# Book Reviews

Jerry Fodor, *The Mind Doesn't Work That Way*, Cambridge, MA: Bradford Books/MIT Press, 2000, 126 pp., ISBN: 0-262-06212-7.

When asked what Cognitive Science was, it used to seem quite appropriate to say that it was whatever Jerry Fodor was up to at the time. This slim book is the latest installment, and one should be grateful that Fodor — like Hume (a comparison Fodor would not appreciate) — writes summaries of his position that get shorter with time. This volume was stimulated in part by a (larger) volume of Pinker's (1997) with a similar title, so this might be no more than a private fight between two major apologists for the Chomskyan position on these matters, in which case why should outsiders join in?

Luckily, the issues dealt with here extend beyond the usual Fodorian list, because Pinker was bold enough to tackle one of the oldest and most serious objections to the whole Chomskyan cosmogony, and Fodor is anxious to set him back on the true path. That issue is language development, outside the individual, that is: many have pointed out since the 1950s that the Chomskyan view, inspired as it was by an automaton that produced all and only the sentences of a language, was a fundamentally synchronic one. On Chomsky's later, constraint-based, view of how to capture all and only the possible human languages by means of innate constraints on the development of language in the individual, the goal was again to capture the set of languages we now have. But any view that languages have evolved to their present states assumes that they were not always as they now are, so how could either the automaton, or the constraints-in-the-brain, capture those other, earlier, languages as well, since they might have been very unlike the ones we now have, and related to quite different innate endowments in the brain?

Yet, if there is language evolution and the earlier languages are not within the set captured by UG (universal grammar), when in human history does UG start to operate? All this is precisely the kind of discussion Chomsky has always refused to have, but denial has ceased to work: Dennett (1995) brought Darwinian evolution back into the discussion, and mind and language began to have histories again, though one must recall that Lieberman raised all these issues in 1984 but was ignored by the establishment at the time. Now Pinker has found it necessary to make some compromise (what Fodor calls a New Synthesis) with a position that seems to have common sense on its side, i.e., that languages do evolve, like everything else, from simpler forms, and those forms will not necessarily have the same structures as modern ones.

*Minds and Machines* **13:** 321–336, 2003.

Before turningto this core argument between Fodor and Pinker, one must point out that the compression of this work, added to the need every academic feels to recite again the basis of his position at sufficient length, have led to some unfortunate simplifications. Some of these are just careless errors, that range from the almost-inspired typo "Santa Clause" (p. 15) to the logical howler (p. 12): "sentences of the form P and Q entail, and are entailed by, the corresponding sentences P, Q."

Fodor restates his all-purpose computational theory of mind (CTM), which he claims is also part of the New Synthesis and which he identifies with Turing: "... a mental process, qua computation, is a formal operation on syntactically structured mental representations" (p. 11); "... it's only if the sufficient conditions for an inference to be truth preserving are syntactic that Turing guarantees that a machine is able to recognise its validity" (p. 13). Fodor seems more or less to believe in CTM, at least as part of the story, though later he hints darkly that the whole Turing game may have to be abandoned.

But all this is very odd, at least in its attribution to Turing, for there is nothing in the latter's papers on Universal Machines (Turing, 1936) or the attribution of mental qualities to machines (Turing, 1950) that is remotely like this: his, rather sceptical, interest was in the possible mentality of machines, not the machinery of mind. For the latter, Craik (1943) is a far more plausible historical precursor for the views Fodor attributes to Turing on mental operations. Turing's concerns with universal computation had little directly to do with theorem proving by machine, which is what Fodor seems to have in mind in the second sentence quoted.

Later on, Turing seems almost a contemporary research collaborator of our author: "It's yet another way of putting Turing's insight that local structures can encode not only grammatical relations between sentences but inferential relations as well" (p. 21), and (p. 23) "Turing's idea that mental processes are computations ... ", and, again, (p. 24) "Turing's idea that cognitive processes are causal only if they are syntactic ... ".

It was once said that Hegel's works should be seen not as philosophy but philosophical novels, and there are elements of that in Fodor's novella: crucial terms are never explained (e.g., supervenience, causal powers) but assumed known from earlier episodes. Although short, the work is turgid and repetitive, sometimes literally incomprehensible, and one must seek insight in the general drift of themes rather than in linear arguments. After all, the text reveals him to be a fan of Martin Amis, who has written a novel backwards. It all gives the impression of being hastily written, possibly dictated, and not revised or edited; far from being too short it is actually too long, more a short story in content. And there are the awful, US Air Force style, acronyms, as in the first section of the book which rests, as I understand it, on the following crucial distinction:

"Principle E: Only essential properties of a mental representation can determine its causal role in mental life. I'll use E(CTM) as a name for the doctrine you get when you read the computational theory of mind as entailing principle E." (p. 24)

"Consider, therefore, the what (sic) I'll call the Minimal Computational Theory of Mind, M(CTM): . . . The role of a mental representation in cognitive processes supervenes on some syntactic facts or other." (p. 29)

Surrounding prose makes it pretty clear that for Fodor the syntax of a representation is essential to it, which could seem pretty odd, since that view takes no account of familiar situations in representation such as the range of equivalent formulations of the propositional calculus: the representations are superficially different (sometimes radically so, e.g., Scheffer's) but, whatever the syntax, they all still have the same representational power, namely that of the propositional calculus. Nonetheless, the above is clearly what Fodor intends, and it retains syntax in the central role he still keeps for it, and is probably connected to his well-known "sentences in the head" view since a sentence's syntax is essential to it, unlike a proposition's.

But what then, give or take some additional explanation, and some ascription of sense to "supervenes on" can be the real or significant difference between M(CTM) and E(CTM) as computational theories of mind based on representational syntax in some way? And yet for him this distinction seems vital, even though he seems to hold both in some form: the former can be true only if cognition is modular (p. 24), and he's been for that since 1983, hasn't he?

What makes ithard for the reader is that, like many modern novelists, Fodor is a tease: positions are worked up as being plausible if not convincing, but we have the feeling (at least, if we know his past work) that he will then suddenly desert some carefully and lovingly constructed straw position for a closely related one, after which we have to look back and reread closely for the clues that his endorsement was never wholehearted. For example, much of the latter half of the book is devoted to showing that language and cognition are not adaptive. Yet, on page 79, we find: "On the other hand, I do think there are reasons why adaptionism should be true of a cognitive architecture in so far as it is (massively or otherwise) modular". And Fodor is best known for the claim that cognition is indeed modular.

None of this makes for clear argument, and one reason is that Fodor is discussing topics well understood and much discussed by AI, but his now well-established style prevents him discussing the issues in GOFAI (Good Old Fashioned AI) terms, although those terms are what links all this endless speculation about mind to some kind of test and implementation. Fodor, give or take a few details, clearly shares a lot of the representational and computational theory of mind shared by most GOFAIers (i.e., not connectionists!). Yet Fodor affects total ignorance of their work, the late David Marr being the only AI reference in the book, apart of course from the late Alan Turing. Is the only good AIer a dead AIer for our author?

How can this be? Part of the problem is Fodor's willful ignorance of anything to do with AI or computation in practice: he gloatingly refers to the AI frame problem and how it has held up any achievement in AI. But that is utter nonsense, as is: " . . . the failure of artificial intelligence to produce successful simulation of routine commonsense cognitive competences is notorious, not to say scandalous. We still don't have the fabled machine that can make breakfast without burning down the house; or the one that can translate everyday English into everyday Italian, or the one that can summarize texts . . . " (p. 37).

There is no market for the breakfast machine yet but the robots that build Fiat cars, often seen on television, are at least as capable and do not burn down the factory. Again, there is less of a market for English to Italian, but any major Italian newspaper passage on a web site can be translated adequately into English on demand by one of two major MT systems, usually for free. As to summarization, he should check the DARPA summarization competitions or the British Telecom summarizer, free on their website until recently and still sometimes turned on for demos.

Fodor's ignorance must be studied not casual: it is perhaps all that insulates him from the necessity of using the dominant GOFAI discourse, and allows him to keep his own unique form in being, since it is not philosophy, nor psychology as normally understood: it is perhaps uniquely Cognitive Science, but at the price of being connected to no well understood norms of evaluation, argument, experiment or implementation.

I once heard Ed Feigenbaum stand up and ask, of an early Cognitive Science audience, what ideas cognitive psychology had ever had that it did not get from AI, and at the time I thought that a bit strong. After reading this book I am not so sure. Fodor launches at one point (p. 56) into a historiography of the term 'module', one he has made very much his own (1983) in some ways, but he includes a "usage proprietary to Noam Chomsky" (p. 57). Yet nowhere does he seem to realise modules were the common currency of AI in the early 1970s, particularly at MIT within the hierarchy/heterarchy discussion, and where Hewitt uttered his immortal "modules shouldn't dicker around with each others' insides", which is as good a definition as Fodor's fumblings for "information encapsulation". The whole module idea in computation and AI came from what was then called "the metaphysics of LISP", where the notion of module followed naturally from the syntax of the function language for representation and manipulation.

In Chapter 3, Fodor launches into a discussion of abductive inference and then the "frame problem". He seems to identify these as two forms of the same thing: as problems in "global" reasoning that cannot be accounted for by a "Turing style", or local, computational theory of mind, and which probably are not, for Fodor, computationally tractable at all. Nowhere is there a hint of the long history of the frame problem in AI (i.e., since McCarthy and Hayes, 1969), nor of computational models of abduction (Hobbs et al., 1990). I do not say any of these have come upon full solutions, but just ignoring them gives a feeling of being in a parallel Fodorian universe in which familiar topics in your field are discussed in a loose, inconclusive, way, usually without any clear statement of what the issue is: neither abduction nor the frame problem is defined or illustrated.

As I noted earlier, I was unable to follow the plot completely, but I think it is something like this: "Turing" stands for a shallow, local, computational theory of mind, one which probably corresponds to innate properties in the phenotype individual, and which is intimately connected to the syntactic properties of the representation of mental contents. In addition, there is the need for a real, rationalist and global theory of mind, and this comes down in some way to abduction (and solving the frame problem!) and this need probably cannot be met by any computational theory at all. But however bad Turing style theories are (which Fodor half believes, though he never actually says so), the main enemy theory is much worse, namely associationism/connectionism, since that cannot even do the easy local stuff properly. The close and more sophisticated enemy (the New Synthesis, alias Pinker and others) has attempted to weld the Turing style theory to Massive Modularity, which means roughly a module for every logical inference rule, and has linked this to an adaptive, Darwinian, style explanation of the evolution of mental processes and content, and this for Fodor is a sell-out by those once firmly in his own tradition.

There is so much in the way of weak premises, arbitrary choices and non-sequiturs here, it is hard to know what to say. Fodor's obsession with "syntactic" muddles together the linguistic syntax of his own "sentence in the head" style of representation to the abstract syntax (versus semantics) of any representation whatever, and which was closer to Turing's concerns. Fodor makes an early point about computing syntactic proofs in the propositional calculus, but fails to see that the semantics of the calculus (i.e., the truth tables) is in fact a much simpler computation, and that syntax and semantics cannot be contrasted in computational terms (as opposed to grounding terms, where the real problems are) in the way he wants.

Why he choosesto identify abduction as the core of rational, holistic, global, inference (as opposed to, say, deduction or any of a range of heuristics in computational proof) is a mystery to me; abduction is no more than a word for the heuristic location of premises to prove a given proposition, and its search can be run as locally or globally as you like. What he and others now call massive modularity is also exceedingly close to the demon-style representation and inference theories identified for decades with Allan Newell and still alive in the SOAR paradigm (e.g., Laird et al., 1987). It has had many successes and Fodor should at least look at it.

When he finally closes with the last enemy, the New Synthesis, he has only very thin arguments as to why adaptivity is not a serious topic for a theory of language and mind; he reports regretfully (p. 118) that even Chomsky is agnostic about massive modularity. His final objection is that adaptivity is not needed because real explanations are synchronic: " ... it's the usual case ... for an explanation that fits a phenomenon into the vast landscape of causation to be largely or solely ahistorical. ... Thus, for example, the aerodynamic explanation of how birds fly doesn't, in and of itself, tell you anything about how birds came to fly" (p. 82).

True, but there was never a theory seeking to constrain the set of all possible flights, as there has been for all possible languages, so that the question of whether earlier developmental forms of language were in such a set cannot be avoided. Later (p. 96), he writes that Chomsky is right to hold "both that human language is innate and modular, and that it is not an adaptation".

In the end it all comes back to Universal Grammar, and Fodor is determined not to give an inch on evolution: what the putative language organ believes is a "General Linguistic Theory" (p. 95), and "there is no particular need for what the language organ believes to have been shaped by natural selection" (p. 96). This is a view that would have played well in Tennessee in the 1930s, but seems a little odd now. The root justification seems to be that modules are "innate databases" (p. 96) and "data isn't useful unless it's true", and those useful needs and truths are that we can learn to communicate with each other because our theories of each other and our language are true. That this observation excludes evolution or adaptation seems to be both a massive non sequitur and utterly untrue to everything else we know about human beings: our "data bases" about sex, smiles, handshakes and so on all show such cooperative features, fortunately, but are not at all immune from evolutionary explanation.

What has heto offer by way of contrast? "Some radical reorganization of global cognitive structure must have occurred in the process of getting from [ape] minds to ours" (p. 97). All the old confusions and teases are here: is this really different from a Darwinian view or could it be nudged towards it, bit by bit? It is certainly not a proto-Creationist view of language origins, but more that of someone who cannot bear the random element inherent in the Darwinian story; Fodor refers disparagingly at one point to permuting names and numbers in phone directories, and Nineteenth Century bishops felt much the same about the evolution of the eye.

As I noted earlier, the book has not been well produced: Searle, a writer with much in common with Fodor, gets only a single footnote, but it is on page 110, not page 109 as the index has it.

## References

Craik, K. (1943). *The Nature of Explanation*, Cambridge: Cambridge University Press.

Dennett, D. (1995). *Darwin's Dangerous Idea*, New York: Simon and Schuster.

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

Hobbs, J., Stickel, M., Appelt, D., and Martin, P. (1990). 'Interpretation As Abduction', SRI Technical note 499. Menlo Park, CA: Stanford Research Institute.

Lieberman, P. (1984). *The Biology and Evolution of Language*, Cambridge, MA: Harvard University Press.

McCarthy, J. and Hayes, P. (1969). 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', in B. Meltzer and D. Michie, eds., *Machine Intelligence* 4, pp. 463–502, Edinburgh: Edinburgh University Press.

Laird, J., Newell, A. and Rosenbloom, P. (1987). 'Soar: An Architecture for General Intelligence', *Artificial Intelligence* 33(1), pp. 1–64.

Pinker, S. (1997). *How the Mind Works*, New York: Norton.
Turing, A. (1950). 'Computing Machinery and Intelligence', *Mind* 59, pp. 433–460.
Turing, A. (1936). 'On computable numbers', *Proc. Lond. Math. Soc.* 2nd Series, 42, pp. 230–265.

YORICK WILKS
*Department of Computer Science*
*University of Sheffield*
*Sheffield, S11 4DP, UK*
*E-mail: yorick@dcs.shef.ac.uk*

John von Neumann, *The Computer and the Brain*, 2nd edition, Mrs. Hepsa Ely Silliman Memorial Lectures, New Haven: Yale University Press, 2000, xxviii + 82 pp., $9.95 (paper), ISBN 0-300-084373-0.

When John von Neumann turned his interest to computers, he was one of the leading mathematicians of his time. In the 1940s, he helped design two of the first stored-program digital electronic computers. He authored reports explaining the functional organization of modern computers for the first time, thereby influencing their construction worldwide (von Neumann, 1945; Burks et al., 1946). In the first of these reports, von Neumann described the computer as analogous to a brain, with an input "organ" (analogous to sensory neurons), a memory, an arithmetical and a logical "organ" (analogous to associative neurons), and an output "organ" (analogous to motor neurons).

His experience with computers convinced him that brains and computers, both having to do with the processing of information, should be studied by a new discipline–automata theory. In fact, according to von Neumann, automata theory would cover not only computers and brains, but also any biological or artificial systems that dealt with information and control, including robots and genes. Von Neumann never formulated a full-blown mathematical theory of automata, but he wrote several important exploratory papers (von Neumann, 1951, 1956, 1966). Meanwhile, besides designing hardware, he developed some of the first programs, programming languages, programming techniques, and numerical methods for solving mathematical problems using computers. (Much of his work on computing is reprinted in Aspray and Burks, 1987.) Shortly before his death in 1956, he wrote an informal synthesis of his views about brains. Though von Neumann left his manuscript sketchy and unfinished, Yale University Press published it as *The Computer and the Brain* in 1958. The 2000 reprint of this small but informative book is an opportunity to learn, or be reminded of, von Neumann's thoughts on the computational organization of the mind-brain.

Von Neumann began by explaining computers, which for him were essentially number crunchers: to compute was "to operate on ... numbers according to a predetermined plan" (p. 3; all page references are to von Neumann's book). In any