This work originally appeared in:

Dennett, Daniel. 1984. Carving the Mind at Its Joints. *Contemporary Psychology: The APA Review of Books* 29(4): 285-286.

It is available online at:
http://psycnet.apa.org/index.cfm?fa=search.displayrecord&uid=2006-06411-005

# Carving the Mind at Its Joints

Jerry A. Fodor
The Modularity of Mind: An Essay on
Faculty Psychology
Canmbridge. MA: MIT Press, 1983.
1 15 pp. $17.50 cloth; $8.50 paper

Review by
Daniel C. Dennett

Jerry A. Fodor; professor of psychology at the Massachusetts Institute of Technology, is author of The Language of Thought. Daniel C. Dennett is professor of philosophy at Tufts University. His books include Brainstorms.

According to one brand of orthodoxy in cognitive psychology, the mind is rather like an apple: Inside the thin skin of "transducers" that form the interface between the mind and the rest of the world there are no very salient or important boundaries to be discerned as you move toward the center (and out the other, efferent side). Where the core begins or ends is almost a matter of the theorist's convenience; there is no sharp line between thinking and perceiving because the perceptual-analysis processes that begin just inboard of the transducers are information-utilizing, partially "topdown," inferential, hypothesis-testing processes that are only marginally less intellectual than the most central sorts of cogitation. Others have thought that the mind is rather more like a peach, with a sharp boundary between the central pit of thought and the surrounding layer of processes that "present the world to thought." In this fascinating monograph, Fodor argues that not only is there such a theoretically crucial boundary between "input systems" and the central forum, but the layer of input systems is itself segmented rather like an orange, "vertically" into "encapsulated" special-purpose modules. These modules are not the "five senses" but modules with much more interesting and unusual boundaries: the sentence-parsing module, the face-recognizing module, the primal-sketch drawing module (Marr & Nishihara, 1978; Marr & Poggio, 1977), for instance.

How interesting or important is this claim? It all depends, as Fodor fully realizes, on how sharply and convincingly he can draw the boundaries unrecognized by orthodoxy. So this is an essay about boundaries, and it is almost always crucial to its success that Fodor accentuate and emphasize boundaries-not only between parts of the mind, but also between theoretical positions held about the mind. For there are two natural ways to dismiss such a thesis: "You have just defined your pet boundaries into existence" and

"Those are real boundaries, but we have always known about them."

Fodor forestalls the latter retort in a lengthy first chapter, "Four Accounts of Mental Structure," in which he presents a somewhat tendentious and idealized taxonomy of distinct (and warring) theoretical positions and attitudes. The chapter is full of insights, but in the author's zeal to leave no view unbranded, it gives off the weird incense of religious war: The True Faith of the Neocartesian is enunciated at length, the Four-Point Creed of the Associationist ("of either the classical mentalist or more recent learning-theoretic" persuasion) is formulated (on p. 27), and we are told, for instance, that "environmentalist biases provide a main motivation for the computational associationist's constructivism" (p. 35). Those who have little faith in refutation-by-classification ("But that's just a variety of -ism!") will take this chapter's many lessons with a grain of salt. But in it Fodor does show that the orthodox views have denied, or at least ignored, the boundaries he is trying to draw.

Rebutting the former charge-that Fodor has simply defined his boundaries into existence-is the burden of the bulk of the monograph. Modular input analyzer systems are, Fodor claims, a "natural kind" recognizable by a cluster of properties. A typical module is

1. "domain specific": The system specializes in processing input on a particular topic or in a particular special domain-such as faces or sentences.
2. "innately specified": Its structure is minimally shaped by any sort of learning process.
3. "hardwired": It is implemented in localized, special-purpose neural systems, not in "relatively equipotential" neural mechanisms.
4. "autonomous": It operates in a self-contained way and does not share cognitive resources-of memory, analysis machinery, or data bases-with other cognitive systems.
5. "not assembled" (a sharpening of point 3): It is not constructed from a stock of more elementary processes; its "virtual architecture map[s] relatively directly onto its neural implementation" (p. 37).

Further typical features follow from these. The activity of a module is recognizable by its being quite automatic or involuntary; it is insensitive to central goals. "You can't hear speech as noise even if you would prefer to" (p. 53).

Moreover, by virtue of its autonomy it must operate in a strongly bottom-up or data-driven manner, because it has scant access to "higher" information that might arguably be of use to it.

The support for these claims is, as it should be, empirical, not a priori. Fodor cites widely from the literature-mostly experimental-in psychology, psycholinguistics, the neurosciences, and artificial intelligence to support his case. I found the support impressive, but I do not have the professional expertise to vouch for the soundness of Fodor's coverage of the relevant fields. Unlike Fodor's earlier books, this one is primarily for psychologists, not philosophers, and for that reason perhaps a psychologist would have been a more appropriate reviewer to tell psychologists how seriously they should take it.

Having made the positive case for input modules, Fodor argues that the "central systems" fed by the modules are vastly different-global, nonautomatic, assembled, the product of learning-and totally untouched to date by any serious theory. Here philosophical implications loom large, but Fodor's presentation is somewhat hurried and unconvincing. It is hard to argue against someone who declares that no one really knows anything about X, and Fodor's rather Romantic pessimism is not a very constructive contribution to the issue; but surely he is closer to the truth than are the wilder optimists of artificial intelligence.

A weak link in Fodor's case is the rather blurry boundary drawn separating the modules from the central arena of thought. For instance, is the output of the language module supposed to be full comprehension of an utterance, or merely some sort of perceptual identification of the utterance? (As anyone who has sat through a boring lecture knows, there is nothing remotely automatic, swift, mandatory, and effortless about understanding the utterances you cannot help but perceive.) Fodor's discussions of this issue (pp. 91-93) and of the related question about when the visual perception modules turn over their results to the global central processes of "belief fixation" (pp. 93-97 and passim) run the risk of winning his point by definition. To any claim about "psychological set" or "theory-leadenness" or "expectation-driven" effects, Fodor can always reply that these are effects on central perceptual belief-fixation but not on the (more peripheral) outputs of the modules-but only at the price of raising our suspicions about the integrity of those output boundaries.

Another weak link is Fodor's insistence that the activity within the modules is "computational." He needs to distinguish his modules from what might be called mere transducers, but the very features that make the input systems modular weaken the grounds for insisting that their operations are properly characterized as computations in any interestingly strong sense. In particular, the utility of a programming-language perspective drops close to zero when one is dealing with hard-wired, nonassembled, special-purpose mechanisms.

Fodor is fond of seeing ironies in the vicissitudes of theory development. Perhaps, then, he actually relishes the ironic implication of his new book: Since cognitive science has made progress only on the relatively peripheral modules, leaving the central systems still totally mysterious, and since only central systems require-if any do-a "language of thought," there is not yet any detailed scientific support arising from cognitive science for the existence of a language of thought.

**References**

Marr, D., & Nishihara, H. (1978, October). Visual information processing: Artificial intelligence and the sensorium of sight. Technology Review, pp. 28-49.

Marr, D., & Poggio, T. (1977). From understanding computation to understanding neural circuitry. Neurosciences Research Progress Bulletin, 15, 470-488.