Once again (given the fallibility of our intuitions of computational tractability) I have to issue a promissory note here which I am myself unable to honor.

play a central role (Sperber and Wilson, 1986/1995). For the twin principles of relevance are: *minimize processing effort* (which roughly amounts to the same as: *seek simplicity*); and: *maximize information generated* (which roughly corresponds to: *seek a combination of fecundity and broad scope*).

The principle of relevance in communication amounts to something like this, then: other things being equal, *adopt the interpretation of the other person's utterance that is simplest and most informative*. And in default circumstances or circumstances where the credibility of an informant isn't in question, the principle of relevance can be formulated as a principle of testimony-acceptance, thus: *believe the interpretation of the other's utterance that is simplest and most informative*.

It is not so very hard to imagine, then, that in contexts where what is in question are self-generated sentences (where interpretation isn't an issue, of course), this same principle might be co-opted to become: *believe the sentence that is simplest and most informative*. And then this, combined with the insistence on consistency and accuracy, would give us all of the main elements of inference to the best explanation. A 'faculty' of inference to the best explanation could then be constructed 'for free' from principles already present and employed in linguistic communication and testimony.

10. Relevance and Computational Tractability

Is relevance maximization computationally tractable, however? As I understand it, the main practitioners of relevance theory don't think that subjects make explicit and direct computations of degrees of relevance (Sperber and Wilson, 1996). However, they do think that judgments of relevance might be made by a semi-independent sub-module of the folk-psychology faculty, which was selected for in human evolution because of its role in language comprehension (Sperber and Wilson, 2002). There are a number of aspects to this account which need to be discussed in turn.

Sperber and Wilson argue that the various modular systems out of which our cognition is constructed, together with the principles according to which they operate and interact, will have been honed by evolution to maximize relevance. For humans occupy what a number of people have described as an *informational niche*. We seek out and store large quantities of information about our environments and social circumstances. Since both storage and maintenance of information are costly, there will have been a premium set on seeking out information that can be economically represented and that is likely to have significant cognitive effects. And since retrieval of information, too, is apt to cost both cognitive resources and time, it is important that salient information should be recoverable swiftly. We are therefore likely to have evolved procedures of attention and resource-allocation which maximize the amount of information that can be acquired and accessed with minimum processing effort; and it is likely that the algorithms now embedded in our various modular systems will be similarly directed towards achieving relevance.

The evolution of human systems of communication (primarily gesture and language) would have been governed by similar constraints. A communicative attempt of some sort can only be successful if its recipients attend to it, and take the time and make the effort to figure out its significance. Would-be communicators therefore need to ensure that their missives are *worth* attending to: in short, they need to ensure that their messages will be *relevant* to their audience. And whatever may be true of the evolutionary background, the proposal that linguistic communication amongst humans today is governed by the goal (amongst speakers) and the presumption (amongst hearers) of relevance, is one that has considerable explanatory power (Sperber and Wilson, 1986/1995).

Let me focus on the standpoint of the hearer, here, since it is this that we need to exploit in understanding how self-directed 'inner speech' might be governed by principles of inference to the best explanation. In assessing relevance, do hearers have to represent all of the available interpretations of an utterance? And do they have to extract all of the information that can be inferred from those interpretations, calculating the degree of difficulty of their own computations as they go? If so, then it might well be the case that judgments of relevance are computationally intractable.

Sperber and Wilson (1996, 2002) argue that nothing of this sort needs to take place, however. Rather, subjects adopt a *satisficing* policy, governed by heuristics. They begin from the interpretation that is most *accessible* and/or *salient*, given the physical and conversational context. They then set their various computational systems to work in extracting information from that interpretation. If they achieve results that are relevant enough, they stop, assuming that this is the intended message. If they don't, then they move on to the next most accessible interpretation and set to work on that. In order for this procedure to be successful, subjects might need to monitor how *hard* they are having to work in extracting information from a given interpretation of a sentence. And they might need to monitor how *much* information they are extracting, where the extraction process is undertaken by a bunch of inferential processes which operate automatically (and for our purposes we can suppose: by the various modular systems to which the sentence is fed as input). But at any rate, it doesn't appear that achieving judgments of relevance should be computationally intractable.

It would seem, then, that the hope we expressed in section 7 above may well turn out to be realized. Not only can principles of inference to the best explanation be understood in terms of a prior set of principles of linguistic testimony and discourse interpretation, but those principles might very well be implemented in computationally tractable ways.

11. Spandrels *versus* Functions

Does the above discussion suggest that distinctively human thinking is a *spandrel*, however? Would it follow that such thinking is a mere by-product of other

selected-for aspects of cognition (a modular language faculty, together with principles of testimony-acceptance and discourse-interpretation)? But then, if so, is it really believable that so much of what is distinctive of our cognition should be a mere by-product, especially when the adaptive consequences seem so vast? (Consider the rapid expansion of *Homo sapiens sapiens* around the globe in the course of a mere 100,000 years or so, and the astonishing expansion of science and technology in just the last 500 years, for examples.)

Certainly, there is an element of happenstance in my proposed account. For there is no intrinsic connection between the evolution of language for the communication of module-specific contents and the appearance of distinctively human thinking. And it is (perhaps) fortunate that the principles deployed in interpretation of speech and the assessment of testimony should be capable of doing double-duty as principles of inference to the best explanation. (I shall challenge this point in a moment.) But there is nothing really surprising or remarkable about all of this. It is a routine finding in biology that items initially selected for one function should become co-opted and re-used in the service of others.

In any case, however, there are a number of respects in which direct selection for aspects of distinctively human thinking will probably have played a role, on my account. Most obviously, the supposition-generator needed to be constructed from scratch. Arguably such a mechanism could be built quite simply using materials that were already available (natural language structures, patterns of semantic association and similarity). But the mechanism itself would have been new, as would have been the behavior (pretend play) that I hypothesize to have been selected for because of its role in helping to strengthen and fine-tune the disposition to think creatively, which lies at the heart of the mechanism.

In addition, it isn't enough that language should already have been *capable* of linking together and combining module-specific contents. There also needed to be a disposition for it to do this on a regular basis. Now admittedly, this may be partially explained by the evolutionary pressures on communication, since combining contents can lead to compression in their mode of expression. But still there needs to be a disposition to generate such sentences in auditory imagination on a regular basis (in 'inner speech'), and to take those sentences or their consequences to be candidates for belief, depending upon their effects. These dispositions would presumably need to have been selected for. (It is hard to imagine how they might be learned behaviors, or culturally transmitted ones.)

Before concluding, however, let me return briefly to a point that was conceded too swiftly above: that there isn't any intrinsic connection between principles of testimony-acceptance and principles of inference to the best explanation. For both would appear to be truth-directed, in fact, as well as being governed by similar pragmatic constraints of cognitive economy. In general, we only want to accept items of testimony that are true, just as we only want to accept theories if they are true, or at least advance us closer to the truth. And both internal inconsistency and conflict with received belief would seem to be signs of falsehood in both testimony and theory-construction.

But what of the preference for simplicity and fecundity? Is there any connection between this as a principle of interpretation and this as a principle of theory choice? Arguably the same goal underlies both domains: the need to maximize useful information. Other things being equal, we want *as much* information as possible (whether in communication or in science), but we want it in a form that is as *useful* as possible, presented in a format that enables us to draw conclusions as and when we need them.

12. Problems and Conclusions

In summary, then, the three main elements of distinctively-human cognition are: (1) flexibility of content; (2) creativity of content; and (3) a capacity for inference to the best explanation. The first can be laid at the door of a modular language faculty. The second requires the postulation of a (computationally pretty simple) supposition generator. The third can be handled by principles that would already have been in place to govern testimony-acceptance and discourse-interpretation; and it would appear that these principles, too, could be realized in computationally tractable ways. Although much detailed work remains to be done, it certainly *seems* that none of these capacities is likely to prove computationally *in*tractable.

A number of outstanding problems remain, of course (quite apart from the manifest sketchiness of my proposals). I shall briefly mention three. The first is that my account assumes that central modules can be directed to work *relevantly and intelligently* to generate information which might aid in the solution of current problems. How can this be achieved without us needing to postulate some sort of over-arching executive system, which would then look uncomfortably like a (computationally intractable) General Problem Solver? There is some reason to hope, however, that the notion of *attention* might be developed in such a way as to come to our aid, at this point.

We are familiar with the idea that background purposes and interests (as well as salience of stimuli; see below) can effect both the inputs to, and the processing undertaken by, modular input-systems such as vision. And no one thinks that this must render the operations of the visual system mysterious. So it might be that something similar can be made to work for central modules, too. Maybe there is something a bit like *attention* at work within conceptual modules, directing them to pay attention to certain inputs, and to be especially on the lookout for certain sorts of information which might be generated from those inputs. At any rate, the idea seems worth exploring.

The second problem left outstanding in my account is that there must be something that *selects* from amongst the various modular outputs and the creations of the suppositional system, choosing just one (or at most a few) such contents to be combined and/or formulated in inner speech. How is this supposed to happen? Does *this* require us to postulate some sort of a-modular executive system within which all of the real intelligence of distinctively-human cognition would be

buried, beyond possibility of understanding? Of course I hope not. And here I suspect that some extension of the notion of *salience* might be co-opted to our aid.

We are also familiar with the idea that some stimuli received as input by the visual system are especially salient and demanding of attention, either through innate preparedness or previous learning (loud noises; naked bodies). Maybe something similar can be at work in connection with the outputs of the conceptual modules (which are competing to be taken as inputs by the language production system), helping to select some of those contents automatically, without the need to postulate any sort of executive. But this is an idea that remains to be worked out.

Finally (and most importantly), I haven't really said anything about how language-involving human thinking can become *practical*, or can have an impact upon our behavior. I have made a start on this problem elsewhere (Carruthers, 2002d, forthcoming); here, unfortunately, there hasn't been the space to pursue it. But let me say in mitigation that Fodor's Problem, as initially formulated, was a problem about human *cognition* (or human 'belief-fixation', in Fodor's terminology), rather than about human mental activity more generally. Of course it was always understood that beliefs, once 'fixed', would need to be available to guide practical reasoning and action. But that was never Fodor's main focus; and nor, accordingly, has it been mine.

In conclusion, then, my claim isn't really that Fodor's Problem has now been solved. It is rather that there is some reason to hope that it *can* be solved, and that we can at least *begin* to understand human cognitive processes in massively modular terms. At any rate, I claim that Fodor's pessimism on the subject is highly premature.

Department of Philosophy
University of Maryland

References

Baars, B. 1988: *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. 1997: *In the Theatre of Consciousness*. Oxford: Oxford University Press.

Barkow, J., Cosmides, L. andTooby, J. (eds.) 1992: *The Adapted Mind*. Oxford: Oxford University Press.

Bryson, J. 2000: Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12, 165–190.

Burge, T. 1993: Content preservation. *The Philosophical Review*, 102, 457–488.

Byrne, R. and Whiten, A. (eds.) 1988: *Machiavellian Intelligence*. Oxford: Oxford University Press.

Carruthers, P. 1998a: Thinking in language?: evolution and a modularist possibility. In P. Carruthers and J. Boucher (eds.), *Language and Thought*. Cambridge: Cambridge University Press.

Carruthers, P. 1998b: Conscious thinking: language or elimination? *Mind & Language*, 13, 323–342.

Carruthers, P. 2000: *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Carruthers, P. 2002a:. Human creativity: its evolution, its cognitive basis, and its connections with childhood pretence. *British Journal for the Philosophy of Science*, 53, 1–25.

Carruthers, P. 2002b: The roots of scientific reasoning: infancy, modularity, and the art of tracking. In P. Carruthers, S. Stich, and M. Siegal (eds.), *The Cognitive Basis of Science*. Cambridge: Cambridge University Press.

Carruthers, P. 2002c: The cognitive functions of language. *Behavioral and Brain Sciences*, 25:6.

Carruthers, P. 2002d: Author's response: Modularity, language, and the flexibility of thought. *Behavioral and Brain Sciences*, 25:6.

Carruthers, P. 2003: Moderately massive modularity. In A. O'Hear (ed.), *Mind and Persons*. Cambridge: Cambridge University Press.

Carruthers, P. forthcoming: Practical reasoning in a modular mind.

Cherniak, C. 1986: *Minimal Rationality*. Cambridge, MA: MIT Press.

Chomsky, N. 2000: *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Coady, C. 1992: *Testimony*. Oxford: Oxford University Press.

Dennett, D. 1991: *Consciousness Explained*. London: Penguin Press.

Fodor, J. 1983: *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. 2000: *The Mind Doesn't Work that Way*. Cambridge, MA: MIT Press.

Fodor, J. and McLaughlin, B. 1990: Connectionism and the problem of systematicity. *Cognition*, 35, 183–204.

Fodor, J. and Pylyshyn, Z. 1988: Connectionism and cognitive architecture. *Cognition*, 28, 3–71.

Gallistel, R. 1990: *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, R. 2000. The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (ed.), *The New Cognitive Neurosciences* (second edition). Cambridge, MA: MIT Press.

Gopnik, A. 1993: How do we know our minds: the illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–14.

Harris, P. 2000: *The Work of the Imagination*. Oxford: Blackwell.

Harris, P. 2002: What do children learn from testimony? In P. Carruthers, S. Stich, and M. Siegal (eds.), *The Cognitive Basis of Science*. Cambridge: Cambridge University Press.

Harris, P. forthcoming: Checking our sources: the origins of trust in testimony. *Studies in History and Philosophy of Science*.

Hermer-Vazquez, L., Spelke, E., and Katsnelson, A. 1999: Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39, 3–36.

Hirschfeld, L. and Gelman, S. (eds.) 1994: *Mapping the Mind: domain specificity in cognition and culture*. Cambridge: Cambridge University Press.

Horgan, T. and Tienson, J. 1996: *Connectionism and the Philosophy of Psychology*. Cambridge, MA: MIT Press.

Jarrold, C., Carruthers, P., Smith, P. and Boucher, J. 1994: Pretend play: is it meta-representational? *Mind & Language*, 9, 445–468.

Kosslyn, S. 1994: *Image and Brain*. Cambridge, MA: MIT Press.

Levelt, W. 1989: *Speaking*. Cambridge, MA: MIT Press.

Liebenberg, L. 1990: *The Art of Tracking: The Origin of Science*. Cape Town: David Philip Publishers.

Marcus, G. 2001: *The Algebraic Mind*. Cambridge, MA: MIT Press.

McDermott, D. 2001: *Mind and Mechanism*. Cambridge, MA: MIT Press.

Mithen, S. 1996: *The Prehistory of the Mind*. London: Thames and Hudson.

Nichols, S. and Stich, S. 2000: A cognitive theory of pretence. *Cognition*, 74, 115–147.

Origgi, G. and Sperber, D. 2000: Evolution, communication and the proper function of language. In P. Carruthers and A. Chamberlain (eds.), *Evolution and the Human Mind*. Cambridge: Cambridge University Press.

Owens, D. 2000: *Freedom within Reason*. London: Routledge.

Paulescu, E., Frith, D. and Frackowiak, R. 1993: The neural correlates of the verbal component of working memory. *Nature*, 362, 342–345.

Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Pinker, S. 1997: *The Way the Mind Works*. London: Penguin Press.

Rey, G. 1997: *Contemporary Philosophy of Mind*. Oxford: Blackwell.

Rizzolatti, G., Fogassi, L. and Gallese, V. 2001: Neuropsychological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661–670.

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

Shergill, S., Brammer, M., Fukuda, R., Bullmore, E. Amaro, E., Murray, R., and McGuire, P. 2002: Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, 16, 219–27.

Sperber, D. 1996: *Explaining Culture*. Oxford: Blackwell.

Sperber, D. 2002: In defense of massive modularity. In I. Dupoux (ed.), *Language, Brain and Cognitive Development*. Cambridge, MA: MIT Press.

Sperber, D. and Wilson, D. 1986: *Relevance: communication and cognition*. Oxford: Blackwell. (2nd Edition 1995.)

Sperber, D. and Wilson, D. 1996: Fodor's frame problem and relevance theory. *Behavioral and Brain Sciences*, 19, 530–532.

Sperber, D. and Wilson, D. 2002: Pragmatics, modularity and mind-reading. *Mind & Language*, 17, 3–23.

Sperber, D., Premack, D. and Premack, A. (eds.) 1995: *Causal Cognition*. Oxford: Oxford University Press.

Tager-Flusberg, H. (ed.) 1999: *Neurodevelopmental Disorders*. Cambridge, MA: MIT Press.

Tooby, J. and Cosmides, L. 1992: The psychological foundations of culture. In Barkow, J., Cosmides, L. and Tooby, J. (eds.), *The Adapted Mind*. Oxford: Oxford University Press.

Varley, R. 1998: Aphasic language, aphasic thought. In P. Carruthers and J. Boucher (eds.), *Language and Thought*. Cambridge: Cambridge University Press.

Varley, R. 2002: Science without grammar: scientific reasoning in severe agrammatic aphasia. In P. Carruthers, S. Stich, and M. Siegal (eds.), *The Cognitive Basis of Science*. Cambridge: Cambridge University Press.