Toward a Cognitive Theory of Consciousness

I

Philosophers of mind and epistemologists have much to learn from recent work in cognitive psychology, but one of philosophy's favorite facets of mentality has received scant attention from cognitive psychologists, and that is consciousness itself: full-blown, introspective, inner-world, phenomenological consciousness. In fact if one looks in the obvious places (the more ambitious attempts at whole theories, overviews of recent research, and more specialized work in such areas as attention and "mental imagery") one finds not so much a lack of interest as a deliberate and adroit avoidance of the issue. I think I know why. Consciousness appears to be the last bastion of occult properties, epiphenomena, immeasurable subjective states — in short, the one area of mind best left to the philosophers, who are welcome to it. Let them make fools of themselves trying to corral the quicksilver of "phenomenology" into a respectable theory.

This would permit an acceptable division of labor were it not for the fact that cognitive psychologists have skirted the domain of consciousness by so wide a margin that they offer almost no suggestions about what the "interface" between the models of cognitive psychology and a theory of consciousness should be. I propose to fill this gap and sketch a theory of consciousness that can be continuous with, and help unify, current cognitivist theories of perception, problem-solving, and language use. I fear that to the extent that the view I put forward is seen to meet these desiderata it will *seem* not to do justice to the phenomena, so it would help if first I said just what I am trying to do justice to. Nagel (1974) has epitomized the

problem of consciousness with the question: "What is it like to be something?" It is certainly not like anything to be a brick or a hamburger; it certainly is like something to be you or me; and it seems to be like something to be a bat or a dog or a dolphin, if only we could figure out what. The question, "is it like something to be an *X*?" may in the end be the wrong question to ask, but it excellently captures the intuitions that constitute the challenge to a theory of consciousness. Until one's psychological or physiological or cybernetic theory explains how it can be like something to be something (or explains *in detail* what is wrong with this demand), one's theory will be *seriously* incomplete. It is open to the theorist, of course, to reject the challenge out of hand. One can emulate those behaviorists who (it has been charged) "feign anesthesia" and categorically deny that anyone *has* an inner life. This course has little or nothing to recommend it. Some behaviorists may find this comfortable ground to defend, but it would be awkward at the very least for the cognitivist, who has to explain what is going on when, for example, one asks one's experimental subjects to form a mental image, or to give an introspective account of problem-solving, or to attend to the sentences in the left earphone rather than the sentences in the right earphone. The cognitivist must take consciousness seriously, but there are relatively noncommittal ways of doing this. One can somewhat paradoxically treat consciousness itself as something of a "black box" from which introspective and retrospective statements issue (with their associated reaction times, and so forth), but how is this black box fastened to the other boxes in one's model? I shall propose an answer to this question, one that will also be a partial account of what is going on inside the black box.

II

There is much that happens to me and in me of which I am *not* conscious, which I do not experience, and there is much that happens in and to me of which I *am* conscious. That of which I am conscious is that to which I have *access*, or (to put the emphasis where it belongs), that to which *I* have access. Let us call this sort of access the access of personal consciousness, thereby stressing that the *subject* of that access (whatever it is) which exhausts consciousness is the *person*, and not any of the person's parts. The first step in charac-

terizing this access is to distinguish it from two other sorts of access
that play important roles in cognitive theories. The first of these can
be called *computational access*. When a computer program is com-
posed of subroutines (typically governed by an "executive" routine)
one can speak of one routine having access to the output of another.
This means simply that there is an information link between them:
the results of computation of one subroutine are available for fur-
ther computation by another subroutine. A variety of interesting
issues can be couched in terms of computational access. For instance,
Marvin Minsky (1974) faults the design of current chess-playing
programs by pointing out that the executive programs typically do
not have enough *access* (of the right sort) to the routines that evalu-
ate the various lines of play considered. Typically, the evaluator
"has to summarize the results of all the search . . . and compress
them into a single numerical quantity to represent the value of being
at node A . . . [but] we want S [the output of the evaluator] to
tell the Move Generator which kinds of moves to consider. But if S
is a mere number, this is unsuitable for much reasoning or analysis."
It would be better if the higher executive had more access to the de-
tails of the line of play evaluated, and not just a summary judgment.

In a very different context, Julesz's (1971) perception experiments
using randomly generated dot displays show that at least some per-
ceptual information about depth, for instance, is computed by a
process that has *access* to highly uninterpreted information about
the pattern of light stimulating the retinas. Lines of computational
access are currently being studied in cognitive psychology and re-
lated fields, and there are useful characterizations of direct and in-
direct access, variable access, gated access, and so forth. Computa-
tional access has nothing *directly* to do with the access of personal
consciousness, for *we* do not have access to many things that vari-
ous parts of our nervous systems are shown to have access to. For
instance, some levels of the visual processing system must have com-
putational access to information about inner ear state changes and
saccadic eye movements, but *we* do not, and *we* have virtually no
access to the information our autonomic nervous systems must have
access to in order to maintain the complex homeostases of health.

The second sort of access to distinguish from both computational
access and the access of personal consciousness might be called

public access. Often it is useful to a programmer to have access to what the computer is doing, so that the computer's progress on the program can be monitored; and to this end a "trace" is provided for in the program so the computer can print out information about the intermediate steps in its own operations. One provides for public access of this sort by designing a print-out subroutine and giving it *computational* access to whatever one wants public access to. This is a nontrivial additional provision in a program, for there is a difference between, say, the access the executive routine has to its subroutines, and the access the print-out routine has to the access the executive routine has. The domain of computational access for a system and the domain of public access for the system user are as distinct as the functions and offices of Secretary of State Henry Kissinger and Press Secretary Ron Nessen. Kissinger has computational access to much information that we the public have no access to because Nessen, our avenue of access, has no computational access to the information. What is used for control is one thing and what is available to the public is another, and there is at best a contingently large overlap between these domains, both in computer systems and in the White House.

The notion of public access seems to bring us closer to the personal access of consciousness, for we are speaking creatures (we have a sort of print-out faculty), and—at least to a first approximation —that of which we are conscious is that of which we can tell, introspectively or retrospectively. There is a problem, however. So far, the *subject* of public access has not been identified. On the one hand we can speak of the *public's* access via print-out or other publication to what is going on in a system, and on the other we can speak of the *print-out faculty's* computational access to the information it publishes; but surely neither of these subjects is the "I" who has access to my contents of consciousness, nor does any more suitable subject appear likely to be found in this neighborhood.[1]

The picture of a human being as analogous to a large organization, with intercommunicating departments, executives, and a public relations unit to "speak for the organization" is very attractive and useful. The basic idea is as old as Plato's *Republic*, but it seems to have a fatal flaw: it is not like anything to be such an organization. What is it like to be the Ford Administration? Nothing, obviously,

even if it is like something to be a certain part of that administration. The whole is a very clever assemblage of coordinated *parts* that at its best acts with a unity not unlike the unity of a single person (Rawls, 1971), but still, it has no soul of its own, even if some of its parts do.

This apparently decisive shortcoming threatens a wide spectrum of theory-building enterprises currently receiving favorable attention in philosophy and psychology. Any philosopher of mind who (like myself) favors a "functionalist" theory of mind (see, for example, Block, this volume—Ed.) must face the fact that the very feature that has been seen to recommend functionalism over cruder brands of materialism—its abstractness and hence neutrality with regard to what could "realize" the functions deemed essential to sentient or Intentional systems—permits a functionalist theory, however realistically biological or humanoid in flavor, to be instantiated not only by robots (an acceptable or even desirable consequence in the eyes of some), but by suprahuman organizations that would seem to have minds of their own only in the flimsiest metaphorical sense.[2] Psychologists cannot escape this embarrassment merely by declining to embrace philosophers' versions of functionalism, for their own theories are vulnerable to a version of the same objection. Functionalist theories are theories of what I have called the subpersonal level. Subpersonal theories proceed by analyzing a person into an organization of subsystems (organs, routines, nerves, faculties, components—even atoms) and attempting to explain the behavior of the whole person as the outcome of the interaction of these subsystems. Thus in the present instance the shortcoming emerged because the two access notions introduced, computational access *simpliciter* and the computational access of a print-out faculty, were defined at the subpersonal level; if introduced into a psychological theory they would characterize relations not between a person and a body, or a person and a state of affairs or a person and anything at all, but rather, at best, relations between *parts* of persons (or their bodies) and other things. So far as I can see, however, every cognitivist theory currently defended or envisaged, functionalist or not, is a theory of the subpersonal level. It is not at all clear to me, indeed, how a psychological theory—as distinct from a philosophical theory—could fail to be a subpersonal theory.[3] So the

functionalists' problem of capturing the person as subject of experience must arise as well for these cognitivist theories. At best a subpersonal theory will seem to give us *no grounds* for believing its instantiations would be subjects of experience, and at worst (as we have seen) a subpersonal theory will seem to permit instantiations that *obviously are not* subjects of experience. Take your favorite inchoate cognitivist theory and imagine it completed and improved along the lines of its infancy; is it not always easy to imagine the completed theory instantiated or "realized" by an entity — an engineer's contraption, for instance, or some kind of zombie — to which we have no inclination to grant an inner, conscious life?

Intuition, then, proclaims that any subpersonal theory must leave out something vital, something unobtainable moreover with subpersonal resources. Intuitions can sometimes be appeased or made to go away, however, and that is the task I set myself here. I propose to construct a full-fledged "I" out of subpersonal parts by exploiting the subpersonal notions of access already introduced.[4]

The first step is to sketch a subpersonal flow chart, a cognitivistic model that by being subpersonal "evades" the question of personal consciousness but, unlike cognitivistic psychologies with which I am familiar, prepares attachment points for subsequent explicit claims about consciousness. The flow chart will be a philosopher's amateur production, oversimplified in several dimensions, but I think it will be fairly clear how one could go about adding complications.

III

For clarity I restrict attention to six of the functional areas to which a theory of consciousness must do justice. (See Figure 1). At the output end we have the print-out component, and since this is our own Ron Nessen analogue I shall call it *PR*. *PR* takes as input *orders to perform speech acts*, or *semantic intentions*, and executes these orders. The details of the organization of the *PR* component are hotly contested by psycholinguists and others, and I do not wish to adjudicate the debates. Roughly, I suppose the breakdown to be as follows: the speech act command gets turned into an *oratio obliqua* command (to say that *p*), and this gets turned into a "deep structure" specification — in "semantic markerese" perhaps — which in turn yields a surface structure or *oratio recta* specification. We
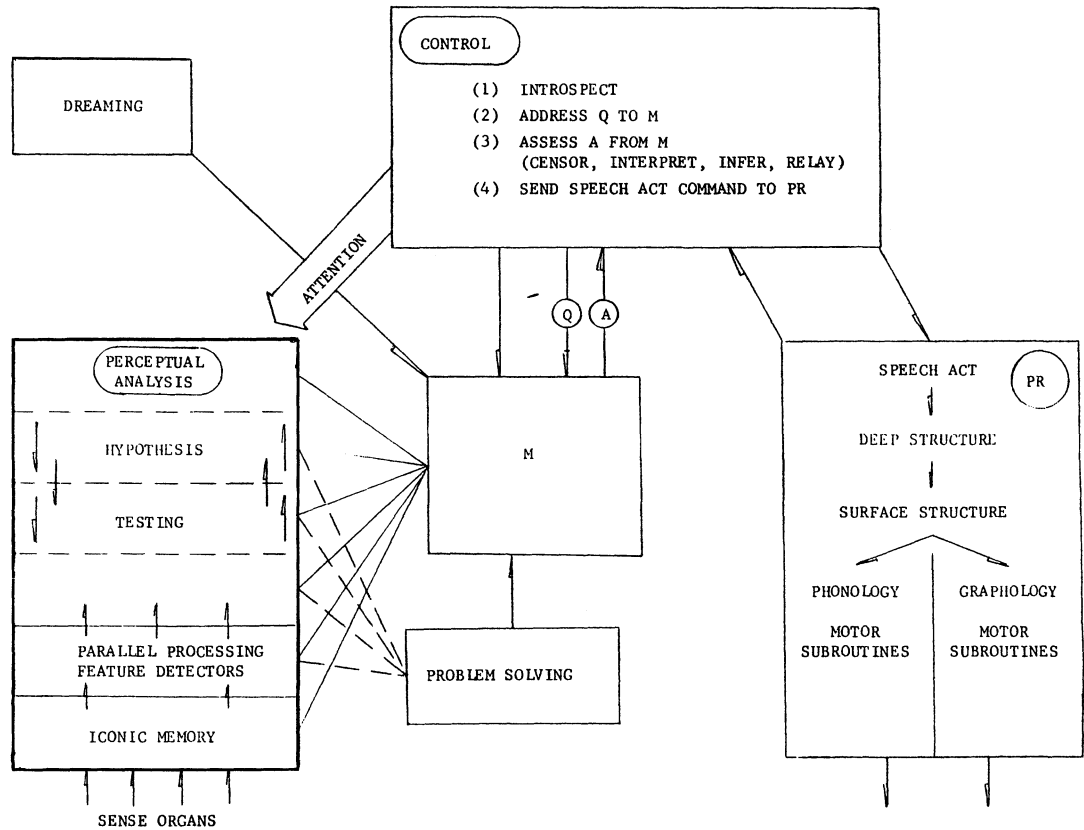
DREAMING

CONTROL

(1) INTROSPECT
(2) ADDRESS Q TO M
(3) ASSESS A FROM M
    (CENSOR, INTERPRET, INFER, RELAY)
(4) SEND SPEECH ACT COMMAND TO PR

ATTENTION

PERCEPTUAL
ANALYSIS

HYPOTHESIS

TESTING

PARALLEL PROCESSING
FEATURE DETECTORS

ICONIC MEMORY

SENSE ORGANS

Q   A

M

PROBLEM SOLVING

SPEECH ACT   PR

DEEP STRUCTURE

SURFACE STRUCTURE

PHONOLOGY   GRAPHOLOGY

MOTOR   MOTOR
SUBROUTINES   SUBROUTINES

Figure 1.

can imagine this to branch into either a phonological or graphological specification, depending on whether the initial command was to speak or write. These specifications, finally, drive motor subroutines that drive the vocal or writing apparatus to yield an ultimate execution of the input intention. There is a good deal of interaction between the levels: if one has difficulty pronouncing a certain word, this may count against its inclusion in the surface structure if one intends to speak but not if one is writing.

PR gets all its directions from a higher executive or *Control* component, but the pool of information to which *PR* has access is a special short-term memory store or buffer memory, which I shall simply call *M*. The lines of communication between Control, *M*, and *PR* are roughly as follows: suppose Control "decides" for various reasons to "introspect":

(1) it goes into its introspection subroutine, in which
(2) it directs a question to *M*;
(3) when an answer comes back (and none may) it assesses the answer: it may
    (a) censor the answer
    (b) "interpret" the answer in the light of other information
    (c) "draw inferences" from the answer, or
    (d) relay the answer as retrieved direct to *PR*
(4) The outcome of any of (a-d) can be a speech command to *PR*.

The point of the buffer memory *M* is that getting some item of information into *M* is a necessary but not sufficient condition for getting it accessed by *PR* in the form of the *content* of some speech act command.

Now what gets into *M* and how? First let us look at perception. I assume a tier of perceptual analysis levels beginning with sense-organ stimulation and arriving ultimately at highly interpreted information about the perceived world, drawing often on more than one sense modality and utilizing large amounts of stored information. The entire process is variably goal-dependent. Again, the details of this stack of processes are controversial, but I shall venture a few relatively safe points. At the lowest levels we have what Neisser (1967) calls "iconic memory,"[5] a very short storage of the stimuli virtually uninterpreted. "Parallel processing" by "feature detectors"

takes us up several levels and yields crude but local-specific information about edges, corners, shapes, patches of color, and so forth. From there a process of "hypothesis generation and confirmation" takes over, a *sequential*, not parallel, process that utilizes both stored "world knowledge" (in the "expectation-driven mode") and the results of the parallel feature detectors (in the "data-driven mode") to determine the generation of hypotheses and their confirmation and disconfirmation. Perhaps the "data structures" at the highest levels of this process are Minsky's "frames" (1974), and perhaps they are not. All the processes of perceptual analysis, but especially the higher, sequential levels are governed by complicated instruction from Control. As Neisser (1967) argues convincingly, with limited cognitive resources with which to perform this sophisticated task of perceptual analysis, Control must budget wisely, allocating the available cognitive resources to the sensory modality or topic of most current importance. This allocation of cognitive resources is the essence of *attention*, Neisser argues, and I partially concur. There is a notion of attention that is very definitely a matter of allocation of cognitive resources.[6] This notion of attention, important as it is, is only very indirectly connected with consciousness, as can be seen at a glance if one considers the fact that any problem-solving or game-playing computer pays attention, in this sense, first to one candidate course of action and then to another, and presumably it would not on this ground be deemed conscious. Or consider the fact that a somnambulist must no doubt allocate considerable cognitive resources to the job of navigating successfully and maintaining balance while being, in some important sense, unconscious (and unconscious *of* all this calculation) at the time. In this sense of attention, *unconscious attention* is no contradiction in terms, and in fact no hints at all have been given to suggest what *conscious* attention might be.

Now the perceptual analysis component sends information to *M from many levels*. Why? Because when one sees a complex scene and analyzes it as, say, a chair and a table in the middle of the room, one sees more than just that there are that chair and table there. One sees the shapes, colors, local details, and periphery too. I do not want to *identify* what one experiences with what one *can say*, but at least *if* one can say something about some current feature of the perceivable world, one has experienced it. This is vividly brought

out by tachistoscopic experiments (Neisser, 1967). If one sees a string of four alphabetic letters flashed for a few milliseconds on a screen, one cannot usually identify them. Although the stimulus pattern persists in iconic storage after the actual external flash has ceased, this storage decays before the higher-level processors can complete their work; and once the data are lost, analysis must terminate. But one sees *something*; one can say that one has seen a flash, or a flash with some dark objects, or even four letters or symbols. Something is experienced even though perceptual analysis is not completed. In such a case, I am supposing, the results of however much analysis gets accomplished normally go to *M*. These results will go other places of great importance as well, no doubt, but for our purposes all that matters is what gets into *M*.[7]

Perception, then, sends a variety of inputs to *M*. Perceptual experience is not the only conscious experience we have, though, so what else must we suppose gets into *M*? We are normally conscious of our thinking when we set out to solve problems, so let us very artificially isolate a *problem-solving component* that sends its results to *M*. (At least for some sorts of problem-solving—"imagistic" problem-solving—it is tempting to suppose the processes utilize a lot of the machinery of perceptual analysis; hence the dotted lines in Figure 1.) We shall return later to this component and its interactions with *M*. Another unit that sends information to *M* is Control itself. A partial record of its goals, plans, intentions, beliefs gets installed in *M* for occasional publication when the situation demands it.

These are the essential units of the system for my purposes here, but just to illustrate how the model could be extended, I add the dream-production unit. It loads *M* as well, and, as I have argued elsewhere (Dennett, 1976), the question whether dreams are experiences is to be answered by assessing the nature of this memory-loading process (the "route taken" by the access arrow).

Before turning to the question of how such a subpersonal model could possibly say anything about consciousness, let me illustrate briefly how it is supposed to handle various phenomena. Fodor (1975) discusses an experiment by Lackner and Garrett (1973). In dichotic listening tests subjects listen through earphones to two different channels and are instructed to *attend* to just one channel. They can typically report with great accuracy what they have heard

through the attended channel, but not surprisingly they can typically say little about what was going on concomitantly in the unattended channel. Thus, if the unattended channel carries a spoken sentence, the subjects can typically report they heard a voice, or even a male or female voice. Perhaps they even have a conviction about whether the voice was speaking in their native tongue, but they cannot report *what was said*. One hypothesis, based on Broadbent's (1958) filtering theory, is that a control decision is made to allocate virtually all the cognitive resources to the analysis of the attended channel, with only low-level ("preattentive") processing being done on the input from the unattended channel. Processing of the unattended channel at the level of semantic analysis, for instance, is on this hypothesis *just not done*. Lackner and Garrett's (1973) experiments disconfirm the Broadbent model in this instance, however. In the *attended* channel subjects heard ambiguous sentences, such as "He put out the lantern to signal the attack." In the unattended channel one group of subjects received disambiguating input (e.g., "He extinguished the lantern"), while another group had neutral or irrelevant input. The former group *could not report* what they heard through the unattended channel, but they favored the suggested reading of the ambiguous sentences more than the control group. The influence of the unattended channel on the interpretation of the attended signal can be explained only on the hypothesis that the unattended input is processed all the way to a semantic level, even though the subjects have no awareness of this—that is, cannot report it. On my model, this suggests that although higher-level processing of the unattended channel goes on, only low-level results are sent to M. This nicely illustrates the independence of computational access for *control* (in this case, influencing perceptual set in the attended channel) from computational access for *publication*, and gives an instance of, and an interpretation of, the well-known unreliability of introspective evidence. The *absence* of introspective evidence that a certain analysis has been performed is never reliable evidence that no such analysis has been performed. The analysis in question may simply be one of the many processes that contribute in other ways to control, perception, and action without loading M with its results. I shall discuss more subtle cases of the relationship between such processing and introspective access later.

To pave the way for this, I want to say a bit more about the interaction proposed between *PR*, Control, and *M*. Suppose *PR* gets a speech act command that for one reason or another it cannot execute. Words fail it. I propose that a failure discovery like this feeds back to Control, which will deal with the situation in a number of ways. It can alter its directions to Perceptual Analysis, producing a new perceptual set. This *may* result in a reinterpretation of the incoming stimulation, producing a *changed* input (at any level) to *M*, and then a changed speech act command to *PR*. Being unable at first to describe one's perceptual experience could lead in this way to a *change* in one's perceptual experience. (This would help explain, I suggest, the heightened capacity to discriminate—and experience—wines that comes from learning to use the exotic vocabulary of the wine-taster. What I am proposing is, of course, a very Kantian bit of machinery, designed in effect to knit intuitions and concepts together. Any psychological theory must address this problem; in some models the Kantian perspective is just more readily seen.) But if perceptual revision did not occur, Control could send to *PR* a direction to say that one finds the experience ineffable or indescribable, and this might be followed by a series of commands to say various things about what the experience was more or less like, about just how one's words are betraying one's true semantic intentions, and so forth. What I am granting is that there is no guarantee that information loaded into *M* has a publication in the native tongue that is acceptable to the system.[8]

What kind of *information* might fail to find expression in one's native tongue? Although *M* has been characterized as an information store, nothing has been said about the *form* the information must take. What sort of "data structures" are involved? Is the information encoded "propositionally" or "imagistically" or "analogically"? These important questions deserve answers, but not here. It is important here, however, to explain why I refrain from answering them, and that will require a digression.

The current debate in cognitive psychology between the propositionalists and the lovers of images[9] is multifariously instructive to philosophers, not only because it contains echoes of philosophic controversies,[10] but also because it clearly illustrates the close and systematic relationship between "pure" philosophy—especially

epistemology — and empirical psychology. Psychologists, faced with the practical impossibility of answering the empirical questions of psychology by brute inspection (how does the human nervous system accomplish perception or cognition?) very reasonably ask themselves an easier preliminary question: how could *any* (physical or mechanical or biological) system accomplish perception or cognition? This question is easier because it is "less empirical"; it is an *engineering* question, a quest for a solution (any solution) rather than a discovery, but it is still dominated by a mountain of empirical facts — in particular, facts about the powers, limits, and idiosyncrasies of actual human perceivers under a wide range of conditions.

The psychological question becomes: how could any system *do all that?* It is a question one is ill-equipped to answer if one does not know what *all that* is — for instance, if one is a philosopher largely unacquainted with the psychologists' data. Yet there is a strong aprioristic element in the psychologists' investigations, because it turns out to be very difficult to compose any model at all that could conceivably *do all that.* What is wrong with most models is that they fail to satisfy some quite general constraint or constraints on all solutions. The charge often leveled against such models is thus not (or *not just*) that they fail to account for some body of data, but that they *could not conceivably* account for human perception or cognition (for instance), since they violate some proclaimed necessary condition on all solutions. This aprioristic thinking is not peculiar to psychology. Engineers can enumerate necessary conditions for being an amplifier or a motor, and biologists can set down constraints on all possible solutions to the problem of the mechanics of genetic inheritance, to give just two examples. Once one decides to do psychology this way at all, one can address oneself to the problems raised by the most particular constraints, by middling constraints, or by the most general constraints. One can ask how any neuronal network of such-and-such features can possibly accomplish human color discriminations, or one can ask how any finite organic system can possibly subserve the acquisition of a natural language, or one can ask, with Kant, how anything at all could possibly experience or know anything. Pure epistemology thus viewed is simply the limiting case of the psychologists' quest, and any constraints the philosopher finds in that most general and abstract in-

vestigation bind all psychological theories as inexorably as constraints encountered in more parochial and fact-enriched environments.

Notice, too, that the philosophers' most abstract question is not asked in a factual vacuum; when we ask aprioristically how experience is possible, or what knowledge is, or how anything can be a symbol or have meaning, we appeal to, and are thus constrained by, an enormous body of commonplace facts: the facts that anchor what we mean by "experience," "symbol," and so forth. All the philosopher need know in the way of facts is what can be learned at mother's knee, but that is not nothing. The psychologist says: the experimental results bear me out, don't they? The philosopher says: that's what it is to understand an utterance, isn't it? If recently many philosophers of mind, knowledge, and language have found it useful or imperative to descend in the direction of more data, the reason is that the issues at the less general levels are proving to be fascinating, manipulable, and apparently useful in illuminating the more abstract level.

This is particularly apparent in the current controversy over propositions and images as vehicles of information in cognitive systems, a controversy of protean guise, sometimes appearing to be pure philosophy (and hence no business of psychologists!), sometimes an abstract engineering question for cyberneticists and the like, and sometimes a question of hard psychological, biological, or phenomenological fact. It has grown popular to the point of becoming second nature to talk of *information*-processing and transmission in the nervous system, but there is uncertainty and disagreement about the a priori constraints on any such talk of information. There is often the illusion that no problems attend the psychologists' talk of information, since information theory has presumably provided a rigorous foundation for such talk, but it is not often that psychologists have in mind any hard-edged information-theoretic sense of the term; usually what is being alluded to is the information or content an event within the system has *for the system as a (biological) whole* (what the frog's eye tells the frog's brain, or better, as Arbib has suggested (in conversation), what the frog's eye tells the frog.) The *content* (in this sense) of a particular vehicle of information, a particular information-bearing event or state, is and must be a function of its function in the system. This is the sense of "information"

utilized in our model (and in psychological models generally); so when I assert that, for instance, there is a transfer of information from some perceptual analysis area to M, I endow that transfer event with content, and the content it has is to be understood as a function of the function within the whole system of that event. So far, that event's function has been only circularly characterized: it has the function of conveying information about the results of analysis at that level to a functional area that is accessible to another functional area whose function is to express in a natural language just that information obtained by that level of perceptual analysis. Such a characterization is circular, but not necessarily viciously circular. The circle is a high-level holding pattern, which permits us to consider the constraints on any theory without descending to the next, more empirical level.

We can say, though, just what that next level down is. The content of a psychological state or event is a function of its function, and its function is—in the end, must be—a function of the *structure* of the state or event and the systems of which it is a part. Not just any structures can realize the functions that we determine must be realized, but the step from functional constraint to structural constraint is treacherous[11] and takes a philosopher quite far from home. When the question of "form of information" takes on this (quite proper) guise as a question of engineering, I leave it—reluctantly— to the engineers. I shall address myself shortly to the question in its "purely philosophical" or "phenomenological" guises. So, to end the digression, it would be best, for the time being, to stay in our circle and talk only of the *content* of states and events, and not the structure of the vehicles of content.

Stopping at a level of description above the stern demands of structural realizations is thus engaging in an extended exercise in more or less well-motivated handwaving,[12] but this handwaving may well be saved by ultimate realizations of these information-processing components, and if it is, it will have been not only not in vain, but an essential propaedeutic to such theorizing.

One can never be sure, however. For instance, the Control component in my model is awfully fancy. It has a superb capacity to address just the right stored information in its long-term memory, a talent for asking M just the right questions, and an ability to or-

ganize its long- and short-term goals and plans in a very versatile way. This is no homunculus that any AI researcher has the faintest idea how to realize at this time. The ever-present worry is that as we devise components — lesser homunculi — to execute various relatively menial tasks near the periphery, we shall be "making progress" only by driving into the center of our system an all-powerful executive homunculus whose duties require an almost Godlike omniscience.[13] I can make no firm claims for the soundness of my components in this model. The most I shall venture for them is that they seem to me not to reproduce the problems at deeper levels, thus merely postponing solution.

IV

With those qualifications and excuses behind me, I turn to the decisive question. Suppose an entity were all wired up in some fashion so as to realize the flow chart in Figure 1. What would it be like (if anything) to be such an entity? At first glance the answer seems to be: not like anything. The whole system has been designed to operate in the dark, as it were, with the various components accomplishing their tasks unperceived and unperceiving. In particular, we have not supposed any inner introspecting eye to be watching the perceptual analysis processes, the control decisions, the efforts of *PR* to execute its orders. And yet to us on the outside, watching such an entity, engaging it in conversation, listening to its efforts to describe the effects on it of various perceptual environments, there will be at least the illusion that it is like something to be the entity. In fact it will tell us (or at least seem to be telling us) just what it is like. But inside it is all darkness, a hoax. Or so it seems. Inside your skull it is also all darkness, and whatever processes occur in your grey matter occur unperceived and unperceiving. Can it be said that just as there is some other point of view that *you* have, there is some other point of view that *it* has?

It is hard to know how to answer that question. But the following may help. Suppose I put forward the bold hypothesis that you are a realization of this flow chart, and that it is in virtue of this fact that it seems — to us and to you — that there is something it is like to be you. Can you give good grounds for denying the hypothesis, and if so, what are they? What personal access do you have, and to

what? Here I must abruptly shift the perspective of this paper and wax phenomenological for a while. I want to draw your attention to a class of phenomena. If you ever had a sudden *presentiment* that someone was looking over your shoulder, or a *premonition* that something dire was about to happen, you are acquainted with the phenomena. These events are propositional episodes, thinkings that *p*; there is normally some inclination to express them (although the inclination is easily suppressed or cancelled), and we may not even express them to ourselves in "inner speech." When they occur in us we have not the faintest idea what their etiology is (unless we have some theory about the causes of premonitions; my point is that "to introspection" they arrive from we know not where). There are other more familiar examples of coming to want to say something without knowing how or why. Witticisms "occur to us," but we do not know how we produce them (the example is Ryle's). Lashley long ago pointed out that if asked to think a thought in dactyllic hexameter we (many of us) can oblige, but we have no awareness of how we do it: the *result* arrives, and that is the extent of our direct access to the whole business. Lashley's provocative comment on his example was that "no activity of the mind is ever conscious," and the interpretation of this I am supporting is that we have access — conscious access — to the *results* of mental processes, but not to the processes themselves.

My contention is that far from being rare and anomalous occurrences, the propositional episodes, these thinkings that *p*, are our normal and continuous avenue to self-knowledge, that they exhaust our immediate awareness, and that the odd varieties, such as the presentiment that someone is looking over one's shoulder, are striking only because of their isolation from the preceding and following presentiments, only because of our inability to follow them up with related propositional episodes about the same topic.[14] Right now it occurs to me that there are pages in front of me, a presentiment whose etiology is not known directly by me, but which is, of course, perfectly obvious. It is my visual system that gives me this presentiment, along with a host of others. I *can say* all sorts of things to elaborate on and supplement my initial report. But if I am put in an abnormal perceptual environment — for instance, in a tachisto-scopic experiment — I may be less sure *why* I want to say what I do.

I sort of have a hunch that is was an English word you flashed on the screen, but did I really *see* it? I cannot say what word it was, or describe it in any detail.

Instead of cajoling you with further phenomenological persuasions, I shall enlarge upon my view by drawing an analogy to Hume (1888). Hume's revolutionary step in the analysis of causation was to suggest that we had it all backwards. Earlier attempts at an account of our belief in causation supposed that when we saw a cause and then an effect we *saw* the necessary connection between them, and *thereupon and therefore* inferred or expected the effect when we saw the cause. Hume examined the cause ("turned it on all sides") and could *find* no such necessary connection to be observed, so he suggested that it was the other way around: having been conditioned, in effect, to infer or expect the effect when seeing the cause, we *found ourselves* drawing the inference, and this gave rise to an illusion of sorts that we were *seeing* a necessary connection that explained and grounded the inference we were compelled to make. The inference itself, Hume says, is psychologically and epistemically prior, and it gives rise to the belief in a "perceived" necessary connection. I am proposing a parallel account of "introspection": we find ourselves *wanting to say* all these things about what is going on in us; this gives rise to *theories* we hold about how we come to be able to do this—for instance, the notorious but homespun theory that we "perceive" these goings on with our "inner eye," and that this *perception* grounds and explains the semantic intentions we have.

Hume might almost have arrived at this extension of his view. He claims (1888, I.iv.2) that there is a difference between our "internal impressions" and our sensations. The latter require the positing of continuously existing external bodies in order to preserve the coherence and constancy of our discontinuous impressions of sensation. (Hume's example is the fire in his study fireplace that gradually burns down, turning slowly to embers in the periods between the interrupted and different sensations he receives from the fire.) But, says Hume (p. 195), "internal impressions" do not require this postulating; "on no occasion is it necessary to suppose that they have existed and operated, when they were not perceived, in order to preserve the same dependance and connexion, of which we have

had experience." This claim is virtually unavoidable for Hume —
given his allegiance to a Lockean doctrine of the "transparency" of
the mind to itself — but it is a fundamental error. Not only must we
come to accept all manner of covert influences, unconscious problem-
solving processes, and the like (recall Lashley's dictum), but *mental
images themselves* are the creatures of a "posit," an inference or
extrapolation exactly analogous to Hume's "posit" about external
bodies. This is graphically illustrated by Shepard's (1971) experi-
ments with "rotating mental images." The subjects in these experi-
ments are shown pairs of line drawings like those in Figure 2, and
asked whether or not the pair are different views of the same shape.
In this case, as you can quickly determine, the answer is Yes. How
did you do it? A typical answer is, "I rotated the image in my mind's
eye." Amazingly, Shepard set out to see if he could determine the
normal angular velocity of rotation of such images! How could he
do this? The subjects were given buttons to press to give their an-
swers. After tentative standard latency times were subtracted from
both ends of the duration between display and answer, Shepard was
left with durations that should, on his hypothesis, vary linearly (ig-
noring acceleration and deceleration) with the degrees of rotation
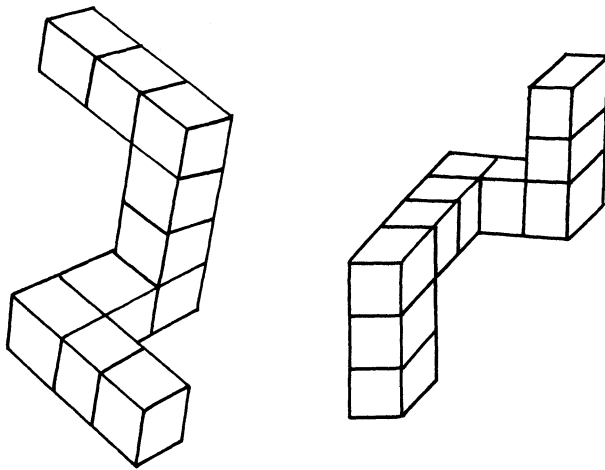required to bring the figures into superimposition. It should take

Figure 2.

roughly twice as long to rotate an image through 100 degrees as through 50 degrees. Shepard claims to have obtained significant positive results: he himself can rotate such mental images at an angular velocity of 62.6 degrees per second.

Now how can my view possibly accommodate such phenomena? Aren't we directly aware of an image rotating in phenomenal space in this instance? No. And that much, I think, you can quickly ascertain to your own satisfaction. For isn't it the case that if you attend to your experience more closely when you say you rotate the image you find it moves in discrete jumps—it flicks through a series of orientations. You cannot gradually speed up or slow down the rotation, can you? But now "look" again. Isn't it really just that these discrete steps are discrete propositional episodes: now it looks like *this*, but if I imagine it turned *that* much, it would look like *that* . . . ah yes, it would eventually look just like the other one. But the flicking, you may insist, is clearly part of a motion observed— the axis of rotation is, perhaps, vertical, not horizontal. But your reason for saying this is just that your intermediate judgments define the rotation. They are judgements that fall in an order that would be the proper order of perceptual judgments in a case of watching a real image rotate around a vertical axis. If you are inclined to argue that only an internal system that did actually proceed by some rotation in space of a representation or image could explain the sequence of judgments and their temporal relations in such cases, you might be right, but your grounds are hardly overwhelming. In fact, these discrete series of judgments bear a striking resemblance to the discrete series of small flashing lights that create the illusions of perceived motion, which have received so much attention from psychologists.[15]

We know that in these situations we all "perceive" motion—even elaborate orbital motions in three dimensions—when there is no motion. When we are confronted by a small group of these sequentially flashing lights we experience an illusion; we are led irresistibly to a nonveridical perceptual judgment that there is a single light moving in a particular way. What I am suggesting is that as the discrete series of flashes is to that nonveridical judgment, so our series of judgments in the image rotation case is to the judgment that something is really rotated in our minds (or in our brains, or anywhere).

There *may* be motion of something "behind" our judgments in the image rotation case, but if there is, it is something quite outside our present ken, and its very existence is suggested only by the most tenuous inference, however psychologically irresistible it may be.

My account of the Shepard phenomenon is that *however* the problem-solving process is accomplished, it yields *results*, both final and intermediate, that are available in M to be accessed by PR. These results, by the time they reach PR, are unproblematically propositional in nature: they are intentions to say that p. They are one product of perception or problem-solving. Another product is ducking when you see a flying object coming at you, but this is neither propositional nor imagistic so far as I can see. These products are perhaps only *indirect* products of perceptual processes; the *direct* or *immediate* product, one might wish to say, is experience itself, and the question is whether experience is propositional or imagistic or something else. My answer, counterintuitive as it may seem at first, is that if that question has any admissible interpretation at all, introspection cannot answer it. We have no direct personal access to the *structure* of contentful events within us.

V

Having given some suggestion about how the model I propose operates with a variety of phenomena, I now want to make some proposals about how the traditional categories of consciousness are to be superimposed on the model. These proposals are not supposed to be a priori truths about consciousness, or the dictates of conceptual analysis of our ordinary concepts, but rather suggestions about the best fit we can achieve between our pretheoretical intuitions (which are not entirely consistent) and a cognitive theory of the sort I have been sketching.

(1) One perceives more than one experiences. Perceptual analysis provides information about the world that is utilized in the control of behavior but is not accessible to introspection or consciousness, on any familiar understanding of these terms. In other words, there is nothing repugnant to theory in the notion of unconscious or subliminal perception or "subception," and any intuitions to the contrary should be discarded.

(2) The content of one's experience includes whatever enters (by

normal routes) the buffer memory M. What one experiences may decay before it is in fact accessed by PR, or it may be garbled in transition to PR, or it may be relatively inaccessible to PR.[16] In virtue of this possibility of error or malfunction between M and PR, *what one wants to say* is not an infallible or incorrigible determinant of what one has experienced or is currently experiencing. So the content of one's experience is given an *objective* characterization, and any intuition we have to the contrary that we are the sole and perfect arbiters of what we experience should be discarded.

(3) One experiences more at any time than one wants to say then. What fills the "periphery," adds detail to one's "percepts," inhabits "fringe consciousness," is, as phenomenologists have insisted, *there*. Where? In M. No more mysterious process of presentation or apprehension of inhabitants of phenomenal space is needed.

(4) One experiences more than one attends to—in either of two senses of attention. One experiences more than what results from higher-level allocations of cognitive resources, and one experiences, as (3) asserts, more than one is currently thinking. These are entirely different ways in which there are unattended contents of consciousness, even though there is a strong contingent link between them. Usually Control fixes things so that what one is attending to in the former sense is what one is attending to in the latter. Put otherwise, our conscious access to what we are attending to is normally excellent.

(5) One's access to one's experience is accomplished via the access relations between M and PR. As Anscombe would put it, we simply *can say* what it is we are experiencing, what it is we are up to. This is accomplished without any inner eye or introspective faculty beyond the machinery invoked in the model.

(6) Our feelings of special authority in offering introspective reports—the basis for all the misbegotten theses of introspective incorrigibility and infallibility—arises from the fact that our semantic intentions, which determine what we want to say, are the standards against which we measure our own verbal productions; hence if we say what we mean to say, if we have committed no errors or infelicities of expression, then our actual utterances cannot fail to be expressions of the content of our semantic intentions, cannot fail to do justice to the access we have to our own inner lives.[17]

VI

Having an inner life—being something it is like something to be
—is on this account a matter of having a certain sort of functional
organization, but the only natural entities that could be expected
to have such functional organizations would be highly evolved and
socialized creatures. The prospect of a robot artificially constructed
to replicate such a functional structure is not ruled out, but when
one reflects on the activities such an entity would have to engage
in to be more than an instantaneous version of such a system, the
claim that it would be conscious loses—at least for me—its im-
plausibility. We might not have the imagination to engage such a
thing in interpersonal relations; it might not *seem*, in its metallic
skin, to have an inner life or any prospect of an inner life. Such ap-
pearances are unreliable, however, for consider the possibility of
their being a truly conscious entity (whatever one supposes this in-
volves) that was just like us except that it operated on a time scale
ten thousand times slower than ours. We would have a very hard
time recognizing any of its day-long emissions as speech acts, let
alone witty, cheery, doleful, heartfelt speech acts, and its ponder-
ous responses to cuts and bruises would not easily enliven our sym-
pathies; but if so, we would *ex hypothesi* be ignoring a genuine,
conscious person among us.

When we wonder if something or someone is conscious, it is
tempting to view this as wondering whether or not a special light is
turned on inside. This is an error, however, as we can see by asking
questions about our own cases: was I conscious (or conscious of X)
at time t? When we see that what settles the issues in our own case
is a consideration of facts about our current capacities and past ac-
tivities, and the best theory that can account for these, we are less
reluctant to let the same considerations settle the issues in the case
of others.

There is no *proving* that something that seems to have an inner
life does in fact have one—if by "proving" we understand, as we
often do, the evincing of evidence that can be seen to establish by
principles already agreed upon that something is the case. In this
paper I set myself the task of constructing an "I," a something it
was like something to be, out of subpersonal parts of the sort en-
countered in cognitivistic theories. I do not now wish to claim that

I have *demonstrably* succeeded in this. Suppose we consider the two questions:

(A) Would an entity instantiating this theory sketch *seem* (to "others," to "us") to have an inner conscious life?

(B) Would such an entity *in fact* have an inner conscious life?
Question (A) is an agreeably straightforward question of engineering. Perhaps the thing whose design I have sketched would impress the keenest skeptic, and perhaps it would be lamentably (or comically) unrealistic or mute or self-defeating. Whatever flaws the design has might have philosophical or psychological significance, or might be rather trivial blunders on my part. (Is my model akin to the blueprint for a perpetual motion machine, or have I merely forgotten to provide a way out for the exhaust gases?) Most if not all objections to *details* in my model can be cast—even if not so intended—as grounds for denying (A), thus:

(C) Such an entity would not even *seem* to have an inner conscious life because

. . . it lacks any provision for such human phenomena as . . .,

. . . it ignores . . .,

. . . it would respond in situation - - - by doing . . .

I must take such objections seriously because part of my goal in this paper is to reveal, by imagined counterinstance, the implausibility of the charge that no entity describable solely by the resources of cognitivistic theory could *possibly* seem to have an inner conscious life. If that charge is nevertheless true (I cannot imagine how that could be *shown*—but perhaps I shall live and learn), then cognitivism is forlorn, and this would be a fact of great importance to philosophy and psychology.

Suppose, however, that some cognitivistic model of consciousness (not mine, no doubt, but its kin, I like to think) encouraged a positive answer to question (A). Suppose some model passed all the appearance tests we could devise. How on earth should one then address question (B)? Is there a better course than mere doctrinaire verificationism on the one hand or shoulder-shrugging agnosticism on the other? This is of course just "the problem of other minds," and I propose that progress can be made on it by reexamining *what one knows about one's own case* in the light of the most promising theories of psychology. What convinces *me* that a cognitivistic theory

could capture all the dear features I discover in my inner life is not any "argument," and not just the programmatic appeal of thereby preserving something like "the unity of science," but rather a detailed attempt to describe to myself exactly those features of my life *and the nature of my acquaintance with them* that I would cite as my "grounds" for claiming that I am — and do not merely seem to be — conscious. What *I* discover are facts quite congenial to cognitivistic theorizing, and my tactic here has been to try, by persuasive redescription, to elicit the same discoveries in others. Skeptics can view the form of the argument, such as it is, as a challenge — to produce a rival description of some feature of conscious experience that is both acceptable to many (better, it should evoke enthusiastic agreement, it should ring a bell) and unassimilable by cognitivistic theorizing. I am aware of the irony of recommending something so reminiscent of the battle of descriptions that embarrassed the early introspectionists to death, but how else could anyone plausibly support the claim that one's theory was a theory of *consciousness?*

Notes

1. There are other worries as well, of course. Nonhuman, nonverbal creatures have no print-out faculties, or at best very rudimentary and unexpressive print-out faculties, yet some philosophers — notably Nagel (1974) — insist that full-blown, phenomenological consciousness is as much their blessing as ours. I think one can be skeptical of this claim without thereby becoming the Village Verificationist, but the issue deserves an unhurried treatment of its own.

2. Davis (1974) has raised a graphic version of this objection with regard to functionalist theories of pain. Let a functionalist theory of pain (whatever its details) be instantiated by a system the subassemblies of which are not such things as C-fibres and reticular systems but telephone lines and offices staffed by people. Perhaps it is a giant robot controlled by an army of human beings that inhabit it. When the theory's functionally characterized conditions for pain are not met we must say, if the theory is true, that the robot is in pain. That is, real pain, as real as our own, would exist in virtue of the perhaps disinterested and businesslike activities of these bureaucratic teams, executing their proper functions. It does *seem* that there must be more to pain than that. See also Shoemaker (1975) for a careful analysis and rebuttal of several kindred objections to such functional theories.

3. Ryle (1949) and Wittgenstein (1953) are the preeminent modern theorists of the personal level. In fact, in their different ways they invent the enterprise, by showing that there is work to be done, that there are questions that arise purely at the personal level, and that one misconceives the questions if one offers subpersonal hypotheses or theories as answers. Typically readers who do not understand, or accept, these difficult claims see them as evading or missing the point, and complain that neither Ryle nor Wittgenstein has any positive psychological theory to offer at all. That is true: the personal level "theory" of persons is not a psychological theory.

4. This unpromising enterprise is forced on me (as students, colleagues, and other critics have insisted over the last few years) if I am to salvage the sort of functionalist theory of the mind I have heretofore defended. Since I have no other theory of the mind to fall back on, since in fact I see no remotely plausible alternatives to tempt me, I accept this problem as mine. It is not mine alone, though, as I hope I have made clear. This is fortunate, for the problem begs for a cooperative solution; my attempt trespasses deep in psychologists' territory, and I would hope to stimulate assistance, not a boundary dispute, from that quarter.

5. Neisser now wishes to banish iconic memory from a position of importance in the thoery of perceptual processes (see his contribution to this volume, and his *Cognition and Realist* (1976); but I cannot see that he can go so far as to deny its existence. Moreover I am not convinced it is not important.

6. Cf. the discussion of "awareness$_{1.5}$" in Arbib (1972).

7. *M* is a special hypothesized memory location, defined functionally by its access relations to *PR*, and it should not be confused with any already familiar functionally or anatomically defined variety of buffer memory, short-term memory, or echoic memory posited by cognitive theories to date. It may, for all I know, *coincide* nicely with some variety of memory already proposed and studied, but eclectic as my model is, I do not intend here to be appropriating any existing notion from psychology.

8. The possibilities for interaction of this sort between *PR* and Control have only been crudely exhibited here, but they can be — and to some extent have been — studied systematically. Relative retrieval times, lexical biases, the reliability of "tip-of-the-tongue" judgments, similarity spaces, and the like can provide an abundance of clues to guide the model builder. Consider James's (1950, pp. 251-2) introspective account of having a forgotten name on the tip of one's tongue:

> There is a gap therein; but no mere gap. It is a gap that is intensely active. A sort of wraith of the name is in it, beckoning us in a given direction, making us at moments tingle with the sense of our closeness, and then letting us sink back without the longed for term. If wrong names are proposed to us, this singularly definite gap acts immediately so as to negate them. They do not fit into the mold. . . . The rhythm of a lost word may be there without a sound to clothe it.

This passage, for all its phenomenological glories, is strikingly suggestive of purely functional interrelationships that might realistically be postulated to hold between the components of the model (or the components of a better model, of course). Suppose a functionalistic model "inspired" by this passage were developed and supported in the usual ways; it would be part of the burden of this paper to mitigate the resistance to the claim that an instantiation of such a theory could *assert* (*knowing* what it meant, and *meaning* what it said) just what James asserts in this passage.

9. An unrepresentative but philosophically valuable sampling of this literature would include Paivio (1971), Shepard and Metzler (1971), Arbib (1972), and Pylyshyn (1973, 1975).

10. The counterpart reference to work in philosophy should include Sellars (1963), (1973), Harman (1973), Armstrong (1973), and especially two new books: Fodor (1975) and Rosenberg (1974). I disparaged images in Dennett (1969) and propositions in Dennett (1975a). The present paper rescinds all (and only) the overstatments in those pieces.

11. Pylyshyn's papers (1973, 1975) give the clearest account of false steps from functional premises to structural conclusions. Minsky (1974) suggests that there are structures

undreamt of by the combatants in the words-versus-pictures debates that may in the end reveal that debate to be misconceived.

12. It only appears to be more specific handwaving when one talks not simply of contentful states and events but of cognitive *maps*, say, as opposed to stored *propositions*. People who like images say they are talking about images but not (of course!) about *pictures in the brain*; people who like propositions say they are talking about propositions – which are not *at all* like images – but also, of course, not *sentences in the brain*. There is plenty of doctrine about what images and propositions are *not*, but very little about what they are.

13. On the prospects and vicissitudes of homunculus theories see Dennett (1975b).

14. Cf. Ryle's (1949) illuminating account of "unstudied utterances" and Sellars's (1963, 1974) treatments of "thinkings out loud" and "proximate propensities" to think out loud. Ryle claims that our unstudied utterances "constitute our primary evidence for making self-comments" (p. 183), and I am claiming that it is our proximate propensities to make unstudied utterances that constitute this primary evidence. Ryle probably would not deny this, for he says (p. 194): "One of the things often signified by 'self-consciousness' is the notice we take of our own unstudied utterances, including our explicit avowals, whether these are spoken aloud, muttered or said in our heads. We eavesdrop on our own voiced utterances and our own silent monologues." Ryle offers no account, however, of just what one is doing when one "eavesdrops" on one's "silent monologue," nor could he give such an account without descending to the subpersonal level he wishes to avoid: *we* do not do anything in order so to eavesdrop. we *just are* aware of our own thinkings.

15. I owe this observation to Michael Hooker.

16. Ryle (1949, p. 160) says: "There is, however, a proper sense in which I can be said generally to know what has just been engaging my notice or half-notice, namely that I generally could give a memory report of it, if there was occasion to do so. This does not exclude the possibility that I might sometimes give a misreport, for even short-term reminiscence is not exempt from carelessness or bias." Ryle permits himself to call this capacity "log-keeping," and my *M* is apparently just Ryle's log-keeping system "paramechanized." (Another precursor of *M* that may have occurred to the reader is Freud's (1962, p. 10) *preconscious*: "The question, 'How does a thing become conscious?' would thus be more advantageously stated: 'How does a thing become preconscious?', and the answer would be: 'Through becoming connected with the word-presentations corresponding to it.' ") But surely Ryle's own remarks on log-keeping, if taken seriously, constitute just the sort of paramechanical hypothesis he typically condemns. Why is Ryle led to such an uncharacteristic account? Not because he is *aware* of keeping logs, and not because he finds himself or observes himself keeping logs (a Rylean would be quick to ask him embarrassing questions about how many entries he writes in his log before breakfast, and how he writes them). Ryle is led to this (happy) lapse, I suppose, because *what must be explained* (viz., our ability to report on so many different things that were just now happening) demands an explanation somehow in terms of an information – or memory – model.

17. I used to claim (Dennett, 1969) that this fact explained how we were, in a very limited and strained sense, incorrigible with regard to the contents of our awareness or consciousness. Now, thanks to the relentless persuasions of John Bender, William Talbott, Thomas Blackburn, Annette Baier, and others, I wish to claim that this fact explains not how we are in fact incorrigible, but rather why people (especially philosophers) so often think we are. The fact does provide for what Gunderson (1972) calls the investigational asymmetry of some first-person claims, but the asymmetry is not profitably to be viewed, as I used to think, as any sort of even limited incorrigibility.

References

Arbib, M. Consciousness: The secondary role of language. *Journal of Philosophy*, October 5, 1972, 69, 579-590.

Armstrong, L. *Belief, truth and knowledge*. Cambridge: Cambridge University Press, 1973.

Broadbent, D. E. *Perception and communication*. Oxford: Pergamon Press, 1958.

Davis, L. Unpublished paper delivered to the Association for Philosophy and Psychology, M.I.T., October 1974.

Dennett, D. C. *Content and consciousness*. London: Routledge & Kegan Paul, 1969, pp. 90-96.

Dennett, D. C. Brain writing and mind reading. In K. Gunderson (Ed.), *Language, mind, and knowledge*. Minneapolis, University of Minnesota Press, 1975. (a)

Dennett, D. C. Why the Law of Effect will not go away. *Journal of Theory of Social Behavior*, October 1975. (b)

Dennett, D. C. "Are dreams experiences?" *Philosophical Review*, April, 1976.

Fodor, J. *The language of thought*. New York: Crowell, 1975.

Freud, Sigmund. *The ego and the id*. New York. Norton, 1962.

Gunderson, K. *Content and consciousness* and the mind-body problem. *Journal of Philosophy*, 1972, 69, 591-603.

Harman, Gilbert. *Thought*. Princeton: Princeton University Press, 1973.

Hume, David. *A Treatise of human nature*, ed. L. A. Selby-Bigge. London: Oxford University Press, 1888.

James, William. *The principles of psychology*, vol. I. New York: Dover Publications, 1950. First published 1890.

Julesz, B. *Foundations of cyclopean vision*. Chicago: University of Chicago Press, 1971.

Lackner, J. R., & Garrett, M. "Resolving ambiguity: effects of biasing context in the unattended ear." *Cognition*, 1973, 1, 359-372.

Minsky, M. A Framework for representing knowledge. M.I.T. AI Lab Memo no. 306, 1974.

Nagel, T. What is it like to be a bat? *Philosophical Review*, 1974, 83, 435-451.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Neisser, U. *Cognition and reality*. San Francisco: W. H. Freeman and Co., 1976.

Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart and Winston, 1971.

Pylyshyn, Z. What the mind's eye tells the mind's brain. A critique of mental imagery. *Psychological Bulletin*, 1973, 1975, 86, 1-24.

Pylyshyn, Z. Representation of knowledge: non-linguistic forms. In *Theoretical issues in natural language processing*, Assoc. for Computational Linguistics Preprint, 1975.

Rawls, J. *Theory of justice*. Cambridge, Mass.: Harvard University Press, 1971.

Rosenberg, J. *Linguistic representation*. Dordrecht: Reidel, 1974.

Ryle, G. *The Concept of Mind*. London: Hutchinson, 1949.

Sellars, W. *Science, perception and reality*. London: Routledge and Kegan Paul, 1963.

Sellars, W. Meaning as functional classification. *Synthèse*, 1974.

Shepard, R. N., and Metzler, J. Mental rotation of three-dimensional objects. *Science*, 1971, 171, 701-3.

Shoemaker, S. Functionalism and qualia. *Philosophical Studies*, 1975, 27, 291-315.

Wittgenstein, L. *Philosophical investigations*. Oxford: Blackwells, 1953.