Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

FACULTY PSYCHOLOGY is getting to be respectable again after centuries of hanging around with phrenologists and other dubious types. By faculty psychology I mean, roughly, the view that many fundamentally different kinds of psychological mechanisms must be postulated in order to explain the facts of mental life. Faculty psychology takes seriously the apparent heterogeneity of the mental and is impressed by such prima facie differences as between, say, sensation and perception, volition and cognition, learning and remembering, or language and thought. Since, according to faculty psychologists, the mental causation of behavior typically involves the simultaneous activity of a variety of distinct psychological mechanisms, the best research strategy would seem to be divide and conquer: first study the intrinsic characteristics of each of the presumed faculties, then study the ways in which they interact. Viewed from the faculty psychologist's perspective, overt, observable behavior is an interaction effect par excellence.

This monograph is about the current status of the faculty psychology program; not so much its evidential status (which I take to be, for the most part, an open question) as what the program is and where it does, and doesn't, seem natural to try to apply it. Specifically, I want to do the following things: (1) distinguish the general claim that there are psychological faculties from a particular version of that claim, which I shall call the *modularity thesis*; (2)

enumerate some of the properties that modular cognitive systems are likely to exhibit in virtue of their modularity; and (3) consider whether it is possible to formulate any plausible hypothesis about which mental processes are likely to be the modular ones. Toward the end of the discussion, I'll also try to do something by way of (4) disentangling the faculty/modularity issues from what I'll call the thesis of *Epistemic Boundedness*: the idea that there are endogenously determined constraints on the kinds of problems that human beings can solve, hence on the kinds of things that we can know.

I shall, throughout, limit my brief to the psychology of cognitive processes, that being the only kind of psychology that I know anything about. Even so, this is going to be a rather long and rambling story, a fault for which I apologize in advance. My excuse is that, though I think the revival of the faculty psychology program has been enormously helpful in widening the range of serious options for cognitive psychologists to pursue, and while I also think that some version of the modularity thesis is very likely to prove true, still the atmosphere in which recent discussions have taken place has been on the steamy side, and a number of claims have been run together that are—or so I'll argue—conceptually distinct and unequally plausible. Moreover, there is quite a lot of ground to cover. A proposed inventory of psychological faculties is tantamount to a theory of the structure of the mind. These are serious matters and call for due expatiation.

PART I
FOUR ACCOUNTS OF MENTAL STRUCTURE

Behavior is organized, but the organization of behavior is merely derivative; the structure of behavior stands to mental structure as an effect stands to its cause. So much is orthodox mentalist doctrine and will be assumed throughout the discussion on which we're now embarked: Canonical psychological explanations account for the organization of behavior by appealing to principles which, they allege, explicate the structure of the mind.

But whereof does the structure of the mind consist? Not, to be sure, the clearest of questions, but nonetheless a pregnant one. I

propose, in this section, to consider faculty psychology as one sort of answer that this question can plausibly receive. (Strictly speaking, I shall regard it as *two* sorts of answer, as will presently emerge.) The primary object of this exercise is to delineate the character of faculty theorizing by contrasting it with several alternative accounts of the mind. My way of carving up these options departs, in some respects, from what I take to be standard, and perhaps the eccentricities will edify. Anyhow, I should say at the start that the positions about to be surveyed need not be understood as mutually exclusive. On the contrary, the view ultimately espoused will be, in a number of respects, quite shamelessly eclectic.

I.1. Neocartesianism: the structure of the mind viewed as the structure of knowledge

As practically everybody knows, Descartes' doctrine of innate ideas is with us again and is (especially under Chomsky's tutelage) explicitly construed as a theory about how the mind is (initially, intrinsically, genetically) structured into psychological faculties or "organs." I am inclined to view this Cartesian revival as very nearly an unmixed blessing. However, I think it is important to distinguish the Neocartesian sort of faculty psychology from other, rather different versions of the doctrine with which it is easily confused and whose rhetoric it has tended to appropriate. In fact, most of this essay will defend a notion of psychological faculty that is rather different from Chomsky's "mental organ" construct, and of which Descartes himself would quite probably have disapproved. The following discussion is by way of sorting out some of these strands.

In a nutshell, the central Neocartesian claim is that "intrinsic (psychological) structure is rich . . . and diverse" (Chomsky, 1980, p. 3). This view is contrasted with all forms of Empiricism, by which it is "assumed that development is uniform across (cognitive) domains and that the intrinsic properties of the initial state (of the mind) are homogeneous and undifferentiated—an assumption found across a spectrum reaching from Skinner to Piaget (who differ on much else)" (ibid.). Issues about innateness will recur, in one or another aspect, through much of what follows. But for now, I want to put them slightly to one side and try to see what notion

PART III
INPUT SYSTEMS AS MODULES

The modularity of the input systems consists in their possession of most or all of the properties now to be enumerated. If there are other psychological systems which possess most or all of these properties then, of course, they are modular too. It is, however, a main thesis of this work that the properties in virtue of which input systems are modular are ones which, in general, central cognitive processes do not share.

III.1. Input systems are domain specific

Let's start with this: how many input systems are there? The discussion thus far might be construed so as to suggest an answer somewhere in the vicinity of six—viz., one for each of the traditional sensory/perceptual 'modes' (hearing, sight, touch, taste, smell) and one more for language. This is *not*, however, the intended doctrine; what is proposed is something much more in the spirit of Gall's bumps. I imagine that within (and, quite possibly, across)[13] the traditional modes, there are highly specialized computational mechanisms in the business of generating hypotheses about the distal sources of proximal stimulations. The specialization of these mechanisms consists in constraints either on the range of information they can access in the course of projecting such hypotheses, or in the range of distal properties they can project such hypotheses about, or, most usually, on both.

Candidates might include, in the case of vision, mechanisms for color perception, for the analysis of shape, and for the analysis of three-dimensional spatial relations.[14] They might also include quite narrowly task-specific 'higher level' systems concerned with the visual guidance of bodily motions or with the recognition of faces of conspecifics. Candidates in audition might include computational systems that assign grammatical descriptions to token utterances; or ones that detect the melodic or rhythmic structure of acoustic arrays; or, for that matter, ones that mediate the recognition of the *voices* of conspecifics. There is, in fact, some evidence for the domain specificity of several of the systems just enumerated, but I suggest the examples primarily by way of indicating the levels of grain at which input systems might be modularized.

What, then, are the arguments for the domain specificity of input systems? To begin with, there is a sense in which input systems are ipso facto domain specific in a way in which computational systems at large are not. This is, however, quite uninteresting, a merely semantic point. Suppose, for example, that the function of the mechanisms of visual perception is to map transduced patterns of retinal excitation onto formulas of some central computational code. Then it follows trivially that their computational domain *qua mechanisms of visual perception* is specific to the class of possible retinal outputs. Correspondingly, if what the language-processing mechanisms do is pair utterance tokens with central formulas, then their computational domains *qua mechanisms of language processing* must be whatever encodings of utterances the auditory transducers produce. In similar boring fashion, the psychological mechanisms that mediate the perception of cows are ipso facto domain specific *qua mechanisms of cow perception*.

From such truisms, it goes without saying, nothing useful follows. In particular, the modularity of a system cannot be inferred from this trivial kind of domain specificity. It is, for example, entirely compatible with the cow specificity of cow perception that the recognition of cows should be mediated by precisely the same mechanisms that effect the perception of language, or of earthquakes, or of three-masted brigantines. For example, all four could perfectly well be accomplished by one and the same set of horizontal faculties. The interesting notion of domain specificity, by contrast, is Gall's idea that there are distinct psychological mechanisms—*vertical* faculties— corresponding to distinct stimulus domains. It is this latter claim that's now at issue.

Evidence for the domain specificity of an input analyzer can be of a variety of different sorts. Just occasionally the argument is quite direct and the demonstrations correspondingly dramatic. For example, there are results owing to investigators at the Haskins Laboratories which strongly suggest the domain specificity of the perceptual systems that effect the phonetic analysis of speech. The claim is that these mechanisms are different from those which effect the perceptual analysis of auditory nonspeech, and the experiments show that how a signal sounds to the hearer does depend, in rather

startling ways, on whether the acoustic context indicates that the stimulus is an utterance. Roughly, the very same signal that is heard as the onset of a consonant when the context specifies that the stimulus is speech is heard as a "whistle" or "glide" when it is isolated from the speech stream. The rather strong implication is that the computational systems that come into play in the perceptual analysis of speech are distinctive in that they operate *only* upon acoustic signals that are taken to be utterances. (See Liberman et al., 1967; for further discussion, see Fodor, Bever, and Garrett, 1974).

The Haskins experiments demonstrate the domain specificity of an input analyzer by showing that only a relatively restricted class of stimulations can throw the switch that turns it on. There are, however, other kinds of empirical arguments that can lead to the same sort of conclusions. One that has done quite a lot of work for cognitive scientists goes like this: If you have an *eccentric* stimulus domain—one in which perceptual analysis requires a body of information whose character and content is specific to that domain—then it is plausible that psychological processes defined over that domain may be carried out by relatively special purpose computational systems. All things being equal, the plausibility of this speculation is about proportional to the eccentricity of the domain.

Comparing perceiving cows with perceiving sentences will help to show what's going on here. I really have no idea how cow perception works, but let's follow the fashions and suppose, for purposes of discussion, that we use some sort of prototype-plus-similarity-metric. That is, the perceptual recognition of cows is effected by some mechanism which provides solutions for computational problems of the form: how similar—how 'close'—is the distal stimulus to a prototypical cow? My point is that *if* that's the way it's done, then cow perception might be mediated by much the same mechanisms that operate in a large variety of other perceptual domains as well—in fact, in any domain that is organized around prototypes. This is because we can imagine a quite general computational system which, given a specification of a prototype and a similarity metric for an arbitrary domain of percepts, will then compute the relevant distance relations in that domain. It seems plausible, that is to say, that procedures for estimating the distance between an input and a perceptual prototype should have pretty much the same computational structure wherever they are encountered.

It is, however, most unlikely that the perceptual recognition of sentences should be mediated by such procedures, and that is because sentence tokens constitute a set of highly eccentric stimuli: All the available evidence suggests that the computations which sentence recognizers perform must be closely tuned to a complex of stimulus properties that is quite specific to sentences. Roughly, the idea is that the structure of the sentence recognition system is responsive to universal properties of language and hence that the system works only in domains which exhibit these properties.

I take it that this story is by now pretty well known. The argument goes like this: Consider the class of *nomologically possible human languages*. There is evidence that this class constitutes quite a small subset of the logically possible linguistic systems. In particular, the nomologically possible human languages include only the ones that satisfy a set of (contingent) generalizations known as the 'linguistic universals.' One way to find out something about what linguistic universals there are is by examining and comparing *actual* human languages (French, English, Urdu, or whatever) with an eye to determining which properties they have in common. Much work in linguistics over the last twenty-five years or so has pursued this strategy, and a variety of candidate linguistic universals have been proposed, both in phonology and in syntax.

It seems quite unlikely that the existence of these universals is merely fortuitous, or that they can be explained by appeal to historical affinities among the languages that share them or by appeal to whatever pragmatic factors may operate to shape communication systems. (By pragmatic factors, I mean ones that involve general properties of communication exchanges as such, including the utilities of the partners to the exchanges. So, for example, Putnam (1961) once suggested that there are grammatical transformations because communicative efficiency is served by the deletion of redundant portions of messages, etc.) The obvious alternative to such accounts is to assume that the universals represent biases of a species-specific language-learning system, and a number of proposals have been made about how, in detail, such systems might be pretuned. It is assumed, according to all these accounts, that the language-learning mechanisms 'know about' the universals and

operate only in domains in which the universals are satisfied. (For a review, see Pinker, 1979.)

Parity of argument suggests that a similar story should hold for the mechanisms of language *perception*. In particular, the perceptual system involved is presumed to have access to information about how the universals are realized in the language it applies to. The upshot of this line of thought is that the perceptual system for a language comes to be viewed as containing quite an elaborate theory of the objects in its domain; perhaps a theory couched in the form of a grammar of the language. Correspondingly, the process of perceptual recognition is viewed as the application of that theory to the analysis of current inputs. (For some recent work on the parsing of natural language, see Marcus, 1977; Kaplan and Bresnan, in press; and Frazier and Fodor, 1978. All these otherwise quite different approaches share the methodological framework just outlined.)

To come to the moral: Since the satisfaction of the universals is supposed to be a property that distinguishes sentences from other stimulus domains, the more elaborate and complex the theory of universals comes to be the more eccentric the stimulus domain for sentence recognition. And, as we remarked above, the more eccentric a stimulus domain, the more plausible the speculation that it is computed by a special-purpose mechanism. It is, in particular, very hard to see how a device which classifies stimuli in respect of distance from a prototype could be recruited for purposes of sentence recognition. The computational question in sentence recognition seems to be not "How far to the nearest prototype?" but rather "How does the theory of the language apply to the analysis of the stimulus now at hand?"

There are probably quite a lot of kinds of relatively eccentric stimulus domains—ones whose perceptual analysis requires information that is highly specific to the domain in question. The organization of sentence perception around syntactic and phonological information does not exhaust the examples even in the case of language. So, for a further example, it is often and plausibly proposed that the processes that mediate phone recognition must have access to an internal model of the physical structure of the vocal apparatus. The argument is that a variety of constancies in speech perception seem to have precisely the effect of undoing garble that its inertial properties produce when the vocal mechanism responds to the phonetic intentions of the speaker. If this hypothesis is correct, then phone recognition is quite closely tuned to the mechanisms of speech production (see note 13). Once again, highly tuned computations are suggestive of special-purpose processors. Analogous points could be made in other perceptual modes. Faces are favorite candidates for eccentric stimuli (see Yin, 1969, 1970; Carey, 1978); and as I mentioned above, Ullman's work has made it seem plausible that the visual recognition of three-dimensional form is accomplished by systems that are tuned to the eccentricities of special classes of rigid spatial transformations.

From our point of view, the crucial question in all such examples is: how good is the inference from the eccentricity of the stimulus domain to the specificity of the corresponding psychological mechanisms? I am, in fact, not boundlessly enthusiastic about such inferences; they are clearly a long way from apodictic. Chess playing, for example, exploits a vast amount of eccentric information, but nobody wants to postulate a chess faculty. (Well, *almost* nobody. It is of some interest that recent progress in the artificial intelligence of chess has been achieved largely by employing specialized hardware. And, for what it's worth, chess is notably one of those cognitive capacities which breeds prodigies; so it is a candidate for modularity by Gall's criteria if not by mine.) Suffice it, for the present to suggest that it is probably characteristic of many modular systems that they operate in eccentric domains, since a likely motive for modularizing a system is that the computations it performs are idiosyncratic. But the converse inference—from the eccentricity of the domain to the modularity of the system—is warranted by nothing stronger than the maxim: specialized systems for specialized tasks. The most transparent situation is thus the one where you have a mechanism that computes an eccentric domain and is also modular by independent criteria; the eccentricity of the domain rationalizes the modularity of the processor and the modularity of the processor goes some way towards explaining how the efficient computation of eccentric domains is possible.

III.2 The operation of input systems is mandatory

You can't help hearing an utterance of a sentence (in a language

you know) as an utterance of a sentence, and you can't help seeing a visual array as consisting of objects distributed in three-dimensional space. Similarly, mutatis mutandis, for the other perceptual modes: you can't, for instance, help feeling what you run your fingers over as the surface of an object.[15] Marslen-Wilson and Tyler (1981), discussing word recognition, remark that ". . . even when subjects are asked to focus their attention on the acoustic-phonetic properties of the input, they do not seem to be able to avoid identifying the words involved. . . . This implies that the kind of processing operations observable in spoken-word recognition are mediated by automatic processes which are obligatorily applied . . . (p. 327).

The fact that input systems are apparently constrained to apply whenever they can apply is, when one thinks of it, rather remarkable. There is every reason to believe that, in the general case, the computational relations that input systems mediate—roughly, the relations between transducer outputs and percepts—are quite remote. For example, on all current theories, it requires elaborate processing to get you from the representation of a proximal stimulus that the retina provides to a representation of the distal stimuli as an array of objects in space.[16] Yet we apparently have no choice but to take up this computational burden whenever it is offered. In short, the operation of the input systems appears to be, in this respect, inflexibly insensitive to the character of one's utilities. You can't hear speech as noise *even if you would prefer to.*

What you can do, of course, is choose not to hear it at all—viz., not attend.[17] In the interesting cases—where this is achieved without deactivating a transducer (e.g., by sticking your fingers in your ears)—the strategy that works best is rather tortuous: one avoids attending to *x* by deciding to concentrate on *y*, thereby taking advantage of the difficulty of concentrating on more than one thing at a time. It may be that, when this strategy is successful, the unattended input system does indeed get selectively 'switched off', in which case there is a somewhat pickwickian sense in which voluntary control over the operation of an input system is circuitously achieved. Or it may be that the unattended input systems continue to operate but lose their access to some central processes (e.g., to those that mediate storage and report). The latter account is favored, at least for the case of language perception, in light of a fair number of results which seem to show relatively high-level processing of the unattended channel in dichotic listening tasks (Lackner and Garrett, 1973; Corteen and Wood, 1972; Lewis, 1970). But since the experimental results in this area are not univocal, perhaps the most conservative claim is this: input analysis is mandatory in that it provides the *only* route by which transducer outputs can gain access to central processes; if transduced information is to affect thought at all, it must do so via the computations that input systems perform.

I suppose one has to enter a minor caveat. Painters, or so I'm told, learn a little to undo the perceptual constancies and thus to see the world in something like the terms that the retina must deliver—as a two-dimensional spread of color discontinuities varying over time. And it is alleged that phoneticians can be taught to hear their language as something like a sound-stream—viz., as something like what the spikes in the auditory nerves presumably encode. (Though, as a matter of fact, the empirical evidence that phoneticians are actually able to do this is equivocal; see, for example, Lieberman, 1965.) But I doubt that we should take these highly skilled phenomenological reductions very seriously as counterexamples to the generalization that input processes are mandatory. For one thing, precisely because they *are* highly skilled, they may tell us very little about the character of normal perceptual processing. Moreover, it is tendentious—and quite possibly wrong—to think of what painters and phoneticians learn to do as getting access to, as it were, raw transducer output. An at least equally plausible story is that what they learn is how to 'correct' perceptually interpreted representations in ways that compensate for constancy effects. On this latter view, "seeing the visual field" or "hearing the speech stream" are *super*sophisticated perceptual achievements. I don't know which of these stories is the right one, but the issue is clearly empirical and oughtn't to be prejudged.

Anyhow, barring the specialized achievements of painters and phoneticians, one simply cannot see the world under its retinal projection and one has practically no access to the acoustics of utterances in languages that one speaks. (You all know what Swedish and Chinese sound like; what does *English* sound like?) In this respect (and in other respects too, or so I'll presently argue) the input mechanisms approximate the condition often ascribed to reflexes: they are automatically triggered by the stimuli that they apply to. And this is true for both the language comprehension mechanisms and the perceptual systems traditionally so-called.

It is perhaps unnecessary to remark that it does *not* seem to be true for nonperceptual cognitive processes. We have only the narrowest of options about how the objects of perception shall be represented, but we have all the leeway in the world as to how we shall represent the objects of *thought*; outside perception, the way that one deploys one's cognitive resources, is, in general, rationally subservient to one's utilities. Here are some exercises that you can do if you choose: think of *Hamlet* as a revenge play; as a typical product of Mannerist sensibility; as a pot-boiler; as an unlikely vehicle for Greta Garbo. Think of sixteen different ways of using a brick. Think of an utterance of "All Gaul is divided into three parts" as an acoustic object. Now try *hearing* an utterance of "All Gaul is divided into three parts" as an acoustic object. Notice the difference.

No doubt there are *some* limits to the freedom that one enjoys in rationally manipulating the representational capacities of thought. If, indeed, the Freudians are right, more of the direction of thought is mandatory—not to say obsessional—than the uninitiated might suppose. But the quantitative difference surely seems to be there. There is, as the computer people would put it, "executive control" over central representational capacities; and intellectual sophistication consists, in some part, in being able to exert that control in a manner conducive to the satisfaction of one's goals—in ways, in short, that seem likely to get you somewhere. By contrast, perceptual processes apparently apply willy-nilly in disregard of one's immediate concerns. "I couldn't help hearing what you said" is one of those clichés which, often enough, expresses a literal truth; and it is what is *said* that one can't help hearing, not just what is *uttered*.

III.3. There is only limited central access to the mental representations that input systems compute

It is worth distinguishing the claim that input operations are mandatory (you can't but hear an utterance of a sentence *as* an utterance of a sentence) from the claim that what might be called 'interlevels' of input representation are, typically, relatively inaccessible to consciousness. Not only must you hear an utterance of a sentence as such, but, to a first approximation, you can hear it *only* that way.

What makes this consideration interesting is that, according to all standard theories, the computations that input systems perform typically proceed via the assignment of a number of intermediate analyses of the proximal stimulation. Sentence comprehension, for example, involves not only acoustic encoding but also the recovery of phonetic and lexical content and syntactic form. Apparently an analogous picture applies in the case of vision, where the recognition of a distal array as, say, a-bottle-on-a-table-in-the-corner-of-the-room proceeds via the recovery of a series of preliminary representations (in terms of visual frequencies and primal sketches inter alia. For a review of recent thinking about interlevels of visual representation, see Zucker, 1981).

The present point is that the subject doesn't have equal access to all of these ascending levels of representation—not at least if we take the criterion of accessibility to be the availability for explicit report of the information that these representations encode. Indeed, as I remarked above, the lowest levels (the ones that correspond most closely to transducer outputs) appear to be completely *in*-accessible for all intents and purposes. The rule seems to be that, even if perceptual processing goes from 'bottom to top' (each level of representation of a stimulus computed being more abstractly related to transducer outputs than the one that immediately preceded), still *access* goes from top down (the further you get from transducer outputs, the more accessible the representations recovered are to central cognitive systems that presumably mediate conscious report).

A plausible first approximation might be that only such representations as constitute the *final* consequences of input processing are fully and freely available to the cognitive processes that eventuate in the voluntary determination of overt behavior. This arrangement of accessibility relations is reasonable enough assuming, on the one hand, that the computational capacities of central cognitive systems are not inexhaustible in their ability to attend to impinging information and, on the other, that it is the relatively abstract products of input-processing that encode most of the news that we are likely to want to know. I said in section III.2 that the operation of input systems is relatively insensitive to the subject's

utilities. By contrast, according to this account, the architectural arrangements that govern exchanges of information between input systems and other mechanisms of cognition do reflect aspects of the organism's standing concerns.

The generalization about the relative inaccessibility of intermediate levels of input analysis is pretty rough, but all sorts of anecdotal and experimental considerations suggest that something of the sort is going on. A well known psychological party trick goes like this:

E: Please look at your watch and tell me the time.
S: (Does so.)
E: Now tell me, without looking again, what is the shape of the numerals on your watch face?
S: (Stumped, evinces bafflement and awe.) (See Morton, 1967)

The point is that visual information which specifies the shape of the numerals must be registered when one reads one's watch, but from the point of view of access to later report, that information doesn't take. One recalls, as it were, pure position with no shape in the position occupied. There are analogous anecdotes to the effect that it is often hard to remember whether somebody you have just been talking to has a beard (or a moustache, or wears glasses). Yet visual information that specifies a beard must be registered and processed whenever you recognize a bearded face. More anecdote: Almost nobody can tell you how the letters and numbers are grouped on a telephone dial, though you use this information whenever you make a phone call. And Nickerson and Adams (1979) have shown that not only are subjects unable to describe a Lincoln penny accurately, they also can't pick out an accurate drawing from ones that get it grossly wrong.

There are quite similar phenomena in the case of language, where it is easy to show that details of syntax (or of the choice of vocabulary) are lost within moments of hearing an utterance, only the gist being retained. (Which did I just say was rapidly lost? Was it the syntactic details or the details of syntax?) Yet it is inconceivable that such information is not registered somewhere in the comprehension process and, within limits, it is possible to enhance its recovery by the manipulation of instructional variables. (For edifying experiments, see Sachs, 1967; Wanner, 1968.)

These sorts of examples make it seem plausible that the relative inaccessibility of lower levels of input analysis is at least in part a matter of how priorities are allocated in the transfer of representations from relatively short- to relatively long-term memory.[18] The idea would be that only quite high-level representations are stored, earlier ones being discarded as soon as subsystems of the input analyzer get the goodness out of them. Or, more precisely, intermediate input representations, when not discarded, are retained only at special cost in memory or attention, the existence of such charges-for-internal-access being itself a prototypical feature of modular systems.

This is, no doubt, part of the story. Witness the fact that in tasks which minimize memory demands by requiring comparison of *simultaneously* presented stimuli, responses that are sensitive to stimulus properties specified at relatively low levels of representation are frequently *faster* than responses to properties of the sort that high-level representations mark. Here, then, the ordering of relative accessibility reverses the top-to-bottom picture proposed above. It may be worth a digression to review some relevant findings.

The classical experimental paradigm is owing to Posner (1978). S's are required to respond 'yes' to visually presented letter pairs when they are *either* font identical (t,t; T,T) *or* alphabetically identical (t,T; T,t). The finding is that when letters in a pair are presented *simultaneously*, response to alphabetically identical pairs that are also font identical is faster than response to pairs that are identical alphabetically but not in font. This effect diminishes asymptotically with increase in the interstimulus interval when the letters are presented *sequentially*.

A plausible (though not mandatory) interpretation is that the representation that specifies the physical shape of the impinging stimulus is computed earlier than representations that specify its alphabetic value. (At a minimum, *some* shape information must be registered prior to alphabetic value, since alphabetic value depends upon shape.) In any event, the fact that representations of shape can drive voluntary responses suggests that they must be available to central processes at *some* point in the course of S's interaction with the stimulus. And this suggests, in turn, that the inaccessibility of font- as compared with alphabetic-information over the relatively long term must be a matter of how memory is deployed rather

than of the intrinsic opacity of low-level representations to high-level processes. It looks as though, in these cases, the relative unavailability of lower levels of input analysis is primarily a matter of the way that the subsystems of the input processors interface with memory systems. It is less a matter of information being unconscious than of its being unrecalled. (See also Crowder and Morton, 1969.)

It is unlikely, however, that this is the whole story about the inaccessibility of interlevels of input analysis. For one thing, as was remarked above, some very low levels of stimulus representation appear to be absolutely inaccessible to report. It is, to all intents and purposes (i.e., short of extensive training of the subject) impossible to elicit voluntary responses that are selectively sensitive to subphonetic linguistic distinctions (or, in the case of vision, to parameters of the retinal projection of distal objects) even though we have excellent theoretical grounds for supposing that such information must be registered somewhere in the course of linguistic (/visual) processing. And not *just* theoretical grounds: *we can often show that aspects of the subject's behavior are sensitive to the information that he can't report.*

For example, a famous result on the psychophysics of speech argues that utterances of syllables may be indistinguishable despite very substantial differences in their acoustic structure *so long as these differences are subphonetic*. When, however, quantitatively identical acoustic differences happen to be, as linguists say, 'contrastive'—i.e., when they mark distinctions between phones—they will be quite discriminable to the subject; as distinguishable, say, as "ba" is from "pa". It appears, in short, that there is a perceptual constancy at work which determines, in a wide range of cases, that only such acoustic differences as have linguistic value are accessible to the hearer in discrimination tasks. (See Liberman, et al., 1967.) What is equally striking, however, is that these 'inaccessible' differences *do affect reaction times*. Suppose a/a and a/b are utterance pairs such that the members of the first pair are literally acoustically identical and the members of the second differ only in *non*contrastive acoustic properties—i.e., the acoustic distinction between a and b is subphonetic. As we have seen, it is possible to choose such properties so that the members of the a/b pair are perceptually indistinguishable (as are, of course, the members of the pair a/a).

Even so, in such cases reaction times to make the 'same' judgment for the a/a pair are reliably faster than reaction times to make the 'same' judgment for the a/b pair. (Pisoni and Tash, 1974.) The subject can't report—and presumably can't hear—the difference between signal a and signal b, but his behavior is sensitive to it all the same.

These kinds of cases are legion in studies of the constancies, and this fact bears discussion. The *typical* function of the constancies is to engender perceptual similarity in the face of the variability of proximal stimulation. Proximal variation is very often misleading; the world is, in general, considerably more stable than are its projections onto the surfaces of transducers. Constancies correct for this, so that in general percepts correspond to distal layouts *better than* proximal stimuli do. But, of course, the work of the constancies would be undone unless the central systems which run behavior were required largely to ignore the representations which encode *un*corrected proximal information. The obvious architectural solution is to allow central systems to access information engendered by proximal stimulation only *after* it has been run through the input analyzers. Which is to say that central processes should have free access only to the *outputs* of perceptual processors, interlevels of perceptual processing being correspondingly opaque to higher cognitive systems. This, I'm claiming, is the architecture that we in fact do find.

There appears, in short, to be a generalization to state about input systems as such. Input analysis typically involves *mediated* mappings from transducer outputs onto percepts—mappings that are effected via the computation of interlevels of representation of the impinging stimulus. These intermediate representations are sometimes absolutely inaccessible to central processes, or, in many cases, they are accessible at a price: you can get at them, but only by imposing special demands upon memory or attention. Or, to put it another way: To a first approximation, input systems can be freely queried by memory and other central systems only in respect of *one* of the levels of representation that they compute; and the level that defines this interface is, in general, the one that is most abstractly related to transduced representations. This claim, if true, is substantive; and if, as I believe, it holds for input systems at large, then that is another reason to believe that the construct *input system* subsumes a natural kind.

III.4. Input systems are fast

Identifying sentences and visual arrays are among the fastest of our psychological processes. It is a little hard to quantify this claim because of unclarities about the individuation of mental activities. (What precisely are the boundaries of the processes to be compared? For example, where does sentence (/scene) *recognition* stop and more central activities take over? Compare the discussion in section III.6, below.) Still, granting the imprecision, there are more than enough facts around to shape one's theoretical intuitions.

Among the simplest of voluntary responses are two-choice reactions (push the button if the *left*-hand light goes on). The demands that this task imposes upon the cognitive capacities are minimal, and a practiced subject can respond reliably at latencies on the low side of a quarter of a second. It thus bears thinking about that the recovery of semantic content from a spoken sentence can occur at speeds quite comparable to those achieved in the two-choice reaction paradigm. In particular, appreciable numbers of subjects can 'shadow' continuous speech with a quarter-second latency (shadowing is repeating what you hear as you hear it) and, contrary to some of the original reports, there is now good evidence that such 'fast shadowers' understand what they repeat. (See Marslen-Wilson, 1973.) Considering the amount of processing that must go on in sentence comprehension (unless all our current theories are *totally* wrongheaded), this finding is mind-boggling. And, mind-boggling or otherwise, it is clear that shadowing latency is an extremely conservative measure of the speed of comprehension. Since shadowing requires *repeating* what one is hearing, the 250 msec. of lag between stimulus and response includes not only the time required for the perceptual analysis of the message, but also the time required for the subject's integration of his verbalization.

In fact, it may be that the phenomenon of fast shadowing shows that the efficiency of language processing comes very close to achieving theoretical limits. Since the syllabic rate of normal speech is about 4 per second, the observed 250 msec. latency is compatible with the suggestion that fast shadowers are processing speech in syllable-length units—i.e., that the initiation of the shadower's response is commenced upon the identification of each syllable-length input. Now, work in the psychoacoustics of speech makes it look quite likely that the syllable is the shortest linguistic unit that *can* be reliably identified in the speech stream (see Liberman et al., 1967). Apparently, the acoustic realizations of shorter linguistic forms (like phones) exhibit such extreme context dependence as to make them unidentifiable on a unit-by-unit basis. Only at the level of the syllable do we begin to find stretches of wave form whose acoustic properties are at all reliably related to their linguistic values. If this is so, then it suggests the following profoundly depressing possibility: the responses of fast shadowers lag a syllable behind the stimulus *not* because a quarter second is the upper bound on the speed of the mental processes that mediate language comprehension, but rather because, if the subject were to go any faster, he would overrun the ability of the speech stream to signal linguistic distinctions.[19]

In the attempt to estimate the speed of computation of visual processing, problems of quantification are considerably more severe. On the one hand, the stimulus is not usually spread out in time, so it's hard to determine how much of the input the subject registers before initiating his identificatory response. And, on the other hand, we don't have a taxonomy of visual stimuli comparable to the classification of utterance tokens into linguistic types. Since the question what type a linguistic token belongs to is a great deal clearer than the corresponding question for visual arrays, it is even less obvious in vision than in speech what sort of response should count as indicating that a given array has been identified.

For all of which there is good reason to believe that given a motivated decision about how to quantify the observations, the facts about visual perception would prove quite as appalling as those about language. For example, in one study by Haber (1980), subjects were exposed to 2,560 photographic slides of randomly chosen natural scenes, each slide being exposed for an interval of 10 seconds. Performance on recognition recall (ability to correctly identify a test slide as one that had been seen previously) approached 90 percent one hour after the original exposure. Haber remarks that the results "suggest that recognition of pictures is essentially perfect." Recent work by Potter (personal communication) indicates that 10 seconds of exposure is actually a great deal more than subjects need to effect a perceptual encoding of the stimulus adequate to mediate this near-perfect performance. According to Potter, S's performance in the Haber paradigm asymptotes at an exposure interval of about 2 seconds per slide.

There are some other results of Potter's (1975) that make the point still more graphically. S is shown a sequence of slides of magazine photographs, the rate of presentation of the slides being the experimentally manipulated variable. Prior to each sequence, S is provided with a brief description of an object or event that may appear in one or another slide—e.g., a boat, two men drinking beer, etc. S is to attend to the slides, responding when he sees one that satisfies the description. Under these conditions, S's respond with better than 70 percent accuracy when each slide is exposed for 125 msec. Accuracy asymptotes (at around 96 percent) at exposure times of 167 msec. per slide. It is of some interest that S's are as good at this task as they are at recognition recall (i.e., at making the global judgment that a given slide is one that they have seen before).

Two first-blush morals should be drawn from such findings about the computational efficiency of input processes. First, it contrasts with the relative slowness of paradigmatic central processes like problem-solving; and, second, it is presumably no accident that these very fast psychologicical processes are mandatory.

The first point is, I suppose, intuitively obvious: one can, and often does, spend hours thinking about a problem in philosophy or chess, though there is no reason to suppose that the computational complexity of these problems is greater than that of the ones that are routinely solved effortlessly in the course of perceptual processing. Indeed, the puzzle about input analysis is precisely that the computational complexity of the problem to be solved doesn't seem to predict the difficulty of solving it; or, rather, if it does, the difference between a 'hard' problem and an 'easy' one is measured not in months but in milliseconds. This dissimilarity between perception and thought is surely so adequately robust that it is unlikely to be an artifact of the way that we individuate cognitive achievements. It is only in trick cases, of the sorts that psychologists devise in experimental laboratories, that the perceptual analysis of an utterance or a visual scene is other than effectively instantaneous. What goes on when you parse a standard psycholinguistic poser like "the horse raced past the barn fell" is, almost certainly, *not* the same sort of processing that mediates sentence recognition in the normal case. They even *feel* different.

Second, it may well be that processes of input analysis are fast *because* they are mandatory. Because these processes are automatic, you save computation (hence time) that would otherwise have to be devoted to deciding whether, and how, they ought to be performed. Compare: eyeblink is a fast response *because* it is a reflex—i.e., because you don't have to *decide* whether to blink your eye when someone jabs a finger at it. Automatic responses are, in a certain sense, deeply unintelligent; of the whole range of computational (and, eventually, behavioral) options available to the organism, only a stereotyped subset is brought into play. But what you save by indulging in this sort of stupidity is *not having to make up your mind*, and making your mind up takes time. Reflexes, whatever their limitations, are not in jeopardy of being sicklied o'er with the pale cast of thought. Nor are input processes, according to the present analysis.

There is, however, more than this to be said about the speed of input processes. We'll return to the matter shortly.

III.5. *Input systems are informationally encapsulated*

Some of the claims that I'm now about to make are in dispute among psychologists, but I shall make them anyway because I think that they are true. I shall run the discussion in this section largely in terms of language, though, as usual, it is intended that the morals should hold for input systems at large.

I remarked above that, almost certainly, understanding an utterance involves establishing its analysis at several different levels of representation: phonetic, phonological, lexical, syntactic, and so forth. Now, in principle, information about the probable structure of the stimulus at any of these levels could be brought to bear upon the recovery of its analysis at any of the others. Indeed, in principle *any* information available to the hearer, including meteorological information, astrological information, or—rather more plausibly—information about the speaker's probable communicative intentions could be brought to bear at any point in the comprehension process. In particular, it is entirely possible that, in the course of computing a structural description, information that is specified only at relatively high levels of representation should be 'fed back' to determine analyses at relatively lower levels.[20] But

though this is possible in principle, the burden of my argument is going to be that the operations of input systems are in certain respects unaffected by such feedback.

I want to emphasize the 'in certain respects'. For there exist, in the psychological literature, dramatic illustrations of the effects of information feedback upon some input operations. Consider, for example, the 'phoneme restoration effect' (Warren, 1970). You make a tape recording of a word (as it might be, the word "legislature") and you splice out one of the speech sounds (as it might be, the 's'), which you then replace with a tape recording of a cough. The acoustic structure of the resultant signal is thus /legi(cough)lature/. But what a subject will *hear* when you play the tape to him is an utterance of /legislature/ with a cough 'in the background'. It surely seems that what is going on here is that the perceived phonetic constituency of the utterance is determined not just by the transduced information (not just by information specified at *sub*-phonetic levels of analysis) but also by higher-level information about the probable lexical representation of the utterance (i.e., by the subject's guess that the intended utterance was probably /legislature/).

It is not difficult to imagine how this sort of feedback might be achieved. Perhaps, when the stimulus is noisy, the subject's mental lexicon is searched for a 'best match' to however much of the phonetic content of the utterance has been securely identified. In effect, the lexicon is queried by the instruction 'Find an entry some ten phones long, of which the initial phone sequence is /legi/ and the terminal sequence is /lature/.' The reply to this query constitutes the lexical analysis under which the input is heard.

Apparently rather similar phenomena occur in the case of visual scotoma (where neurological disorders produce a 'hole' in the subject's visual field). The evidence is that scotoma can mask quite a lot of the visual input without creating a phenomenal blind spot for the subject. What happens is presumably that information about higher-level redundancies is fed back to 'fill in' the missing sensory information. Some such process also presumably accounts for one's inability to 'see' one's retinal blind spot.

These sorts of considerations have led to some psychologists (and many theorists in AI) to propose relentlessly top-down models of input analysis, in which the perceptual encoding of a stimulus is determined largely by the subject's (conscious or unconscious) beliefs and expectations, and hardly at all by the stimulus information that transducers provide. Extreme examples of such feedback-oriented approaches can be found in Schank's account of language comprehension, in Neisser's early theorizing about vision, and in 'analysis by synthesis' approaches to sentence parsing. Indeed, a sentimental attachment to what are known generically as 'New Look' accounts of perception (Bruner, 1973) is pervasive in the cognitive science community. It will, however, be a main moral of this discussion that the involvement of certain sorts of feedback in the operation of input systems would be incompatible with their modularity, at least as I propose to construe the modularity thesis. One or other of these doctrines will have to go.

In the long run, which one goes will be a question of how the data turn out. Indeed, a great deal of the empirical interest of the modularity thesis lies in the fact that the experimental predictions it makes tend to be diametrically opposed to the ones that New Look approaches license. But experiments to one side, there are some prima facie reasons for doubting that the computations that input systems perform could have anything like unlimited access to high-level expectations or beliefs. These considerations suggest that even if there are *some* perceptual mechanisms whose operations are extensively subject to feedback, there must be others that compute the structure of a percept largely, perhaps solely, in isolation from background information.

For one thing, there is the widely noted persistence of many perceptual illusions (e.g., the Ames room, the phi phenomenon, the Muller-Lyre illusion in vision; the phoneme restoration and click displacement effects in speech) even in defiance of the subject's explicit knowledge that the percept is illusory. The very same subject who can tell you that the Muller-Lyre arrows are identical in length, who indeed has seen them measured, still finds one looking longer than the other. In such cases it is hard to see an alternative to the view that at least *some* of the background information at the subject's disposal is inaccessible to at least some of his perceptual mechanisms.

An old psychological puzzle provides a further example of this kind. When you move your head, or your eyes, the flow of images across the retina may be identical to what it would be were the head and eyes to remain stationary while the scene moves. So: why don't we experience apparent motion when we move our eyes? Most psychologists now accept one or other version of the "corollary discharge" answer to this problem. According to this story, the neural centers which initiate head and eye motions communicate with the input analyzer in charge of interpreting visual stimulations (See Bizzi, 1968). Because the latter system knows what the former is up to, it is able to discount alterations in the retinal flow that are due to the motions of the receptive organs.

Well, the point of interest for us is that this visual-motor system is informationally encapsulated. Witness the fact that, if you (gently) push your eyeball with your finger (as opposed to moving it in the usual way: by an exercise of the will), you *do* get apparent motion. Consider the moral: when you voluntarily move your eyeball with your finger, you certainly are possessed of the information that it's your eye (and not the visual scene) that is moving. This knowledge is absolutely explicit; if I ask you, you can *say* what's going on. But this explicit information, available to you for (e.g.) report, is *not* available to the analyzer in charge of the perceptual integration of your retinal stimulations. That system has access to corollary discharges from the motor center *and to no other information that you possess.* Modularity with a vengeance.

We've been surveying first blush considerations which suggest that at least some input analyzers are encapsulated with respect to at least some sorts of feedback. The next of these is a point of principle: feedback works only to the extent that the information which perception supplies is redundant; and it *is* possible to perceptually analyze arbitrarily unredundant stimulus arrays. This point is spectacularly obvious in the case of language. If I write "I keep a giraffe in my pocket," you are able to understand me despite the fact that, on even the most inflationary construal of the notion of context, there is nothing in the context of the inscription that would have enabled you to predict either its form or its content. In short, feedback is effective only to the extent that, *prior* to the analysis of the stimulus, the perceiver knows quite a lot about what the stimulus is going to be like. Whereas, the *point* of perception is, surely, that it lets us find out how the world is even when the world is some way that we *don't* expect it to be. The teleology of perceptual capacities presupposes a considerably-less-than-omniscient-organism; they'd be no use to God. If you already *know* how things are, why look to *see* how things are?[21]

So: The perceptual analysis of *un*anticipated stimulus layouts (in language and elsewhere) is possible only to the extent that (a) the output of the transducer is insensitive to the beliefs/expectations of the organism; and (b) the input analyzers are adequate to compute a representation of the stimulus from the information that the transducers supply. This is to say that the perception of novelty depends on bottom-to-top perceptual mechanisms.

There is a variety of ways of putting this point, which is, I think, among the most important for understanding the character of the input systems. Pylyshyn (1980) speaks of the "cognitive impenetrability" of perception, meaning that the output of the perceptual systems is largely insensitive to what the perceiver presumes or desires. Pylyshyn's point is that a condition for the *reliability* of perception, at least for a fallible organism, is that it generally sees what's there, not what it wants or expects to be there. Organisms that don't do so become deceased.

Here is another terminology for framing these issues about the direction of information flow in perceptual analysis: Suppose that the organism is given the problem of determining the analysis of a stimulus at a certain level of representation—e.g., the problem of determining which sequence of words a given utterance encodes. Since, in the general case, transducer outputs underdetermine perceptual analyses,[22] we can think of the solution of such problems as involving processes of nondemonstrative inference. In particular, we can think of each input system as a computational mechanism which projects and confirms a certain class of hyputheses on the basis of a certain body of data. In the present example, the available hypotheses are the word sequences that can be constructed from entries in the subject's mental lexicon, and the perceptual problem is to determine which of these sequences provides the right analysis of the currently impinging utterance token. The mechanism which solves the problem is, in effect, the realization of a *confirmation function*: it's a mapping which associates with each pair of a lexical hypothesis and some acoustic datum a value which expresses the degree of confirmation that the latter bestows upon the former. (And similarly, mutatis mutandis, for the nondemonstrative inferences that the other input analyzers effect.) I emphasize that construing the situation this way involves no commitment to a detailed theory of the operation of perceptual systems. *Any* nondemonstrative inference can be viewed as the projection and confirmation of a hypothesis, and I take it that perceptual inferences must in general be nondemonstrative, since their underdetermination by sensory data is not in serious dispute.

Looked at this way, the claim that input systems are informationally encapsulated is equivalent to the claim that the data that can bear on the confirmation of perceptual hypotheses includes, in the general case, considerably less than the organism may know. That is, the confirmation function for input systems does not have access to all of the information that the organism internally represents; there are restrictions upon the allocation of internally represented information to input processes.

Talking about the direction of information flow in psychological processes and talking about restrictions upon the allocation of information to such processes are thus two ways of talking about the same thing. If, for example, we say that the flow of information in language comprehension runs directly from the determination of the phonetic structure of an utterance to the determination of its lexical content, then we are saying that only phonetic information is available to whatever mechanism decides the level of confirmation of perceptual hypotheses about lexical structure. On that account, such mechanisms are encapsulated with respect to *non*phonetic information; they have no access to such information; not even if it is *internally represented, accessible to other cognitive processes* (i.e., to cognitive processes other than the assignment of lexical analyses to phone sequences) and *germane* in the sense that if it *were* brought to bear in lexical analysis, it would affect the confirmation levels of perceptual hypotheses about lexical structure.

I put the issue of informational encapsulation in terms of constraints on the data available for hypothesis confirmation because doing so will help us later, when we come to compare input systems with central cognitive processes. Suffice it to say, for the moment, that this formulation suggests another possible reason why input systems are so fast. We remarked above that the computations that input systems perform are mandatory, and that their being so saves time that would otherwise have to be used in executive decision-making. We now add that input systems are bull-headed and that this, too, makes for speed. The point is this: to the extent that input systems are informationally encapsulated, of all the information that might *in principle* bear upon a problem of perceptual analysis only a portion (perhaps only quite a small and stereotyped portion) is actually admitted for consideration. This is to say that speed is purchased for input systems by permitting them to ignore lots of the facts. Ignoring the facts is not, of course, a good recipe for problem-solving in the general case. But then, as we have seen, input systems don't function in the *general* case. Rather, they function to provide very special kinds of representations of very specialized inputs (to pair transduced representations with formulas in the domains of central processes). What operates in the general case, and what is sensitive, at least in principle, to *everything* that the organism knows, are the central processes themselves. Of which more later.

I should add that these reflections upon the value of bull-headedness do not, as one might suppose, entirely depend upon assumptions about the speed of memory search. Consider an example. Ogden Nash once offered the following splendidly sane advice: "If you're called by a panther/don't anther." Roughly, we want the perceptual identification of panthers to be very fast and to err, if at all, only on the side of false positives. If there is a body of information that must be deployed in such perceptual identifications, then we would prefer not to have to recover that information from a large memory, assuming that the speed of access varies inversely with the amount of information that the memory contains. This is a way of saying that we do not, on that assumption, want to have to access panther-identification information from the (presumably *very* large) central storage in which representations of background-information-at-large are generally supposed to live. Which is in turn to say that we don't want the input analyzer that mediates panther identification to communicate with the central store on the assumption that large memories are searched slowly.

Suppose, however, that random access to a memory is *insensitive* to its size. *Even so* panther-identification (and, mutatis mutandis, other processes of input analysis) had better be insensitive to much of what one knows. Suppose that we can get at *everything* we know about panthers *very fast*. We still have the problem of deciding, for each such piece of information retrieved from memory, *how much inductive confirmation it bestows upon the hypothesis that the presently observed black-splotch-in-the-visual-field is a panther*. The point is that in the rush and scramble of panther identification, there are many things I know about panthers whose bearing on the likely pantherhood of the present stimulus *I do not wish to have to consider*. As, for example, that my grandmother abhors panthers; that every panther bears some distant relation to my Siamese cat Jerrold J.; that there are no panthers on Mars; that there is an Ogden Nash poem about panthers . . . etc. Nor is this all; for, in fact, the property of being 'about panthers' is not one that can be surefootedly relied upon. Given enough context, practically everything I know can be construed as panther related; and, *I do not want to have to consider everything I know* in the course of perceptual panther identification. In short, the point of the informational encapsulation of input processes is not—or not solely—to reduce the memory space that must be searched to find information that is perceptually relevant. The primary point is to so restrict the number of *confirmation relations* that need to be estimated as to make perceptual identifications fast. (I am indebted to Scott Fahlman for raising questions that provoked the last two paragraphs.)[23]

The informational encapsulation of the input systems is, or so I shall argue, the essence of their modularity. It's also the essence of the analogy between the input systems and reflexes; reflexes are informationally encapsulated with bells on.

Suppose that you and I have known each other for many a long year (we were boys together, say) 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. To say, as we did above, that the blink reflex is mandatory is to say, inter alia, that it has no access to what you know about my character or, for that matter, to any other of your beliefs, utilities and expectations. For this reason the blink reflex is often produced when sober reflection would show it to be uncalled for; like panther-spotting, it is prepared to trade false positives for speed.

That is what it is like for a psychological system to be informationally encapsulated. If you now imagine a system that is encapsulated in the way that reflexes are, but also computational in a way that reflexes are not, you will have some idea of what I'm proposing that input systems are like.

It is worth emphasizing that being modular in this sense is not quite the same thing as being autonomous in the sense that Gall had in mind. For Gall, if I read him right, the claim that the vertical faculties are autonomous was practically equivalent to the claim that there are no horizontal faculties for them to share. Musical aptitude, for example, is autonomous *in that* judging musical ideas shares no cognitive mechanisms with judging mathematical ideas; remembering music shares no cognitive mechanisms with remembering faces; perceiving music shares no cognitive mechanisms with perceiving speech; and so forth.

Now, it is unclear to what extent the input systems *are* autonomous in *that* sense. We do know, for example, that there are systematic relations between the amount of computational strain that decoding a sentence places on the language handling systems and the subject's ability to perform simultaneous nonlinguistic tasks quickly and accurately. 'Phoneme monitor' (Foss, 1970) techniques, and others, can be used to measure such interactions, and the results suggest a picture that is now widely accepted among cognitive psychologists: Mental processes often compete for access to resources variously characterized as attention, short-term memory, or work-space; and the result of allocating such resources to one of the competing processes is a decrement in the performance of the others. How general this sort of interaction is is unclear in the present state of the art (for contrary cases, suggesting isolated work spaces for visual imagery on the one hand and verbal recall on the other, see Brooks, 1968). In any event, where such competition does obtain, it is a counterexample to autonomy in what I am taking to be Gall's understanding of that notion.[24]

On the other hand, we can think of autonomy in a rather different way from Gall's—viz., in terms of informational encapsulation. So, instead of asking what access language processes (e.g.) have to *computational resources* that other systems also share, we can ask what access they have to the *information* that is available to other systems. If we do look at things this way, then the question "how much autonomy?" is the same question as "how much constraint on information flow?" In a nutshell: one way that a system can be autonomous is by being encapsulated, by not having access to facts that other systems know about. I am claiming that, whether or not the input systems are autonomous in Gall's sense, they are, to an interesting degree, autonomous in this informational sense.

However, I have not yet given any arguments (except some impressionistic ones) to show that the input systems actually are informationally encapsulated. In fact, I propose to do something considerably more modest: I want to suggest some caveats that ought to be, but frequently aren't, observed in interpreting the sorts of data that have usually been alleged in support of the contrary view. I think that many of the considerations that have seemed to suggest that input processes are cognitively penetrable—that they are importantly affected by the subject's belief about context, or his background information, or his utilities—are, in fact, equivocal or downright misleading. I shall therefore propose several ground rules for evaluating claims about the cognitive penetrability of input systems; and I'll suggest that, when these rules are enforced, the evidence for 'New Look' approaches to perception begins to seem not impressive. My impulse in all this is precisely analogous to what Marr and Pogio say motivates their work on vision: ". . . to examine ways of squeezing the last ounce of information from an image before taking recourse to the descending influence of high-level interpretation on early processing" (1977, pp. 475–476).

(a) Nobody doubts that the information that input systems provide must somehow be reconciled with the subject's background knowledge. We sometimes know that the world can't really be the way that it looks, and such cases may legitimately be described as the correction of input analyses by top-down information flow. (This, ultimately, is the reason for refusing to identify input analysis with perception. The point of perception is the fixation of belief, and the fixation of belief is a *conservative* process—one that is sensitive, in a variety of ways, to what the perceiver already knows. Input analysis may be informationally encapsulated, but perception surely is not.) However, to demonstrate *that* sort of interaction between input analyses and background knowledge is not, in and of itelf, tantamount to demonstrating the cognitive penetrability of the former; you need also to show that the locus of the top-down effect is *internal* to the input system. That is, you need to show that the information fed back interacts with interlevels of input-processing and not merely with the final results of such processing. The penetrability of a system is, by definition, its susceptibility to top-down effects at stages *prior* to its production of output.

I stress this point because it seems quite possible that input systems specify only relatively shallow levels of representation (see the next section). For example, it is quite possible that the perceptual representation delivered for a token sentence specifies little more than the type to which the token belongs (and hence does *not* specify such information as the speech act potential of the token, still less the speech act performed by the tokening). If this is so, then data showing effects of the hearer's background information on, e.g., his estimates of the speaker's communicative intentions would *not* constitute evidence for the cognitive penetration of the presumptive language-comprehension module; by hypothesis, the computations involved in making such estimates would not be among those that the language-comprehension module *per se* performs. Similarly, mutatis mutandis, in the case of vision. There is a great deal of evidence for context effects upon certain aspects of visual object recognition. But such evidence counts for nothing in the present discussion unless there is independent reason to believe that these aspects of object recognition are part of visual input analysis. Perhaps the input system for vision specifies the stimulus only in terms of "primal sketches" (for whose cognitive impenetrability there is, by the way, some nontrivial evidence. See Marr and Nishihara (1978).) The problem of assessing the degree of informational encapsulation of input systems is thus not independent of the problem of determining how such systems are individuated and what sorts of representations constitute their outputs. I shall return to the latter issue presently; for the moment, I'm just issuing caveats.

(b) Evidence for the cognitive penetrability of some computational mechanism that does what input systems do is not, in and of itself, evidence for the cognitive penetrability of input systems.

To see what is at issue here, consider some of the kinds of findings that have been taken as decisively exhibiting the effects of background expectations upon language perception. A well known way of estimating such expectations is the use of the so-called Cloze procedure. Roughly, S is presented with the first *n* words of a sentence and is asked to complete the fragment. Favored completions (as, for example, "salt" in the case of the fragment "I have the pepper, but would you please pass the ----") are said to be "high Cloze" and are assumed to indicate what the subject would expect a speaker to say next if he had just uttered a token of the fragment. An obvious generalization allows the estimation of the Cloze value at each point in a sentence, thereby permitting experiments in which the average Cloze value of the stimulus sentences is a manipulated variable.

It is quite easy to show that relative Cloze value affects S's performance on a number of experimental tasks, and it is reasonable to infer from such demonstrations that whatever mechanisms mediate the performance of these tasks must have access to S's expectations about what speakers are likely to say, hence not just to the 'stimulus' (e.g., acoustic) properties of the linguistic token under analysis. (For an early review of the literature on redundancy effects in sentence processing, see Miller and Isard, 1963.) So, for example, it can be shown that the accuracy of S's perception of sentences heard under masking noise is intimately related to the average Cloze value of the sentences: high Cloze sentences can be understood under conditions of greater distortion than the perception of low Cloze sentences tolerates. (Similarly, high Cloze sentences are, in general, more easily remembered than low Cloze sentences; recognition thresholds for words that are high Cloze in a context are lower than those for words that are low Cloze in that context; and so forth.)

The trouble with such demonstrations, however, is that although they show that there exist *some* language-handling processes that have access to the hearer's expectations about what is likely to be said, they do *not* show that the input systems enjoy such access. For example, it might be argued that, in situations where the stimulus is acoustically degraded, the subject is, in effect, encouraged to guess the identity of the material that he can't hear. (Similarly, mutatis mutandis, in memory experiments where a reasonable strategy for the subject is to guess at such of the material as he can't recall.) Not surprisingly, in such circumstances, the subject's background information comes into play with measurable effect. The question, however, is whether the psychological mechanisms deployed in the slow, relatively painful, highly attentional process of reconstructing noisy or otherwise degraded linguistic stimuli are the same mechanisms which mediate the automatic and fluent processes of normal speech perception.

That this question is not merely frivolous is manifested by results such as those of Fishler and Bloom (1980). Using a task in which sentences are presented in clear, they found only a marginal effect of high Cloze on the recognition of test words, and such effects vanished entirely when the stimuli were presented at high rates. (High presentation rates presumably discourage guessing; guessing takes time.) By contrast, words that are 'semantically anomalous' in context showed considerable inhibition in comparison with neutral controls. This last finding is of interest because it suggests that at least some of the effects of sentence context in speech recognition must be, as psychologists sometimes put it, 'post-perceptual'. In our terminology, these processes must operate *after* the input system has provided a (tentative) analysis of the lexical content of the stimulus. The point is that even if the facilitation of redundant items is mediated by predictive, expectation-driven mechanisms, the inhibition of contextually anomalous items cannot be. It is arguable that, in the course of speech perception, one is forever making such predictions as that 'pepper' will occur in 'salt and ----; but surely one can't also be forever predicting that 'dog', 'tomorrow', and all the other anomalous expressions will *not* occur there.[25] The moral is: some processes which eventuate in perceptual identifications are, doubtless, cognitively penetrated. But this is compatible with the informational encapsulation of the input systems themselves. Some traditional enthusiasm for context-driven perceptual models may have been prompted by confusion on this point.

(c) The claim that input systems are informationally encapsulated must be very carefully distinguished from the claim that there is top-down information flow *within* these systems. These issues are very often run together, with consequent exaggeration of the well-groundedness of the case against encapsulation.

Consider, once again, the phoneme restoration effect. Setting aside the general caution that experiments with distorted stimuli provide dubious grounds for inferences about speech perception in clear, phoneme restoration provides considerable prima facie

evidence that phone identification has access to what the subject knows about the lexical inventory of his language. If this interpretation is correct, then phoneme restoration illustrates top-down information flow in speech perception. It does *not*, however, illustrate the cognitive penetrability of the language input system. To show that that system is penetrable (hence informationally unencapsulated), you would have to show that its processes have access to information that is not specified at any of the levels of representation that the language input system computes; for example, that it has generalized access to what the hearer knows about the probable beliefs and intentions of his interlocutors. If, by contrast, the 'background information' deployed in phoneme restoration is simply the hearer's knowledge of the words in his language, then that counts as top-down flow within the language module; on any remotely plausible account, the knowledge of a language includes knowledge of its lexicon.

The most recent work in phoneme restoration makes this point with considerable force. Samuel (1981) has shown that both information about the lexical inventory and 'semantic' information supplied by sentential context affect the magnitude of the phoneme restoration effect. Specifically, you get more restoration in words than in (phonologically possible) nonwords, and you get more restoration when a word is predictable in sentence context than when the context is neutral. This looks like the penetration of phone recognition by both lexical and 'background' information, but the appearance is misleading. In fact, Samuel's data suggest that, of the two effects, *only the former* is strictly perceptual, the latter operating in consequence of a response bias to report predictable words as intact. (Detection theoretically: the word/nonword difference affects d', whereas the neutral context/predictive context difference affects β.) As Samuel points out, the amount of restoration is inversely proportional to S's ability to distinguish the stimulus word with a phone missing from an undistorted token of the same type; and, on Samuel's data, this discrimination is actually *better* for items that are highly predictable in context than for items that aren't. Another case, in short, where what had been taken to be an example of context-driven prediction in perception is, in fact, an effect of the biasing of post-perceptual decision processes.

The importance of distinguishing cognitive penetration from intramodular effects can be seen in many other cases where predictive analysis in perception is demonstrable. It is, for example, probable (though harder to show than one might have supposed) that top-down processes are involved in the identification of the surface constituent structure of sentences (see Wright, 1982). For example, it appears that the identification of nouns is selectively facilitated in contexts like T A ------, the identification of verbs is selectively facilitated in contexts like T N -------, and so forth. Such facilitation indicates that the procedures for assigning lexical items to form classes have access to information about the general conditions upon the well-formedness of constituent structure trees.

Now, it is a question of considerable theoretical interest whether, and to what extent, predictive analysis plays a role in parsing; but this issue must be sharply distinguished from the question whether the parser is informationally encapsulated. Counterexamples to encapsulation must exhibit the sensitivity of the parser to information that is not specified internal to the language-recognition module, and constraints on syntactic well-formedness are paradigms of information that does *not* satisfy this condition. The issue is currently a topic of intensive experimental and theoretical inquiry; but as things stand I know of *no* convincing evidence that syntactic parsing is ever guided by the subject's appreciation of semantic context or of 'real world' background. Perhaps this is not surprising; there are, in general, so many syntactically different ways of saying the same thing that even if context allowed you to estimate the *content* of what is about to be said, that information wouldn't much increase your ability to predict its *form*.[26]

These questions about where the interacting information comes from (whether it comes from inside or outside the input system) take on a special salience in light of the following consideration: it is possible to imagine ways in which mechanisms *internal* to a module might contrive to, as it were, mimic effects of cognitive penetration. The operation of such mechanisms might thus invite overestimations of the extent to which the module has access to the organism's general informational resources. To see how this might occur, let's return to the question of contextual facilitation of word recognition; traditionally a parade case for New Look theorizing, but increasingly an area in which the data are coming to seem equivocal.

Here are the bare bones of an ingenious experiment of David Swinney's (1979; for further, quite similar, results, see Tannenhaus, Leirnau, and Seidenberg, 1979). The subject listens to a stimulus sentence along the lines of "Because he was afraid of electronic surveillance, the spy carefully searched the room for bugs." Now, we know from previous research that the response latencies for 'bugs' (say, in a word/nonword decision task) will be faster in this context, where it is relatively predictable, than in a neutral context where it is acceptable but relatively low Cloze. This seems to be—and is traditionally taken to be—the sort of result which demonstrates how expectations based upon an intelligent appreciation of sentential context can guide lexical access; the subject predicts 'bugs' before he hears the word. His responses are correspondingly accelerated whenever his prediction proves true. Hence, cognitive penetration of lexical access.

You can, or so it seems, gild this lily. Suppose that, instead of measuring reaction time for word/nonword decisions on 'bugs', you simultaneously present (flashed on a screen that the subject can see) a different word belonging to the same (as one used to say) 'semantic field' (e.g., 'microphones'). If the top-down story is right in supposing that the subject is using semantic/background information to predict lexical content, then 'microphones' is as good a prediction in context as 'bugs' is, so you might expect that 'microphones', too, will exhibit facilitation as compared with a neutral context. And so it proves to do. Cognitive penetration of lexical access with bells on, or so it would appear.

But the appearance is misleading. For Swinney's data show that if you test with 'insects' instead of 'microphones', you get the same result: facilitation as compared with a neutral context. Consider what this means. 'Bugs' has two paraphrases: 'microphones' and 'insects'. But though only one of these is contextually relevant, *both are contextually facilitated*. This looks a lot less like the intelligent use of contextual/background information to guide lexical access. What it looks like instead is some sort of associative relation among lexical forms (between, say, 'spy' and 'bug'); a relation pitched at a level of representation sufficiently superficial to be *insensitive* to the semantic content of the items involved. This possibility is important for the following reason: If facilitation is mediated by merely interlexical relations (and not by the interaction of background information with the semantic content of the item and its context), then the information that is exploited to produce the facilitation can be represented *in the lexicon; hence internal to the language recognition module*. And if that is right, then contextual facilitation of lexical access is *not* an argument for the cognitive penetration of the module. It makes a difference, as I remarked above, where the penetrating information comes from.

Let's follow this just a little further. Suppose the mental lexicon is a sort of connected graph, with lexical items at the nodes and with paths from each item to several others. We can think of accessing an item in the lexicon as, in effect, exciting the corresponding node; and we can assume that one of the consequences of accessing a node is that excitation spreads along the pathways that lead from it. Assume, finally, that when excitation spreads through a portion of the lexical network, response thresholds for the excited nodes are correspondingly lowered. Accessing a given lexical item will thus decrease the response times for items to which it is connected. (This picture is familiar from the work of, among others, Morton, 1969, and Collins and Loftus, 1975; for relevant experimental evidence, see Meyer and Schvaneveldt, 1971.)

The point of the model-building is to suggest how mechanisms internal to the language processor could mimic the effects that cognitive penetration would produce if the latter indeed occurred. In the present example, what mimics the background knowledge that (roughly) spies have to do with bugs is the existence of a connection betweeen the node assigned to the word 'spy' and the node assigned to the word 'bug'. Facilitation of 'bug' in spy contexts is affected by the excitation of such intralexical connections.

Why should these intralexical connections exist? Surely not just in order to lead psychologists to overestimate the cognitive penetrability of language-processing. In fact, if one works the other way 'round and assumes that the input systems are encapsulated, one might think of the mimicry of penetration as a way that the input processors contrive to make the best of their informational isolation. Presumably, what encapsulation buys is speed; and, as we remarked above, it buys speed at the price of unintelligence. It would, one supposes, take a lot of time to make reliable decisions about whether there is the kind of relation between spies and bugs that makes it on balance likely that the current token of 'spy' will be followed by a token of 'bug'. But that is precisely the kind of decision that the subject would have to make if the contextual facilitation of lexical access were indeed an effect of background knowledge interacting with the semantic content of the context. The present suggestion is that no such intelligent evaluation of the options takes place; there is merely a brute facilitation of the recognition of 'bug' consequent upon the recognition of 'spy'. The condition of this brute facilitation buying anything is that it should be possible, with reasonable accuracy, to mimic what one knows about connectedness *in the world* by establishing corresponding connections among entries in the mental lexicon. In effect, the strategy is to use the structure of interlexical connections to mimic the structure of knowledge. The mimicry won't be precise (a route from 'spy' to 'insect' will be generated as a by-product of the route from 'spy' to 'bug'). But there's no reason to doubt that it may produce savings over all.

Since we are indulging speculations, we might as well indulge this one: It is a standing mystery in psychology why there should be interlexical associations at all; why subjects should exhibit a reliable and robust disposition to associate 'salt' with 'pepper', 'cat' with 'dog', 'mother' with 'father', and so forth. In the heyday of associationism, of course, such facts seemed quite *un*mysterious; they were, indeed, the stuff of which the mental life was supposed to be made. On one account the utterance of a sentence was taken to be a chained response, and associations among lexical items were what held the links together. According to still earlier tradition, the postulation of associative connections between Ideas was to be the mechanism for reconstructing the notion of degree of belief. None of this seems plausible now, however. Belief is a matter (not of association but) of *judgment*; sentence production is a matter (not of association but) of *planning*. So, what on earth are associations *for*?

The present suggestion is that associations are the means whereby stupid processing systems manage to behave as though they were smart ones. In particular, interlexical associations are the means whereby the language processor is enabled to act as though it knows that spies have to do with bugs (whereas, in fact, it knows no such thing). The idea is that, just as the tradition supposed, terms for things frequently connected in experience become themselves connected in the lexicon. Such connection is *not* knowledge; it is not even judgment. It is simply the mechanism of the contextual adjustment of response thresholds. Or, to put the matter somewhat metaphysically, the formation of interlexical connections buys the synchronic encapsulation of the language processor at the price of its cognitive penetrability *across time*. The information one has about how things are related in the world is inaccessible to modulate lexical access; that is what the encapsulation of the language processor implies. But one's experience of the relations of things in the world does affect the structure of the lexical network—viz., by instituting connections among lexical nodes. If the present line of speculation is correct, these connections have a real, if modest, role to play in the facilitation of the perceptual analysis of speech. The traditional, fundamental, and decisive objection to association is that it is too stupid a relation to form the basis of a mental life. But stupidity, when not indulged in to excess, is a virtue in fast, peripheral processes; which is exactly what I have been supposing input processes to be.

I am not quite claiming that all the putative effects of information about background (context, etc.) on sentence recognition are artifacts of connections in the lexical network (though, as a matter of fact, such experimental attempts as I've seen to demonstrate a residual effect of context after interlexical/associative factors are controlled for strike me as not persuasive). I am claiming only that the possibility of such artifacts contaminates quite a lot of the evidence that is standardly alleged. The undoubted fact that "semantically" coherent text is relatively easy to process does not, in and of itself, demonstrate that the input system for language has access to what the organism knows about how the world coheres. Such experimental evidence as supported early enthusiasms for massively top-down perceptual models was, I think, sexy but inconclusive; and the possibility of a modular treatment of input processes provides motivation for its reconsideration. The situation would seem to be paradigmatically Kuhnian: the data look different to a jaundiced eye.

Consider the provenance of New Look theorizing. Cognitive psychologists in the '40s and '50s were faced with the proposal that perception is *literally* reflexive; for example, that the theory of perception is reducible without residue to the theory of discriminative operant response. It was natural and admirable in such circumstances to stress the 'intelligence' of perceptual integration. However, in retrospect it seems that the intelligence of perceptual integration may have been seriously misconstrued by those who were most its partisans.

In the ideal condition—one approached more frequently in the textbooks than *in rerum naturae*, to be sure—reflexes have two salient properties. They are computationally simple (the stimulus is "directly connected" to the response), and they are informationally encapsulated (see above). I'm suggesting that New Look theories failed to distinguish these properties. They thus assumed, wrongly, that the disanalogy between perceptual and reflexive processes consisted in the capacity of the former to access and exploit background information. From the point of view of the modularity thesis, this is a case of the right intuition leading to the wrong claim. Input systems *are* computationally elaborated. Their typical function is to perform inference-like operations on representations of impinging stimuli. Processes of input analysis are thus unlike reflexes in respect of the character and complexity of the operations that they perform. But this is quite compatible with reflexes and input processes being similar in respect of their informational encapsulation; in this latter respect, both of them contrast with "central processes"—problem-solving and the like—of which cognitive penetrability is perhaps the most salient feature, or so I shall argue below. To see that informational encapsulation and computational elaboration are compatible properties, it is only necessary to bear in mind that unencapsulation is the exploitation of information from *outside* a system; a computationally elaborated sytem can thus be encapsulated if it stores the information that its computations exploit. Encapsulation is a matter of foreign affairs; computational elaboration begins at home.

It may be useful to summarize this discussion of the informational encapsulation of input systems by comparing it with some recent, and very interesting, suggestions owing to the philosopher Steven Stich (1978). Stich's discussion explores the difference between belief and the epistemic relation that is alleged to hold between, for example, speaker/hearers and the grammar of their native language (the relation that Chomsky calls 'cognizing'). Stich supposes, for purposes of argument, that the empirical evidence shows that speakers *in some sense* 'know' the grammar of their native language; his goal is to say something about what that sense is.

Let us call the epistemic relation that a native speaker has to the grammar of his language *subdoxastic* belief.[27] Stich suggests that there are two respects in which subdoxastic beliefs differ from beliefs strictly so-called. In the first place, as practically everybody has emphasized, subdoxastic beliefs are *unconscious*. But, Stich adds, subdoxastic beliefs are also typically "inferentially unintegrated." The easiest way to understand what Stich means by this is to consider one of his examples.

> If a linguist believes a certain generalization to the effect that no transformation rule exhibits a certain characteristic, and if he comes to (nonsubdoxastically) believe a given transformation which violates the generalization, he may well infer that the generalization is false. But merely having the rule stored (in the way that we are assuming all speakers of the language do) does not enable the linguist to draw the inference. . . . Suppose that for some putative rule, you have come to believe that if r then Chomsky is seriously mistaken. Suppose further that, as it happens, r is in fact among the rules stored by your language processing mechanism. The belief along with the subdoxastic state will not lead to the belief that Chomsky is seriously mistaken. By contrast, if you believe (perhaps even mistakenly) that r, then the belief that Chomsky is seriously mistaken is likely to be inferred. [pp. 508–509]

Or, as Stich puts the argument at another point, "It is characteristic of beliefs that they generate further beliefs via inference. What is more, beliefs are inferentially promiscuous. Provided with a suitable set of supplementary beliefs, almost any belief can play a role in the inference to any other. . . . (However) subdoxastic states, as contrasted with beliefs, are largely inferentially isolated from the large body of inferentially integrated beliefs to which a subject has (conscious) access."

Now, as Stich clearly sees, the proposal that subdoxastic states are typically both unconscious and inferentially unintegrated raises a question—viz., *Why should these two properties co-occur?* Why should it be, to put it in my terminology, that subdoxastic states

are typically encapsulated with respect to the processes which affect the inferential integration of beliefs?

Notice that there is a kind of encapsulation that follows from unconsciousness: an unconscious belief cannot play a role as a premise in the sort of reasoning that goes on in the conscious drawing of inferences. Stich is, however, urging something more interesting than this trivial truth. Stich's claim is that subdoxastic beliefs are largely inaccessible even to *unconscious* mental processes of belief fixation. If this claim is true, the question does indeed arise why it should be so.

I want to suggest, however, that the question doesn't arise because, as a matter of act, subdoxastic beliefs are not in general encapsulated; or, to put it more precisely, they are not in general encapsulated *qua* subdoxastic. Consider, as counterexamples, one's subdoxastic views about inductive and deductive warrant; for example, one's subdoxastic acquiescence in the rule of *modus ponens*. On the sort of psychological theory that Stich has in mind, subdoxastic knowledge of such principles must be accessible to practically all mental processes, since practically all inferential processes exploit them in one way or another. One's subdoxastic beliefs about validity and confirmation are thus quite unlike one's subdoxastic beliefs about the rules of grammar; though both are unconscious, the former are paradigms of promiscuous and unencapsulated mental states. So the connection between unconsciousness and encapsulation cannot be *intrinsic*.

Nevertheless, I think that Stich is onto something important. For, though much unconscious information must be widely accessible to processes of fixation of belief, it is quite true that very many of the examples of unconscious beliefs for which there is currently good empirical evidence are encapsulated. This is because most of our current cognitive science is the science of input systems, and, as we have seen, *informational encapsulation is arguably a pervasive feature of such systems*. Input systems typically do not exchange subdoxastic information with central processes or with one another.

Stich almost sees this point. He says that "subdoxastic states occur in a variety of separate, special purpose cognitive systems" (p. 508). True enough; but they must also occur in integrated, general purpose systems (in what I'm calling "central" systems), assuming that much of the fixation of belief is both unconscious and subserved by inferential mechanisms of that kind. The point is: subdoxastic states are informationally encapsulated *only* insofar as they are states of special purpose systems (e.g., states of input analyzers). Practically all psychologically interesting cognitive states are unconscious; but it is only the beliefs accessible to modules that are subdoxastic by the second of Stich's criteria as well.

III.6. Input analyzers have 'shallow' outputs.

The question where to draw the line between observation and inference (in the psychological version, between perception and cognition) is one of the most vexed, and most pregnant, in the philosophy of science. One finds every opinion from the extreme 'foundationalist' view, which restricts observation to processes that issue in infallible introspective reports, to the recent revisionism which denies that the distinction is in any respect principled. (Hanson, 1958, for example, holds that a physicist can *see* that the cloud chamber contains a proton track in the same sense of 'see' that is operative when Smith sees that there's a spot on Jones' tie.) Sometimes the argument for this sort of view is based explicitly on accounts of perception borrowed from New Look psychology, which suggests that *all* perception is ineliminably and boundlessly theory laden; see Goodman (1978).

Philosophers have cared about the observation/inference distinction largely for epistemological reasons; what is (nondemonstratively) inferred is supposed to run an inductive risk from which what is observed is supposed to be free. And it has seemed important to some epistemologists that whatever count as the data statements of a science should be isolated from such risk, the idea being that unless some contingent truths are certain, no empirical theory can compel rational belief.

I am not myself much moved by the idea that inductive warrant is inherited upward in science from a base level of indubitable truths; and barring some such assumption, the philosophical problem of making the observation/theory distinction rigorous seems less consequent than was once supposed. However, the corresponding psychological problem of saying where perceptual processes interface with cognitive ones must be addressed by anyone who takes the postulation of modular input systems seriously. For

one thing, it is a point of definition that distinct functional components cannot interface *everywhere* on pain of their ceasing to be distinct. It is this consideration that flow-chart notation captures by drawing boxes around the processing systems it postulates. That only the inputs and outputs of functionally individuated systems can mediate their information exchanges is tautological.

Moreover, we have seen that the plausibility of claims for the informational encapsulation of an input system depends very much on how one draws the distinction between its *outputs* and its *interlevels* of representation. Since it is common ground that there must be *some* mental processes in which perception interacts with background knowledge and with utilities, the issue about informational encapsulation is whether such interactions take place *internal* to the input systems. But the question what is internal to a system, and the question what is to count as the output of the system, are patently two ways of asking the same thing.

In general, the more constrained the information that the outputs of perceptual systems are assumed to encode—the shallower their outputs, the more plausible it is that the computations that effect the encoding are encapsulated. If, for example, the visual analysis system can report only upon the shapes and colors of things (all higher-level integrations being post-perceptual) it is correspondingly plausible that all the information that system exploits may be represented internal to it. By contrast, if the visual system can deliver news about protons (as a psychologized version of the Hanson story would suggest), then the likelihood that visual analysis is informationally encapsulated is negligible. Chat about protons surely implies free access to quite a lot of what I have been calling 'background knowledge'.

In this section I want to make a few, highly speculative suggestions about how the outputs of the language and visual processors might be characterized—that is, about the level of representation at which these systems interface with central processes. I shall rely heavily on the assumptions that input computations are very fast, and that their outputs are typically phenomenologically salient (see above). Consonant with these assumptions, I shall argue that there are some reasonable proposals to make about how to distinguish visual and linguistic perception from the cognitive processes with which they interface. It turns out, however, that there is nothing *episte-*

mologically special about the levels of representation which constitute the outputs of the visual (/linguistic) processing mechanisms. So if, in the spirit of epistemology naturalized, one leaves it to psychologists to draw the observation/theory distinction, then, according to these proposals, there is nothing epistemologically interesting about that distinction. For example, it does *not* correspond to the distinction between what we infallibly know and what we merely justifiably surmise. This seems to me, if anything, to argue in favor of drawing the line where I propose to draw it; still this version of naturalized epistemology may strike some epistemologists as far too deflationary.

What representation of an utterance does the language input processor compute? Or, to put the question in the context of the preceding discussion, which phenomenologically accessible properties of an utterance are such that, on the one hand, their recovery is mandatory, fast, and relevant to the perceptual encoding of the utterance and, on the other, such that their recovery might be achieved by an informationally encapsulated computational mechanism? Clearly, there is a wide choice of properties of utterances that *could* be computed by computational systems whose access to background information is, in one way or another, interestingly constrained—the duration of the utterance, e.g. For all that, there is, in the case of language, a glaringly obvious galaxy of candidates for modular treatment—viz., those properties that utterances have in virtue of some or other aspects of their linguistic structure (where this means, mostly, grammatical and/or logical form). Making these notions clear is notoriously hard; but the relevant intuitions are easy enough to grasp.

Whether John's utterance of "Mary might do it, but Joan is above that sort of thing" is ironical, say, is a question that can't be answered short of using a lot of what you know about John, Mary, and Joan. Worse yet, there doesn't seem to be any way to say, in the general case, how much, or precisely what, of what you know about them might need to be accessed in making such determinations. *Maybe* an interestingly encapsulated system could reliably recognize the irony (sincerity, metaphoricalness, rhetoricalness, etc.) of utterances, but there are certainly no plausible proposals about how this might be so. It looks as though recognizing such properties of utterances is typically an exercise in "inference to the best explanation": given what I know about John, and about what John thinks about Mary and Joan, he *couldn't* have meant that literally . . . etc. These are, of course, precisely the sorts of inferences that you would *not* expect encapsulated systems to perform. The "best" explanation is the one you want to accept *all things considered*, and encapsulated systems are prohibited by definition from considering all things.

Compare the computational problems involved in the recognition of linguistic form. The idea here is that the grammatical and logical structure of an utterance is uniquely determined (or, more precisely, uniquely determined up to ambiguity) by its phonetic constituency; and its phonetic constituency is uniquely determined in turn by certain of its acoustic properties (mutatis mutandis, the linguistic properties of written tokens are uniquely determined by certain properties of their shapes). "Acoustic" properties, according to this usage, are ipso facto transducer-detectable; so an input system that has access to the appropriate transduced representations of an utterance knows everything about the utterance that it needs to know to determine which sentential type it is a token of and, probably, what the logical form of the utterance is.[28] In short, if you are looking for an interesting property of utterances that might be computed by rigidly encapsulated systems—indeed, a property that might even be computed by largely bottom-to-top processors—then the type-identity of the utterance, together, perhaps, with its logical form would seem to be a natural candidate.

It is thus worth stressing that type-identity and at least some aspects of logical form are phenomenologically salient and are patently recognized 'on line'; moreover, the computation of type-identity is clearly an essential part of the overall process of language comprehension. In the general case, you can't understand what the speaker has said unless you can at least figure out which sentence he has uttered.

Is there, then, an encapsulated analyzer for logical and grammatical form? All the arguments are indirect; but, for what it's worth, it's rather hard to see how some of the processes that recognize logical and grammatical form could be anything but encapsulated. Background information can be brought to bear in perceptual analysis only where the property that is recognized is, to some significant extent, redundant in the context of recognition. But, as we remarked above, there doesn't seem to be much redundancy between context variables and the *form* of an utterance, however much context may predict its *content*. Even if you know precisely what someone is going to say—in the sense of knowing precisely which proposition he is going to assert—the knowledge buys you very little in predicting the type/token relation for his utterance; there are simply too many linguistically different ways of saying the same thing.

It is not, therefore, surprising that the more extreme proposals for context-driven language recognizers do *not* generally proceed by using contextual information to identify grammatical relations. Instead, they proceed whenever possible directly from a lexical analysis to a "conceptual" analysis—one which, in effect, collapses across synonymous tokens regardless of their linguistic type. It is unclear to me whether such models are proposed as serious candidates for the explanation of human communicative capacities, though sometimes I fear that they may be. (See, e.g., Schank and Abelson, 1975; for experimental evidence that linguistic form continues to have its effect as semantic integration increases, precisely as one would expect if the recovery of logical syntactic form is mandatory, see Forster and Olberi, 1973.) To put the point in a nutshell: linguistic form recognition can't be context-driven because context doesn't determine form; if linguistic form is recognized at all, it must be by largely encapsulated processes.

So the present proposal is that the language-input system specifies, for any utterance in its domain, its linguistic and maybe its logical form. It is implicit in this proposal that it does no more than that[29]—e.g., that it doesn't recover speech-act potential (except, perhaps, insofar as speech-act potential may be correlated with properties of form, as in English interrogative word order). As I suggested, the main argument for this proposal is that, on the one hand, type/token relations surely must be computed in the course of sentence comprehension and, on the other, it is hard to see how anything much richer than type/token relations could be computed by an informationally encapsulated processor. All this comports with the strong intuition that while there could perhaps be an algorithm for parsing, there surely could not be an algorithm for estimating communicative intentions in anything like their full diversity. Arguments about what an author *meant* are thus able to be interminable in ways in which arguments about what he *said* are not.

This is all pretty loose. Most dicussions in linguistics and psycholinguistics have been primarily interested in establishing *minimal* conditions on the output of the sentence processor, e.g., by demonstrating that one or another level of linguistic representation is "psychologically real" and recovered on line. By contrast, the problem that arises in discussions of modularity is typically of the form: What is the *most* that an encapsulated processor should be supposed to compute? Which aspects of the input can plausibly be recognized without generalized appeal to background data? There is, however, one area of language research in which issues of this latter sort have been extensively discussed. It may be worth a brief recapitulation here, since it provides quite a clear illustration of what problems about determining the level of the perception/cognition interface are like.

Consider again the question of the vocabulary of an utterance (as opposed to its logicosyntactic form on the one hand and its propositional content on the other). Since I have assumed that input-processing yields type identifications, I am committed to the claim that the language processor delivers, for each input utterance, a representation which specifies its lexical constituents inter alia. (Utterances which differ in their lexical constituents are, of course, ipso facto distinct in type.) The present question is whether it is plausible to suppose that the language-input system provides still deeper representations at the lexical level.

A view that has been influential in both linguistics and psychology suggests that it does. According to this view, understanding an utterance involves recovering the *definitions* of such definable lexical items as it may contain. So, for example, understanding a token of "John is a bachelor" involves representing the utterance as containing a word that means *unmarried man*. Note that this is a claim about processes of *comprehension* and not, e.g., about inferential operations which may be applied to the internal representation of the utterance *after* it has been understood. It is thus natural to interpret the claim as implying that the recovery of definitions of lexical items takes place during input processing (viz., *interior* to the putative language module). We would thus expect, if the claim is true, that the recovery of definitional information should exhibit the typical properties of input processes: it should happen fast, it should be mandatory (insensitive to task demands), etc.

The alternative view is that the "surface" vocabulary of an utterance is preserved at the level of representation where the language processor interfaces with cognitive processes at large. There should thus be no level of analysis specified by the language-input system at which ". . . bachelor . . ." and ". . . unmarried man . . ." receive identical representations (though, of course, *post*comprehension inferential processes may indeed identify them as synonymous. One could imagine that such postcomprehension inferences might be mediated by the application of "meaning postulates" in something like the sense of Carnap (1960); for discussion, see Kintsch (1974), Fodor, Fodor and Garrett (1975).)

The currently available experimental evidence supports the latter view. (See Fodor et al., 1980.) In fact, so far as I know, there have been *no* convincing data in favor of the claim that representations of definitional content engage *any* sentence-comprehension process. The importance of imposing appropriate task demands in experimental tests of this claim can, however, hardly be overemphasized. There is, e.g., no doubt at all that definitionally related sentences tend to be conflated in experiments that require not just comprehension but recall as well. This is quite consonant with the view that memory is an inferential process par excellence (see Bartlett, 1932).

If these observations are correct, they strongly suggest that input-processing for language provides no semantic analysis "inside" lexical items. Or, to put it another way, the functionally defined level *output of the language processing module* respects such *structurally* defined notions as *item in the morphemic inventory of the language*. It is of primary importance to see that there is no a priori reason why this should be true.[30] That is, there is no a priori reason why the representations of utterances that are computed by fast, mandatory, informationally encapsulated, etc., etc., processes should constitute a representational level by *any* independent criteria. But, in the case of language at least, there is some a posteriori reason to believe that they do: on the one hand, there is strong evidence that such notions as *morphemic level* and *syntactic level* pick out coherent classes of representations; and, on the other, there are at least reasonable grounds for supposing that it is representations at these sorts of levels that the input system delivers.

By the way, the (presumptive) fact that the representations which

input systems recover constitute linguistic natural kinds is a strong argument that the concept *input process* itself picks out a natural kind. Suppose that the representations of utterances that are recovered by fast, informationally encapsulated, mandatory, etc. processes turned out to specify, e.g., the second phoneme of the third word of each utterance, the intonation contour of its last five syllables, and the definitions of all the words that it contains which begin with 'u'. Since this collection of properties has no theoretical interest whatever, we would be inclined to infer that there is, to that extent, nothing interesting about the class of psycholinguistic processes that are fast, mandatory, and informationally encapsulated. But, apparently, that is *not* the sort of thing that we find. What we find instead is that the fast, mandatory . . . etc. processes deliver representations of utterances which make perfectly good sense considered *as* representations of utterances; representations which specify, for example, morphemic constituency, syntactic structure, and logical form. This is just the sort of thing you would expect if the fast, mandatory . . . etc. processes form a system that is functionally relevant to language comprehension. In particular, it is just what you would expect if language comprehension is effected by the sort of system that I am calling a module.

If I am inclined to harp on these points, it is because the opposed view—that sentence-processing grades off insensibly into inference and the appreciation of context; into general cognition in short—is actually predominant in the field. (Especially on the West Coast, where gurus teach that the All is One.) Suffice it to say that the choice between these pictures is empirical—not a matter of taste—and that such evidence as is actually germane seems not unfavorable to the modularity view.

The preceding discussion provides a context for raising analogous issues about vision. If the modularity story is to be plausible here, the output of the visual processor must be reasonably shallow (it should not categorize visual stimuli in such terms as *proton trace*), and it must form a level of representation by some independent criterion—i.e., there should be interesting things to say about the output representations other than that they are, de facto, the kinds of representations that the visual processor puts out.

Moreover, various candidates that satisfy the shallowness test and the levels test must nevertheless be rejected on grounds of phenomenological inaccessibility.[31] I am thinking of such representations as Marr's 'primal', '2.5 D', and '3 D' sketch (Marr and Nishihara, 1978). Such representations are certainly shallow enough. Indeed, they would seem to be too shallow. If we accept them as defining visual processor outputs, we shall have to say that even object recognition is not, strictly speaking, a phenomenon of visual perception, since, at these levels of representation, only certain geometric properties of the stimulus are specified. But, surely, from the point of view of phenomenological accessibility, perception is above all the recognition of objects and events. Shallower systems of representation can therefore constitute only interlevels of input analysis. What, then, is its output?

One of the most interesting ideas in recent cognitive theorizing is that there is a level of 'basic' perceptual objects (or, to use a slightly less misleading terminology, of basic perceptual categories). This notion is explored extensively in Brown (1958) and in Rosch et al. (1976), but a quick presentation may make the point. Consider a category hierarchy like *poodle, dog, mammal, animal, physical object, thing*. Roughly, the following seems to be true of such sets of categories: they effect a taxonomy of objects at increasing levels of abstractness, such that a given entity may belong to any or all of them, and such that the potential extensions of the categories increase as you go up the hierarchy (there are, as it were, more possible dogs than possible poodles; more possible animals than possible dogs; and so forth). Moreover, this is an *implicational* hierarchy in the sense that it is somehow *necessary* that whatever satisfies a category at the *nth* level of abstraction must always satisfy every category at higher-than-*n* levels of abstraction. (I don't care, for present purposes [actually, I don't think I care at all] whether this necessity is analytic or even whether it is linguistic. Suffice it that it is no accident that every poodle is a dog.)

The idea of *basic* categories is that some of the levels of abstraction in such implicational hierarchies have peculiar psychological salience. Intuitively, salience clusters at the "middle" levels of abstraction (in the present case, *dog* rather than *poodle* or *thing*). There is, alas, no independent definition of "middle," and it is quite conceivable that intuitions about which levels are in the middle just *are* intuitions of relative salience. Still, the fact seems to be that the following cluster of psychological properties tend to converge on the same member (or members) of each implicational hierarchy; that is, whatever member(s) of a hierarchy has one of them is also quite likely to have the rest. A category that has them all is paradigmatically basic.

(a) The basic category of a hierarchy often turns out to correspond to the high-frequency item in vocabulary counts; "dog" is thus a higher-frequency lexical item than either "animal" or "poodle."

(b) The word for the basic category of a hierarchy tends to be learned earlier than words that express other levels in the hierarchy (Anglin, 1979).

(c) The basic category is often the least abstract member of its hierarchy that is monomorphemically lexicalized. Compare "Sheraton wing-back armchair"; "armchair"; "*chair*"; "furniture"; "artifact"; "physical object . . ." In some domains there is evidence that the monomorphemic lexicalization of the basic category is universal—for example, there are few or no languages that have a single word for what we would call "a washed-out pinkish red" while coding what we would call plain "red" polymorphemically. (See Berlin and Kay, 1969.) As with (a) and (b), it seems natural to interpret (c) as a linguistic reflex of the relative psychological salience of the basic category as compared with other members of its hierarchy.

(d) Basic categories are natural candidates for ostensive introduction. "Dog" is ostensively definable for a child who hasn't learned "poodle," but it is probably not possible to teach "poodle" ostensively to a child who hasn't got "dog"; and it probably is not possible to teach "animal" ostensively to a child who hasn't got at least some animal words at the same level as "dog." This becomes glaringly obvious if one thinks about the relative ostensive definability of, e.g., "pale red," "red," and "color." Once again, it seems plausible to connect the relative ostensive definability of a word with the relative psychological salience of the property that the word expresses. (For a discussion of the implications of the correlation between basicness and ostensive definability, see Fodor, 1981a, chap. 10.)

(e) Basic categorizations yield 'information peaks' in the following sense. Ask a subject to list all the properties that come to mind when he thinks of *animals*; then ask him to list all the properties that come to mind when he thinks of *dogs*; and then ask him to list all the properties that come to mind when he thinks of *poodles*. One finds that one gets quite a lot more properties for *dog* than for *animal*, whereas the properties listed for *poodles* include very few more than one got for *dog*.[32] (See Rosch, et al., 1976.) It seems that—in some sense that is admittedly not very clear—basic categorizations are the ones that encode the most information per unit judgment. Taken together with Paul Grice's "maxim of quantity" (be informative) and his "maxim of manner" (be succinct), this observation predicts the following bit of pragmatics:

(f) Basic categories are the natural ones to use for describing things, ceteris paribus. "Ceteris paribus" means something like 'assuming that there are no special task demands in play'. You say to me, 'What do you see out the window?; I reply, 'A lady walking a dog', (rather than, e.g., 'A lady walking an animal' on the one hand, or 'A lady walking a silver-grey, miniature, poodle bitch', on the other. The point to notice here is that, all things being equal, the first is the preferred level of description even where I may happen to know enough to provide the third.

I assume that these linguistic facts are surface reflections of a deeper psychological reality, to wit:

(g) Basic categorizations are phenomenologically *given*; they provide, as it were, the natural level for describing things *to oneself*. A glance out the window thus reveals: a lady walking a *dog*, rather than a lady walking a silver-grey, miniature. . . etc. (Of course, sustained inspection alters all this. But phenomenological salience is accessibility *without* sustained inspection.) You might predict from these intuitions that perceptual identifications which involve the application of basic categories ought to be fast as compared to applications of either more or less abstract members of their implication hierarchies. There is, in fact, experimental evidence that this is true. (See Intraub, 1981.)

(h) Basic categories are typically the most abstract members of their implication hierarchies which subtend individuals of approximately similar appearance (Rosch, et al., 1976). So, roughly, you can draw something that is just a dog, but you can't draw something that is just an animal; you can draw something that is just a chair, but you can't draw something that is just furniture.

This observation suggests that, to a first approximation, basic categorizations (unlike categorizations that are more abstract) can

be made, with reasonable reliability, on the basis of the visual properties of objects. It thus returns us to the issue of perception. Since input systems are, by assumption, informationally encapsulated (no generalized top-down access to background information), the categorizations such systems effect must be comprehensively determined by properties that the visual transducers can detect: shape, color, local motion, or whatever. Input systems aren't, of course, confined to encoding properties like shape and color, but they *are* confined—in virtue of their informational encapsulation—to categorizations which can be inferred, with reasonable accuracy, from such "purely visual" properties of the stimulus.[33] (Compare: the language processor is confined to recovering properties of the input token that can be inferred, with reasonable accuracy, from its acoustic properties—hence to recovering linguistic form rather than, say, the speaker's metaphorical intent.)

Putting it all together, then: basic categorizations are typically the most abstract members of their inferential hierarchies that *could* be assigned by an informationally encapsulated visual-input analyzer; more abstract categorizations are not reliably predicted by *visual* properties of the distal stimulus. And basic categorizations are the ones that you would want the input systems to deliver assuming that you are interested in maximizing the information per unit of perceptual integration (as, presumably, you are). So, the suggestion is that the visual-input system delivers basic categorizations.[34]

A lot follows from this suggestion: for example, that in one useful sense of the observation/theory distinction, dogs but not protons count as observed; that the outputs of the visual processor—like the outputs of the language processor—constitute a level of representation on grounds independent of the fact that they happen to be the set of representations that some input system delivers; that it is no accident that the phenomenologically accessible categorizations are expressed by ostensively definable words. And so forth. I leave it to the reader to draw the morals. Suffice it that the notion that visual analyses are computed by an informationally encapsulated system leads to the prediction that there should be some set of representations which are (roughly) shape-assignable on the one hand, and which, on the other hand, play a specially central role in the mental life of the organism. The pregnancy of the basic category construct suggests that this prediction is true.