

Fodor Encapsulated

J. A. Fodor, *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology* (Cambridge, MA: MIT Press).

JOHN COLLINS

Evolutionary psychology is on the rise, both in academia and the public arena. The governing metaphor is that the mind is a Swiss army knife: it consists of a set of specialised computational devices (modules) that subserve our domain specific competencies in language, face recognition, *theory of mind*, cheater detection, *et al.* We might even have a module dedicated to the extirpation of stones from the hooves of horses. Evolution (natural selection, in particular) is the designer of this *massively modular* architecture: each module is an adaptation to a problem faced by our hunting and gathering ancestors; we discover the content of a module by determining what it was selected *for* back on the Savannah. Fodor, *pace* Pinker (1998) and a host of others, thinks the mind doesn't work this way. We have some good ideas about how some parts of the mind *might* work, Fodor avers, but we haven't got a clue about how most of the mind works *salva* some new scientific insights.

Unlike Dennett, Penrose and Pinker, serious thinkers all, Fodor is not a populist, as may be judged by his happily svelte volume (that popular books about cognition should also serve as doorstops is regrettable). The arguments, in the core of the book at least, are dense and technical; there is no surfeit of intuition pumps nor (heaven forbid) analogies drawn from *Star Trek*. If, then, you are new to the area, much of the book will prove difficult, although Ch.5, based largely on Fodor's (1998) review of Pinker (1998) and Plotkin (1997), is more polemical than its predecessors.

As with green readers, those who are seeking a manifesto will be disappointed. Fodor doesn't kick God while he's on his knees, still less

is he concerned to cup his palms around the ineffable spark of humanity as the blind processes of selection blow all around. One of the great things about Fodor, like that other contemporary rationalist, Chomsky, is that in matters scientific he cares only for the truth; bother being tough-minded. Putting the boot in, for sure, is fun, if hardly edifying, but the truth is much more interesting.

Still, I recommend Fodor to everyone who is seriously concerned with cognitive science and the putative fecundity offered by evolutionary theory. In the sequel, I shall outline the shape of Fodor's argument and then offer some thoughts of my own.

For Fodor, Alan Turing is the only guy who has had a good idea about cognition. The idea is that the rationality (truth preservation) characteristic of thought can be captured by a syntactically driven causal process. What links content to cause is syntax or *logical form* (not to be confused with the linguist's LF). Thinking, on this view, is computation defined over syntactically constituent structures. The syntax of the structures (shape, size, whatever) determines their causal role; the semantic interpretability of the structures means that the causal role is also a rational one that may be captured by our familiar intentional generalisations. If Smith believes that *A & B*, he will also believe that *A* and that *B*. This is because, so the story goes, to have the first belief is to token a complex representation with constituents corresponding to the semantically relevant parts of the complement clause, which, *ipso facto*, is to token representations with the content of *A* and *B*. An intentional law is thus instantiated by a mechanical (syntactic) process. More complicatedly, Smith goes for his umbrella when he *desires* not to get wet and *believes* that his umbrella will achieve such an end *because* Smith's mind is such that the mental representations about rain, umbrellas, etc. tokened in Smith's belief and desire boxes (think of a flow chart) have a syntax which realises a causal role that eventuates in going-for-umbrella behaviour.

In short, a computational theory of mind (CTM) shows us how content (information) can get its hands on the steering wheel. But, and this is a big 'but', CTM cannot be true *in toto*. The problem is this: only the syntactic properties of a representation determine its effects. Consider a vending machine. If you put in a fifty pence piece, out pops a *coke*. But the coin's property of *being worth 50p* didn't cause the *coke* to be deposited. How *Coca-Cola corp.* wish that it would! The causally relevant properties of your coin are its size, shape, density, etc., *viz.*, its *intrinsic* (local) properties *P*, not its *extrinsic* ones. You need the police

and other state organisations to make it so that *being worth 50p* supervenes on *P*. But no amount of law enforcement will constitute nomological necessity. Just so for mental representations, as construed by CTM. Their causal properties are local, not global; but the mind is up to its ghostly ears in abduction, context sensitive inference, and the assessment of relevance and overall soundness. This worry is not new to Fodor, although it is easily missed, especially by those who didn't get to the end of *The Modularity of Mind* (1983). There Fodor argued that CTM is good idea about modules because they have a proprietary database over which the specialised processes are defined; such is what it means for modules to be domain specific: they are computationally local *qua encapsulated* from the rest of cognition. As for global cognition... Sorry, I haven't a clue. Try to design a vending machine that will deposit a coke *only if* a coin worth 50p is put into its slot. (*Cognoscenti* will note that I am conflating external extrinsic properties with internal ones; the difference does not here matter.)

Massive modularity might seem a veritable boon. Modular processing is by definition restricted to proprietary databases. CTM provides a nice theory of the working of the modules, for Turing's story is that computation is a local process *qua* determined by representations' intrinsic properties. Thus, the more restricted the content is to which the computations have access, the more that content can be understood to enter into the causal explanation of the system's behaviour. Conversely, the less restricted the information is, the less we understand how Turing computations can be defined over it; that is, we don't have a computational story of global cognition. Someone committed to CTM, therefore, had better hope that massive modularity is true; for if modularity is not the norm, then it would seem that CTM can at best be the partial truth about cognition. Two problems: the massive modularity thesis is false and a lot of cognition *is* global anyhow. Bother Darwin, there is good reason to think that the mind is not a Swiss army knife independent of one's favourite story about the phylogeny of cognition. Such is Fodor's argument.

This is all very deep, and I agree with its general thrust; still, I have reservations.

Fodor is very good at showing why massive modularity is attractive. Indeed, Fodor's *The Modularity of Mind* is largely responsible for the popularity of modular theorising. Fodorian modules, however, are peripheral input/output systems: fast, dedicated, unreflective, hardwired devices that respond to stimulus. They trade veridicality and depth of

output for speed and shallowness of output. This makes evolutionary sense (false negatives are a nuisance, false positives lead to extinction). For Fodor, the *sine qua non* of modularity is encapsulation: a module's computations are restricted to a proprietary database; that is why they are fast, dedicated, etc. Fodorian modules are not central. Now, one might say: 'If this is what modules are like, then obviously globality is a real headache. But the shift to thinking about central modules necessitates a relaxation on encapsulation; modules co-penetrate, exploiting the resources of each other. What looks like global cognition is really the ensemble effect of a whole bunch of modules'. Such a story, I think, is the one any thinking evolutionary psychologist (which is not a pleonasm) would want to tell. Using Milikan's (1984) jargon, adopted by Gigerenzer (1997) and Sperber (1994), one distinguishes between a *proper* and *normal* domain. The former is that content for which the module was selected to compute over; the latter is that content which the module currently computes over. The idea is that the mind way back before we even hunted and gathered comprised a whole load of little dedicated gizmos, each with their proper domains. This ensemble, under selective pressure, evolved to be an increasingly integrated system. Alternatively, syntax might have dropped from the sky 100,000 years ago and provided us with a metarepresentational *lingua franca*. For all we know, dogma aside, this might be true (Bickerton (1995); but see Bickerton and Calvin (2000) for a less catastrophic phylogeny). Either way, eventually, we end up with minds like ours with normal domains more or less smeared over sub-ensembles. Is this okay?

Well, anyone who tells this tale should be careful. Massive modularity is meant to be a story about cognition as it is now, not a story about how it used to be. A module that is totally unencapsulated just ain't a module. Fodor (p.63) may be read as saving time by ignoring the kind of story set out above. What Fodor should not ignore, though, is the idea that modules do not need to be totally encapsulated. One can, it seems to me, believe that central modules come in ensembles. A disciplined multi-path boxology is required lest the theory fall into vacuity, but everyone has their problems. Furthermore, there are good cases of putative ensembles.

Consider language. Fodor (1983) thinks of the language module as a parser, and, to my knowledge, has not retracted the idea. It is clear, however, that parsibility is not a condition on the output representations of the language module. Sentences with centrally embedded relative clauses (*the boat the sailor the dog bit built sank*) and 'garden paths'

(*the horse raced past the barn fell*) tend not to be parsible, although they are perfectly grammatical, as can be seen on reflection (hint: insert the missing complementisers). Even so, only the benighted think that linguistic competence is not modular based (*contra* Cowie (1999), poverty of stimulus considerations, *inter alia*, really do show that a dedicated mechanism armed with universal grammar is the only idea we have). Yet our language faculty, it would seem, makes use of information not employed in parsing, i.e., it is not quite encapsulated. More speculatively, I (Collins (2000)) have conjectured that the Theory of Mind (ToM) module is interfaced with the language faculty; the thesis explains some interesting ontogenetic and pathological data and many features of propositional content (opacity, embeddness, *et al.*). In general, the language module appears to take its place within a broader system of modules.

Fodor really does not consider this kind of modular ensemble story. He does, however, deploy an argument that should give everyone serious pause for thought (pp. 71-8). In a nutshell: we can understand how peripheral modules are triggered because they are hardwired to our sensorium, as it were. What of central modules? Presumably, they need some module(s) to input the relevant representations. In simple terms, a central module needs to know when it is time to do its thing. The problem is that, by definition, the inputting module(s) will be *less* modular than the receiving module; for, however the story is told, the inputting module(s) must sort between representations that carry information that is and is not proper to the receiving module. Alternatively, the inputting module(s) may not be less modular, but such a case requires an infinite number of them. This is a splendid argument and one I have pondered myself. I'll come to my thoughts presently; first, however, Fodor adds a twist.

One way to stop the threatened regress of modules is to understand the sensorium as the ultimate source of all triggers. This move, though, is only available to the mad dog empiricist; for it amounts to the claim that *every* cognitive distinction is an empirical one, i.e., *all* selective sorting of representations is made on the basis of empirical features. We on the side of the angles have long despaired of connectionist groupies and the anti-Chomskyians for just this reason. As Fodor (p.77) points out, there appears to be no sensory feature that tells the language perception module that language is abroad. (Just think about the perfectly acceptable languages, ASL and varieties of *home sign*). More grossly, what on earth is the perceptual feature that tells your cheater

module that you might be being fleeced? If there were such a trigger, cheating, for good or ill, would, contrary to fact, be nigh impossible.

As I said, I take this argument very seriously, and the massively modular crowd have not so much as recognised it, still less answered it. What, then, of my story above? Here's a speculation about ToM in particular (see Collins (2000) for details). What tells ToM that there is a cognitive agent abroad is its face recogniser; that is, anything with a face is, *ceteris paribus*, apt to trigger the attribution of mental states. Something like this may explain why we happily attribute mental states to animals while realising that, say, Rover really does not believe that if he can get me to think that he has not been fed, then I shall be duped in feeding him extras. Likewise, *once* the child has fixated on a grammar, then sensory input will suffice to mark language from non-language. The *real* problem for empiricism is that the child cannot fixate on a grammar in the first place without already knowing the options which need to be decided. The point in both cases is that the cognitive distinctions (language or not; cognitive agent or not) aren't empirical ones; rather, given enough innate information, all we need are *cues* to downstream modules that dutifully select as default certain upstream modular processes. I'm not certain about the soundness of this proposal, but I cannot see how it is obviously incorrect. That said, this proposal will not work for many of the proposed central modules, such as a cheater detector. I can happily live with this; it has long seemed to me that a cheater detector module was only slightly less silly than a line dancing one.

What has any this got to do with the globality of cognition? Well, quite a lot. The thing about modularised cognition is that it is mandatory; but surely, even the least of us *thinks* occasionally. Fodor does not stress this point, although I suspect it is behind much of his argument, for a clear way for thinking not to be mandatory is for it to be global. As we saw, the central modularist is obliged to ease off on encapsulation and allow mutual cognitive penetration, but not so much so that modularity becomes vacuous. To such an end I proposed my disciplined boxology, yet this also can only realise mandatory processes.

Here's another speculation. Global cognition is (reflexively) conscious; mandatory cognition is unconscious. What makes thinking global is that it is *Quinean*, i.e., any information is potentially relevant to the assessment of any claim. Only consciousness can realise this kind of isotropic thinking. Why? Because consciousness is basically off-line, we deliberate for as long as we like and then, if at all, act. Mandatory

cognition is by definition immune to any of our reflective strategies; it issues its output whether we like it or not. Modularity is a good story about mandatory cognition, a bad one about global cognition. We get from one to the other via consciousness. Here's what I think consciousness amounts to. The language faculty enables meta-representational thinking that gives us (partial) non-inferential access to the information otherwise involved in mandatory cognition. What makes the access non-inferential is that the content is linguistically encoded; for example, visual representations are not, I should guess, linguistic, but their content is already apt to be linguistically encoded in as much as it can be the object of judgement. Put differently, all reflexive propositional thinking is linguistic. Globality is all to do with the conscious linguistic exploitation of a modular ensemble that serves mandatory cognition.

The reader will have noted that I have heretofore said nothing about evolution; my silence is pregnant. Evolution (read: natural selection) fits into the story because it is often claimed that only massive modularity is evolutionarily possible. Lots of other similarly ill-judged conjectures are made in the same specious modality. I shall leave the reader to the joys of Fodor's Ch.5, where he gleefully demolishes such twaddle. I have my reservations on some points of detail, especially his remarks on Chomsky towards the end of the chapter, but such quibbles are for another time. I shall end, in the spirit of Fodor, with what I think is the real problem with evolutionary psychology.

It is not to be denied that we should like to know how human cognition evolved from the mind/brain of the ancestor we share with the chimpanzee. Further, let us assume that natural selection is the only story we have for the evolution of a system of the complexity of the human mind. (I'm as fond of I'm-the-only-president-you've-got arguments as the next guy, but, *pace* the zealots, they are not demonstrative; would that they were.) Somewhere down line, then, we shall be obliged to tell, as far as we can see, some selectionist story; so, why not tell it now? My dispute with this line is not so much substantive; I do not, for example, have a story about what happens when you cram 10^{11} neurones into the volume of a football; for all I know, a human mind might emerge, as they say, although I really don't think so. The complaint I have, rather, is methodological. I shall illustrate this with the case of language.

Three interesting questions about language are (i) What does a speaker/hearer know? (Humboldt's problem); (ii) How is this knowledge acquired? (Plato's problem); and (iii) How is this knowledge used? (Descartes's problem). Now considerations of evolution simply do not

impinge on these questions in the sense that we could answer all three without having a clue as to *why*, say, English, but not Italian, has obligatory subjects, or *why* children *never* regularise auxiliaries, or *why* lateralisation of linguistic competence after left-side brain damage sometimes precludes passivisation. I should not say that an evolution story fails *a priori* to shed any light here, but the logic of selection appears to be the wrong one: we are interested in *what* computational system underlies linguistic competence and *how* it works. I just cannot see the difference it would make to the acceptability of answers to these questions if language gradually evolved or if Alpha Centurions altered our DNA back when we mainly hunted and gathered. Unsurprisingly, given the otiosity of evolutionary considerations, when one approaches language with an evolvability constraint, muddled thinking ensues. This is the real problem.

When it is asked what language is *for*, the answer most often returned is 'communication'. (Incidentally, I don't think that language is *for* anything at all; nor are beliefs. We want our linguistic exchanges to be communicative, but they are not *supposed* to be; ditto for beliefs *vis-à-vis* truth.) An evolvability constraint tends to lead to an externalist story about organismic-environmental interaction; it is the environment that shapes the cognition so that it may better deal with the external demands. Naturally, this leads to externalist accounts of language, such as Clark's (1997, 1998) trading spaces model, where language is some kind of environmental feature onto which we can off-load cognitive resources. Dennett (1995, 1997) tells a similar tale. These stories are quite jejune, fashioned, as they apparently are, in complete ignorance of, or disregard for, the findings of linguistics and psycholinguistics. No-one, and I mean *no-one*, has the faintest idea how the abstract hierarchical structure of language may be encoded into a statistically retrievable phonetic form (to say nothing of sign language or the developmental alacrity of the blind). It seems that if knowing a language is being able to retrieve structured information from our environment, then we must already know the structure. As it happens, the vast majority of the data points to such *internalism*, and the rest is equivocal. I really am at a loss at the presumption of theorists to know what they are talking about when they turn to language. I would not dream of declaiming about electromagnetism without tailoring my thoughts to what the physicists have to say. Why do people think that they can have so much as coherent thoughts about language without caring what the linguists have to say? Evolutionary theory, I am afraid to say, must shoulder much of the

responsibility. Significantly, Pinker and Bloom (1990), who both know their linguistics, commend an evolutionary story but do not fall into the bog of externalism. Such is perhaps why they judge their thesis to be "boring", for it affects not a jot the practice of linguistics. This is as it should be.

Similar methodological complaints, I think, can be made against the imposition of an evolvability constraint to any cognitive domain, language simply offers perhaps the best example. Let us, then, with Fodor, eschew all evolvability constraints: bother being tough-minded; bother materialism; bother atheism. Let us pursue the truth.

JCollins42@compuserve.com

References

- Bickerton, D. (1995). *Language and Human Behaviour*. London: UCL Press.
- Bickerton, D, and Calvin, W. (2000). *Lingua Ex Machina*. Cambridge, MA: MIT Press.
- Clark, A. (1997). *Being There*. Cambridge, MA: MIT Press.
- Clark, A. (1998). 'Magic Words: How Language Augments Human Computation.' In P. Carruthers and J. Boucher (eds.), *Language and Thought*. Cambridge: Cambridge University Press.
- Collins, J. (2000). 'Theory of Mind, Logical Form, and Eliminativism.' *Philosophical Psychology*, 13: 465-490.
- Cowie, F. (1999). *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.
- Dennett, D. (1995). *Darwin's Dangerous Idea*. London: Allen Lane.
- Dennett, D. (1997). 'How to Do Things With Words.' In J. Preston (ed.), *Thought and Language*. Cambridge: Cambridge University Press.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fodor, J. (1998). 'Review of Stephen Pinker's *How the Mind Works* and Henry Plotkin's *Evolution in Mind*'. In *In Critical Condition*. Cambridge, MA: MIT Press.
- Gigerenzer, G. (1997). 'The Modularity of Social Intelligence.' In A. Whiten and R. Byrne (eds.), *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

- Pinker, S. (1998). *How the Mind Works*. London: Penguin Books.
- Pinker, S. and Bloom, P. (1990). 'Natural Language and Natural Selection.' *Brain and Behavioural Sciences*, 13: 707-27.
- Plotkin, H. (1997). *Evolution in Mind*. London: Allen Lane.