Printed in the United States of America

Précis of The Modularity of Mind

Jerry A. Fodor

Department of Psychology, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Abstract: The Modularity of Mind proposes an alternative to the "New Look" or "interactionist" view of cognitive architecture that has dominated several decades of cognitive science. Whereas interactionism stresses the continuity of perceptual and cognitive processes, modularity theory argues for their distinctness. It is argued, in particular, that the apparent plausibility of New Look theorizing derives from the failure to distinguish between the (correct) claim that perceptual processes are inferential and the (dubious) claim that they are unencapsulated, that is, that they are arbitrarily sensitive to the organism's beliefs and desires. In fact, according to modularity theory, perceptual processes are computationally isolated from much of the background knowledge to which cognitive processes have access. The postulation of autonomous, domain-specific psychological mechanisms underlying perceptual integration connects modularity theory with the tradition of faculty psychology, in particular, with the work of Franz Joseph Gall. Some of these historical affinities, and some of the relations between faculty psychology and Cartesianism, are discussed in the book.

Keywords: Cartesianism; cognition; faculty psychology; interactionism; language; modularity; neuropsychology; perception; phrenology

Everybody knows that something is wrong. But it is uniquely the achievement of contemporary philosophy – indeed, it is uniquely the achievement of contemporary analytical philosophy – to have figured out just what it is. What is wrong is that not enough distinctions are being made. If only we made all the distinctions that there are, then we should all be as happy as kings. (Kings are

notoriously very happy.)

The Modularity of Mind (henceforth Modularity) is a monograph much in the spirit of that diagnosis. I wanted to argue there (and will likewise argue here) that modern Cognitivism failed, early on, to notice a certain important distinction: roughly, a distinction between two ways in which computational processes can be "smart." Because it missed this distinction, Cognitivism failed to consider some models of mental architecture for which a degree of empirical support can be marshaled, models that may, indeed, turn out to be true. If these models are true, then standard accounts of the nature of cognition and perception — and of the relations between them — are seriously misled, with consequences that can be felt all the way from artificial intelligence to epistemology. That was my story, and I am going to stick to it.

"What," you will ask, "was this missed distinction; who missed it; and how did missing it lead to these horrendous consequences?" I offer a historical reconstruction in the form of a fairy tale. None of what follows actually happened, but it makes a good story and has an edifying

moral.

So then: Once upon a time, there was a Wicked Behaviorist. He was, alas, a mingy and dogmatic creature of little humor and less poetry; but he did keep a clean attic. Each day, he would climb up to his attic and throw things out, for it was his ambition eventually to have almost nothing in his attic at all. (Some people whispered that this was his only ambition, that the Wicked Behav-

iorist was actually just a closet Ontological Purist. For all I know, they were right to whisper this.)

Anyhow, one day when the Wicked Behaviorist was upstairs cleaning out his attic, the following Very Interesting Thought occurred to him. "Look," he said to himself, "I can do without perceptual processes." (Because he had been educated in Vienna, the Wicked Behaviorist usually thought in the formal mode. So what actually occurred to him was that he could do without a theory of perceptual processes. It comes to much the same thing.) "For," it continued to occur to him, "perceptual identification reduces without residue to discriminative responding. And discriminative responding reduces without residue to the manifestation of conditioned (as it might be, operant) reflexes. And the theory of conditioned reflexes reduces without residue to Learning Theory. So, though learning is one of the things that there are, perceptual processes are one of the things there aren't. There also aren't: The True, or The Beautiful, or Santa Claus, or Tinkerbell; and unicorns are metaphysically impossible and George Washington wore false teeth. So there. Grrr!" He really was a very Wicked Behaviorist.

Fortunately, however, in the very same possible world in which the WB eked out a meager existence as a value of a bound variable (for who would call that living?), there was also a Handsome Cognitivist. And whereas the WB had this preference for clean attics and desert landscapes, the HC's motto was: "The more the merrier, more or less!" It was the HC's view that almost nothing reduces to almost anything else. To say that the world is so full of a number of things was, he thought, putting it mildly; for the HC, every day was like Christmas in Dickens, ontologically speaking. In fact, far from wishing to throw old things out, he was mainly interested in turning new things up. "Only collect," the HC was often heard to say.

Above all – and this is why I'm telling you this story – the HC wanted mental processes in general, and perceptual processes in particular, to be part of his collection.

Moreover, the HC had an argument. "Perceptual processes," he said, "can't be reflexes because, whereas reflexes are paradigmatically dumb, perceptual processes are demonstrably smart. Perception is really a part of cognition; it involves a kind of thinking."

"And what demonstrates that perceptual processes are

smart?" grumbled the Wicked Behaviorist.

"I will tell you," answered the Handsome Cognitivist. "What demonstrates that perceptual processes are smart is Poverty of The Stimulus Arguments." [A Poverty of The Stimulus Argument alleges that there is typically more information in a perceptual response than there is in the proximal stimulus that prompts the response; hence perceptual integration must somehow involve the contribution of information by the perceiving organism. [See Chomsky: "Rules and Representations" BBS 3(1) 1980.] No one knows how to quantify the relevant notion of information, so it is hard to show conclusively that this sort of argument is sound. On the other hand, such. phenomena as the perceptual constancies have persuaded almost everybody - except Gibsonians and Wicked Behaviorists [see Ullman: "Against Direct Perception" BBS 3(3) 1980, and Rachlin: "Pain and Behavior," this issue) - that Poverty of The Stimulus Arguments have to be taken very seriously. I shall assume, in what follows, that that is so.] "Poverty of The Stimulus Arguments," continued the HC, "show that perceptual identifications can't be reflexive responses to proximal stimulus invariants. In fact, Poverty of The Stimulus Arguments strongly suggest that perceptual identifications depend on some sort of computations, perhaps on computations of quite considerable complexity. So, once we have understood the force of Poverty of The Stimulus Arguments, we see that there probably are perceptual processes after all." "And," the HC added in a rush, "I believe that there are Truth and Beauty and Santa Claus and Tinkerbell too (only you have to read the existential quantifier leniently). And I believe that for each drop of rain that falls / A flower is born. So there." (Some people whispered that the Handsome Cognitivist, though he was very handsome, was perhaps just a little wet. For all I know, they were right to whisper that, too.) End of fairy tale.

My point is this: Modern Cognitivism starts with the use of Poverty of The Stimulus Arguments to show that perception is smart, hence that perceptual identification can't be reduced to reflexive responding. However – and I think this is good history and not a fairy tale at all – in their enthusiasm for this line of argument, early Cognitivists failed to distinguish between two quite different respects in which perceptual processes might be smarter than reflexes. Or, to put it the other way around, they failed to distinguish between two respects in which perception might be similar to cognition. It's at precisely this point that Modularity seeks to insert its wedge.

Reflexes, it is traditionally supposed, are dumb in two sorts of ways: They are noninferential and they are encapsulated. To say that they are noninferential is just to say that they are supposed to depend on "straight-through" connections. On the simplest account, stimuli elicit reflexive responses directly, without mediating mental processing. It is my view that the HC was right about

perceptual processes and reflexive ones being different in this respect; Poverty of The Stimulus Arguments do make it seem plausible that a lot of inference typically intervenes between a proximal stimulus and a perceptual identification.

By contrast, to describe reflexes as encapsulated is to say that they go off largely without regard to the beliefs and utilities of the behaving organism; to a first approximation, all that you need do to evoke a reflex is to present the appropriate eliciting stimulus. Here's how Modu-

larity put this point:

Suppose that you and I have known each other for many a long year . . . and you have come fully to appreciate the excellence of my character. In particular, you have come to know perfectly well that under no conceivable circumstances would I stick my finger in your eye. Suppose that this belief of yours is both explicit and deeply felt. You would, in fact, go to the wall for it. Still, if I jab my finger near enough to your eyes, and fast enough, you'll blink. . . . [The blink reflex] has no access to what you know about my character or, for that matter, to any other of your beliefs, utilities [or] expectations. For this reason the blink reflex is often produced when sober reflection would show it to be uncalled for. . . . (p. 71)

In this respect reflexes are quite unlike a lot of "higher cognitive" behavior, or so it would certainly seem. Chess moves, for example, aren't elicited willy-nilly by presentations of chess problems. Rather, the player's moves are determined by the state of his utilities (is he trying to win? or to lose? or is he, perhaps, just fooling around?) and by his beliefs, including his beliefs about the current state of the game, his beliefs about the structure of chess and the likely consequences of various patterns of play, his beliefs about the beliefs and utilities of his opponent, his beliefs about the beliefs of his opponent about his beliefs and utilities, and so on up through ever so many orders of intentionality.

So, then, cognition is smart in two ways in which reflexes are dumb. Now the question arises: What is perception like in these respects? Modularity offers several kinds of arguments for what is, really, a main thesis of the book: Although perception is smart like cognition in that it is typically inferential, it is nevertheless dumb like reflexes in that it is typically encapsulated. Perhaps the most persuasive of these arguments - certainly the shortest - is one that adverts to the persistence of perceptual illusions. The apparent difference in length of the Mueller-Lyer figures, for example, doesn't disappear when one learns that the arrows are in fact the same size. It seems to follow that at least some perceptual processes are insensitive to at least some of one's beliefs. Very much wanting the Mueller-Lyer illusion to go away doesn't make it disappear either; it seems to follow that at least some perceptual processes are insensitive to at least some of one's utilities. The ecological good sense of this arrangement is surely self-evident. Prejudiced and wishful seeing makes for dead animals.

This sort of point seems pretty obvious; one might wonder how Cognitivist enthusiasm for "top down," "cognitively penetrated" perceptual models managed to survive in face of it. I think we have already seen part of the answer: Cognitivists pervasively confused the question about the encapsulation of perception with the ques-

tion about its computational complexity. Because they believed – rightly – that Poverty of The Stimulus Arguments settled the second question, they never seriously considered the issues implicit in the first one. You can actually see this confusion being perpetrated in some of the early Cognitivist texts. The following passage is from Bruner's "On Perceptual Readiness":

Let it be plain that no claim is being made for the utter indistinguishability of perceptual and more conceptual inferences. . . . I may know that the Ames distorted room that looks so rectangular is indeed distorted, but unless conflicting cues are put into the situation . . . the room still looks rectangular. So too with such compelling illusions as the Mueller-Lyer: In spite of knowledge to the contrary, the line with the extended arrowheads looks longer than the equallength line with arrowheads inclined inward. But these differences, interesting in themselves, must not lead us to overlook the common feature of inference underlying so much of cognitive activity. (Bruner 1973, p. 8; emphasis added)

The issue raised by the persistence of illusion is not, however, whether some inferences are "more conceptual" than others – whatever, precisely, that might mean. Still less is it whether perception is in some important sense inferential. Rather, what's at issue is: How rigid is the boundary between the information available to cognitive processes and the information available to perceptual ones? How much of what you know/believe/desire actually does affect the way you see? The persistence of illusion suggests that the answer must be: "at most, less than all of it."

So far, my charge has been that early Cognitivism missed the distinction between the inferential complexity of perception and its cognitive penetrability. But, of course, it's no accident that it was just that distinction that Cognitivists confused. Though they are independent properties of computational systems, inferential complexity and cognitive penetrability are intimately related – so intimately that, unless one is very careful, it's easy to convince oneself that the former actually entails the latter. [For discussion see Pylyshyn: "Computation and Cognition" BBS 3(1) 1980.]

What connects inferential complexity and cognitive penetrability is the truism that inferences need premises. Here's how the argument might seem to go: Poverty of The Stimulus Arguments show that the organism must contribute information to perceptual integrations; "perceptual inferences" just are the computations that effect such contributions. Now, this information that the organism contributes - the premises, as it were, of its perceptual inferences - must include not just sensory specifications of current proximal inputs but also "background knowledge" drawn from prior experience or innate endowment; for what Poverty of The Stimulus Arguments show is precisely that sensory information alone underdetermines perceptual integrations. But, surely, the availability of background knowledge to processes of perceptual integration is the cognitive penetration of perception. So if perception is inferentially elaborated, it must be cognitively penetrated. Q.E.D.

What's wrong with this argument is that it depends on what one means by cognitive penetration. One might mean the availability to perceptual integration of some

information not given in the proximal array. Because Poverty of The Stimulus Arguments show that some such information must be available to perceptual integration, it follows that to accept Poverty of The Stimulus Arguments is to accept the cognitive penetrability of perception in this sense. But one might also mean by the cognitive penetrability of perception that anything that the organism knows, any information that is accessible to any of its cognitive processes, is ipso facto available as a premise in perceptual inference. This is a much more dramatic claim; it implies the continuity of perception with cognition. And, if it is true, it has all sorts of interesting epistemic payoff (see Fodor 1984). Notice, however, that this stronger claim does not follow from the inferential complexity of perception.

Why not? Well, for the following boring reason. We can, in principle, imagine three sorts of architectural arrangements in respect of the relations between cognition and perception: no background information is available to perceptual integration; some but not all background information is available to perceptual integration. Because Poverty of The Stimulus Arguments imply the inferential elaboration of perception, and because inferences need premises, the first of these architectures is closed to the Cognitivist. But the second and third are still open, and the persistence of illusions is prima facie evidence that the second is the better bet.

We arrive, at last, at the notion of a psychological module. A module is (inter alia) an informationally encapsulated computational system - an inference-making mechanism whose access to background information is constrained by general features of cognitive architecture, hence relatively rigidly and relatively permanently constrained. One can conceptualize a module as a specialpurpose computer with a proprietary database, under the conditions that: (a) the operations that it performs have access only to the information in its database (together, of course, with specifications of currently impinging proximal stimulations); and (b) at least some information that is available to at least some cognitive process is not available to the module. It is a main thesis of Modularity that perceptual integrations are typically performed by computational systems that are informationally encapsulated in this sense.

Modularity has two other main theses, which I might as well tell you about now. The first is that, although informational encapsulation is an essential property of modular systems, they also tend to exhibit other psychologically interesting properties. The notion of a module thus emerges as a sort of "cluster concept," and the claim that perceptual processes are modularized implies that wherever we look at the mechanisms that effect perceptual integration we see that this cluster of properties tends to recur. The third main thesis is that, whereas perceptual processes are typically modularized - hence encapsulated, hence stupid in one of the ways that reflexes are - the really "smart," really "higher" cognitive processes (thinking, for example) are not modular and, in particular, not encapsulated. So Modularity advocates a principled distinction between perception and cognition in contrast to the usual Cognitivist claims for their continuity.

Since Modularity goes into all of this in some detail, I

don't propose to do so here; otherwise, why would you buy the book? But I do want to stress the plausibility of the picture that emerges. On the one hand, there are the perceptual processes; these tend to be input driven, very fast, mandatory, superficial, encapsulated from much of the organism's background knowledge, largely organized around bottom-to-top information flow, largely innately specified (hence ontogenetically eccentric), and characteristically associated with specific neuroanatomical mechanisms (sometimes even with specific neuroanatomical loci). They tend also to be domain specific, so that - to cite the classic case - the computational systems that deal with the perception/production of language appear to have not much in common with those that deal with, for example, the analysis of color or of visual form (or, for that matter, the analysis of nonspeech auditory signals). So strikingly are these systems autonomous that they often rejoice in their proprietary, domain-specific pathologies: compare the aphasias and agnosias. Modularity takes the view that it is high time to praise Franz Joseph Gall for having predicted the existence of psychological mechanisms that exhibit this bundle of properties. (Gall was approximately a contemporary of Jane Austen's, so you see how far we have come in cognitive psychology - and in the novel, for that matter.) It is precisely in the investigation of these "vertical faculties" that modern Cognitivism has contributed its most important insights, and Modularity suggests that this is no accident. Precisely because the perceptual mechanisms are encapsulated, we can make progress in studying them without having to commit ourselves about the general nature of the cognitive mind.

On the other hand, there are the true higher cognitive faculties. So little is known about them that one is hardput even to say which true higher cognitive faculties there are. But "thought" and "problem solving" are surely among the names in the game, and here Modularity's line is that these are everything that perception is not: slow, deep, global rather than local, largely under voluntary (or, as one says, "executive") control, typically associated with diffuse neurological structures, neither bottom-totop nor top-to-bottom in their modes of processing, but characterized by computations in which information flows every which way. Above all, they are paradigmatically unencapsulated; the higher the cognitive process, the more it turns on the integration of information across superficially dissimilar domains. Modularity assumes that in this respect the higher cognitive processes are notably similar to processes of scientific discovery indeed, that the latter are the former writ large. Both, of course, are deeply mysterious; we don't understand nondemonstrative inference in either its macrocosmic or its microcosmic incarnation.

If much of the foregoing is right, then mainstream Cognitive science has managed to get the architecture of the mind almost exactly backwards. By emphasizing the continuity of cognition with perception, it missed the computational encapsulation of the latter. By attempting to understand thinking in terms of a baroque proliferation of scripts, plans, frames, schemata, special-purpose heuristics, expert systems, and other species of domain-specific intellectual automatisms – jumped-up habits, to put it in a nutshell – it missed what is most characteristic, and most puzzling, about the higher cognitive mind: its

nonencapsulation, its creativity, its holism, and its passion for the analogical. One laughs or weeps according to one's temperament. It was, perhaps, Eeyore who found precisely the right words: "Pathetic,' he said, 'That's what it is, pathetic.'"

Well, yes, but is much of this right? I want at least to emphasize its plausibility from several different points of view. Perception is above all concerned with keeping track of the state of the organism's local spatiotemporal environment. Not the distant past, not the distant future, and not - except for ecological accidents like stars - what is very far away. Perception is built to detect what is right here, right now - what is available, for example, for eating or being eaten by. If this is indeed its teleology, then it is understandable that perception should be performed by fast, mandatory, encapsulated, . . . etc. systems that - considered, as it were, detection-theoretically - are prepared to trade false positives for high gain. It is, no doubt, important to attend to the enternally beautiful and to believe the eternally true. But it is more important not to be eaten.

Why, then, isn't perception even stupider, even less inferential than it appears to be? Why doesn't it consist of literally reflexive responses to proximal stimulations? Presumably because there is so much more variability in the proximal projections that an organism's environment offers to its sensory mechanisms than there is in the distal environment itself. This kind of variability is by definition irrelevant if it is the distal environment that you care about - which, of course, it almost always is. So the function of perception, from this vantage point, is to propose to thought a representation of the world from which such irrelevant variability has been effectively filtered. What perceptual systems typically "know about" is how to infer current distal layouts from current proximal stimulations: the visual system, for example, knows how to derive distal form from proximal displacement, and the language system knows how to infer the speaker's communicative intentions from his phonetic productions. Neither mechanism, on the present account, knows a great deal else, and that is entirely typical of perceptual organization. Perceptual systems have access to (implicit or explicit) theories of the mapping between distal causes and proximal effects. But that's all they have.

If the perceptual mechanisms are indeed local, stupid, and extremely nervous, it is teleologically sensible to have the picture of the world that they present tempered, reanalyzed, and - as Kant saw - above all integrated by slower, better informed, more conservative, and more holistic cognitive systems. The purposes of survival are, after all, sometimes subserved by knowing the truth. The world's deep regularities don't show in a snapshot, so being bullheaded, ignoring the facts that aren't visible on the surface - encapsulation in short - is not the cognitive policy that one wants to pursue in the long run. The surface plausibility of the Modularity picture thus lies in the idea that Nature has contrived to have it both ways, to get the best out of fast dumb systems and slow contemplative ones, by simply refusing to choose between them. That is, I suppose, the way that Nature likes to operate: "I'll have some of each" - one damned thing piled on top of another, and nothing in moderation, ever.

It will have occurred to you, no doubt, that Cognitivism could quite possibly have hit on the right doctrine, even if it did so for the wrong reasons. Whatever confusions may have spawned the idea that perception and cognition are continuous, and however plausible the encapsulation story may appear to be a priori, there is a lot of experimental evidence around that argues for the effects of background knowledge in perception. If the mind really is modular, those data are going to have to be explained away. I want to say just a word about this.

There are, pretty clearly, three conditions that an experiment has to meet if it is to provide a bona fide counter-instance to the modularity of a perceptual system.

1. It must, of course, demonstrate the influence of background information in some computation that the system performs. But, more particularly, the background information whose influence it demonstrates must be exogenous from the point of view of the module concerned. Remember, each module has its proprietary database; whatever information is in its database is ipso facto available to its computations. So, for example, it would be no use for purposes of embarrassing modularity theory to show that words are superior to nonwords in a speech perception task. Presumably, the language processing system has access to a grammar of the language that it processes, and a grammar must surely contain a lexicon. What words are in the language is thus one of the things that the language module can plausibly be assumed to know consonant with its modularity.

2. The effect of the background must be distinctively perceptual, not postperceptual and not a criterion shift. For example, it is of no use to demonstrate that utterances of "implausible" sentences are harder to process than utterances of "plausible" ones if it turns out that the mechanism of this effect is the hearer's inability to believe that the speaker could have said what it sounded like he said. No one in his right mind doubts that perception interacts with cognition somewhere. What's at issue in the disagreement between modularity theory and "New Look" Cognitivism (e.g., Bruner 1973) is the locus of this interaction. In practice, it usually turns out that the issue is whether the recruitment of background information in perception is predictive. Modularity theory says almost never; New Look Cognitivism says quite a lot of the time.

3. The cognitively penetrated system must be the one that shoulders the burden of perceptual analysis in normal circumstances, and not, for example, some backup, problem-solving type of mechanism that functions only when the stimulus is too degraded for a module to cope with. Therefore, it is of no use to show that highly redundant lexical items are easier to understand than less redundant ones when the speech signal is very noisy – unless, of course, you can also show that the perception of very noisy speech really is bona fide speech perception.

So far as I know, there is very little in the experimental literature that is alleged to demonstrate the cognitive penetration of perception that meets all three of these conditions (to say nothing of replicability). This isn't to claim that such experiments cannot be devised or that, if devised, they might not prove that New Look Cognitivism is right after all. I claim only that, contrary to the textbook story, the empirical evidence for the continuity of perception with cognition is not overwhelming when contemplated with a jaundiced eye. There is, in any event, something for laboratory psychology to do for the

next twenty years or so: namely, try to develop some designs subtle enough to determine who's right about all this

"But look," you might ask, "why do you care about modules so much? You've got tenure; why don't you take off and go sailing?" This is a perfectly reasonable question and one that I often ask myself. Answering it would require exploring territory that I can't get into here and raising issues that Modularity doesn't even broach. But roughly, and by way of striking a closing note: The idea that cognition saturates perception belongs with (and is, indeed, historically connected with) the idea in the philosophy of science that one's observations are comprehensively determined by one's theories; with the idea in anthropology that one's values are comprehensively determined by one's culture; with the idea in sociology that one's epistemic commitments, including especially one's science, are comprehensively determined by one's class affiliations; and with the idea in linguistics that one's metaphysics is comprehensively determined by one's syntax. All these ideas imply a sort of relativistic holism: because perception is saturated by cognition, observation by theory, values by culture, science by class, and metaphysics by language, rational criticism of scientific theories, ethical values, metaphysical world-views, or whatever can take place only within the framework of assumptions that - as a matter of geographical, historical, or sociological accident - the interlocutors happen to share. What you can't do is rationally criticize the framework.

The thing is: I hate relativism. I hate relativism more than I hate anything else, excepting, maybe, fiberglass powerboats. More to the point, I think that relativism is very probably false. What it overlooks, to put it briefly and crudely, is the fixed structure of human nature. (This is not, of course, a novel insight; on the contrary, the malleability of human nature is a doctrine that relativists are invariably much inclined to stress. See, for example, John Dewey in Human Nature and Conduct [1922].) Well, in cognitive psychology the claim that there is a fixed structure of human nature traditionally takes the form of an insistence on the the heterogeneity of cognitive mechanisms and on the rigidity of the cognitive architecture that effects their encapsulation. If there are faculties and modules, then not everything affects everything else; not everything is plastic. Whatever the All is, at least there is more than One of it.

These are, as you will have gathered, not issues to be decisively argued – or even perspicuously formulated – in the course of a paragraph or two. Suffice it that they seem to be the sorts of issues that our cognitive science ought to bear on. And they are intimately intertwined: surely, surely, no one but a relativist would drive a fiberglass powerboat.

Coming in our next installment: "Restoring Basic Values: Phrenology in an Age of License." Try not to miss it!

NOTES

I. See, for example, Cregory (1970, p. 30): "perception involves a kind of problem-solving; a kind of intelligence." For a more recent and comprehensive treatment that runs along the same lines, see Rock (1983).

2. I don't at all care whether these "traditional assumptions" about reflexes are in fact correct, or even whether they were traditionally assumed. What I want is an ideal type with which to compare perception and cognition.