CONTENT AND CONSCIOUSNESS:
THE SECONDARY ROLE OF LANGUAGE

Michael A. Arbib

Philosophical Report 71 P-1
Computer and Information Science
University of Massachusetts
Amherst, MA. 01002

(April 19, 1972)

A paper to be presented at the December 1972 meeting of the American Philosophical Association.

CONTENT AND CONSCIOUSNESS: THE SECONDARY ROLE OF LANGUAGE

Michael A. Arbib

University of Massachusetts at Amherst

Many philosophers have claimed that it is impossible to paraphrase an intensional sentence such as "A believes it is raining" by an extensional sentence which captures its full meaning, and have argued that this refutes a materialist theory of mind. In his recent book, "Content and Consciousness"[1], Dennett has argued that considerations from the study of brain, evolution, and the design of artificial intelligences suggest that in fact our increasing scientific knowledge is allowing us to bridge the gap between the intensional and the extensional, and has set forth the methodological view that if, in one's account of brain functioning

> one does ascribe content to events, the system of ascription in no way interferes with whatever physical theory of function one has at the extensional level... [Thus] a centralist theory would consist of two levels of explanation: the extensional account of the interaction of functional structures, and an intensional characterisation of these structures, the events occurring within them, and states of the system resulting from these (79-80).

My purpose in this article is to advance this methodology while suggesting that one may go even further than Dennett has gone if one adopts the viewpoint that language arose as an elaboration of more fundamental functions of perception. This does not deny that the evolution of languages has greatly enriched human behaviour, but does suggest the danger of trying to understand the human mind by regarding language as its primary attribute.

I do not intend to include explicit arguments against the intentionalist's thesis of irreducibility, since I agree with Dennett's general position on this issue. However, I would suggest that many such arguments stem from a

pulley-and-lever view of mechanism with the concomitant lack of understanding of how complex the behaviour of a mechanism such as a computer can be. I firmly believe that a study of computer science and cybernetics should be required of every philosopher!

I

We often let our emphasis on language delude us into thinking that our interaction with the world must involve a verbal mediation of somewhat the following kind:

1. Seeing an object, name it
2. Using the name of the object, name the appropriate action
3. Act as designated.

though this simple scheme is elaborated by such considerations as the need to weigh alternatives, etc. This is the approach taken by many philosophers. It has also been taken by many artificial intelligence researchers because they sought methods amenable to string-processing on conventional "serial" computers, rather than trying to understand the brain mechanisms underlying human intelligence.

In the linguistic approach, all sensory input is viewed as funneling down to a few well-chosen words which can be processed by some centralized executive to yield a command which can then control the musculature. This may well be a useful model for analyzing much of human behaviour. My point, however, is that there exist sophisticated strategies for sensory-motor relations which do not require an arch-controller, and that we may do well to regard verbal mediation as providing a higher level in a hierarchical structure in which verbally formulable commands can control the output directly, but can leave the control of much of intelligent behaviour to lower centres, which need not make explicit use of verbal coding of input to

compute actions.

The primary purpose of perceiving objects is not, I suggest, to name them, but rather to be prepared to interact with them. I would thus caricature perception as gaining access to programs for interaction with an object rather than as the execution of a program: you perceive an object and yet may still leave it alone. In gaining access to a program, the perceiver only gives it potential command, further processing being required to determine whether or not the organism will act. Further - as when a cat is perceived both as 'something-to-be-stroked' and 'something-to-be-pushed-away' - there will be redundancy of potential command, with computation required to determine which of the accessed programs will actually control the ensuing actions of the organism. If we have been scratched frequently by cats, the "push program" may take command.

Evolution has given each animal a basic repertoire of skills for survival. But in all animals to some extent, more so in mammals and perhaps exeptionally so in man, these evolutionary skills are augmented by individually acquired skills and memories. These experiences, skills and memories cohere into what has been called an internal model of the world - the memory structure that, for example, lets us walk into a strange room and, on the basis of visual stimuli from a brownish rhombus, know that a table is present and that we may put the papers we are carrying on that table without risk that they will fall to the floor. The distinction between potential and direct action, and the role of perception in model-updating[2] makes it naive to assert that

> in the brain, discrimination of afferents according to their significance just is the production of efferent effects in differential response to afferents... effects on behavioural controls are conceptually required for there to be discrimination by

> significance, and... a stimulus, as a physical event, can have no intrinsic significance but only what accrues to it in virtue of the brain's discrimination, [and therefore] the problem-ridden picture of a stimulus being recognised by an animal, meaning something to the animal, prior to the animal's determining what to do about the stimulus is a conceptual mistake (74-76).

Our cybernetic framework suggests that Dennett is mistaken, for an animal may have access to a program, and this gaining of access is the animal's perception, prior to the animal's 'decision' whether or not to act on this program. Note, too, that it is confusing to talk of perception of an object x_1 if one somehow implies a total awareness of all its attributes - perception involves access to relations x_2 between the perceiver and its world and the relations with an object which are represented in the animal's brain will vary when it perceives the object on various occasions. This point is, in fact, well made by Dennett in his Chapter X on Language and Understanding:

> Only a being that is non-verbally active in the world could meet our requirements for understanding. A computer whose only input and output was verbal...might 'grasp' all the verbal connections, but it would lack 'acquaintance' with the things the words are about... The tests for understanding in each particular case involve a 'family' of behavioural capacities, some of which must be demonstrated.... [Moreover] what tests must be passed depend largely on what else the person knows and understands (181-2).

The statement "I am a doctor" means more to an adult than a child, the statement "salt is sodium chloride" means more to a chemist than a layman. No one need ever gain complete understanding of such statements, whose full meaning lies in a sociohistorical matrix, and a piece of information has meaning to a system only in the way in which it coheres in its 'internal model'.

Dennett states - perhaps somewhat inconsistently with his views just cited from Chapter X - that "assigning content to an event must be relating the event to a particular verbal expression (82)". But this seems to be so only if the assignation is done by a human in talking or writing. An animal's

actions, or a human's pantomime may serve as well. Recognising a face constitutes more than recalling the name of its owner, and yet few of us can verbalise the content of that recognition. Or when we say "the frog snapped at the fly", our assignation of content is but approximate, for the frog would not have snapped at the fly had the fly been immobile, but would have snapped at a wiggling pencil tip. Perhaps this is why, considering the example of a dog seeing a piece of steak lying on thin ice, Dennett notes that

> what the dog recognises [the steak]...as is something for which there is no English word, which should not surprise us - why should the differentiations of a dog's brain match the differentiations of dictionary English? ...If the centralist can say, roughly, that some feature of the dog's cerebral activity accounts for his desire to get the steak, and some other feature accounts for his fear (inculcated by certain past experinces) of what he takes to be thin ice, he will be matching imprecision for imprecision, which is the best that can be hoped for. Precision would be a desideratum if it allowed safe inferences to be drawn from particular ascriptions of content to subsequent ascriptions of content and eventual behaviour, but in fact no such inferences at all can be drawn from a particular ascription (85).

However, this viewpoint seems somewhat defeatist for - as in existing robots[3] - a precise specification at the intensional level might allow identification of what internal program was currently controlling action, and thus yield specifications of how the animal would behave.

In other words, <u>there is nothing which precludes description at the intensional level from expressing causal sequences providing our intensional language is extended to allow us to provide descriptions with the flexibility of a program</u>, rather than a statement of general tendencies. I reiterate that everyday language must be extended to face the task.

Dennett thus seems to go too far when he asserts that "people do not <u>do</u> anything to arrive at conclusions, but their brains must (149)" which is hard to reconcile with "People can reason, but brains cannot (149)". People can be

aware of various facts and can marshall them, compare them, combine them and arrive at conclusions. This does not deny that much goes on at the subpersonal level to determine which facts are brought to awareness, etc., so that people certainly do things to arrive at conclusions, but, as with all phenomena at the personal level, much is inexplicable without reference to the subpersonal level. It is perhaps because the incompleteness is in some sense particularly embarrassing when we talk of reasoning that Dennett takes the tack he does:

> Our inability to analyse introspectively our own problem-solving activities below a certain level of simplicity strongly suggests an analogy with certain sorts of problem-solving computer programmes... Ordinarily the print-out [of a computer providing a trace of protocol of a problem-solving computation] is in a high-order language rather than in basic machine language, and hence the computer is unequipped to report the truly atomic steps of its computations (151-152).

One should add that the machine protocols omit all the heuristic search[4] that underlies the selection of the next operator to be applied - this is perhaps an even more important point than Dennett's that each operator application requires a whole program in "machine language", for it suggests that intensional language of ordinary usage is not only short on details - which is, I think, relatively uninteresting - but lacks vital terminology, of the kind included in a programming language, for the explicit specification of context-dependent causal sequences. This emphasises the inadequacy of our verbal reports of our mental states to do justice to the richness that states must exhibit to play the role prescribed for them in system theory. It will be no easy task to elaborate our robot models to the point where they can do justice to such a rich 'background'. Nonetheless, I think our observations on heuristic search, parallel processing, etc., do indeed establish how much of what is beyond awareness - what we might identify with Polanyi's "Tacit

Dimension" - is already explicable, at least in principle, in the cybernetic framework.

If we know a computer is operating as designed, we can explain the fact that a computer added two numbers by saying that it executed a clear add and an add instruction. This does not preclude a deeper level of explanation which shows how the circuitry is set up which ensures that the gating configurations of the machine representation of these two instructions does indeed lead to the formation in the machine of the representation of the number which is the sum of the two numbers encoded by gating configurations at designated loci in the machine.[5] However, once one has guaranteed that machine language behaves as advertised, it would be tedious and unenlightening to explain the computer's behaviour at a level of explanation lower than that in terms of a sequence of executions of instructions. We thus arrive at a position consonant with Dennett's:

> While the peripheralist hopes to characterize behavioral events and stimulation extensionally from the beginning, and arrive at extensional laws relating these, [we take the approach of] the centralist [who] makes his initial characterization intensional, describing the events to be related in law like ways using... intensional expressions. He then hopes that an adequate physical basis can be found among the internal states and events of the organism so that 'reductions' of intensional sentences of the theory to extensional sentences of the theory is possible (41-42).

II

Dennett holds that a human's introspective reports are infallible, and asks:

> How can introspective utterances be so related to certain internal conditions that they can be viewed as error-free indications of these internal conditions?

I would suggest that this question is the wrong one to ask. If one is thinking of a face, the thought might be totally non-verbal - it is only being asked to describe the face that "projects" one into the verbal mode, and in this case usually with a great loss of detail. The phrase is but a projection of the thought, not the thought itself. Thus it is misleading to think that the thought is isomorphic to the phrase, just awaiting recoding to be uttered. This gap is made clear by Dennett's vision of a perceiving machine

> the output from [whose]...cameras could be recoded in any regular way to fit the input requirements of an immense neural net analyser which then fed <u>its</u> output into a 'speech centre' computer. The speech centre computer would be programmed to transform the output of the analyser into printed English 'reports', like 'I see a man approaching' (108).

My point (cf. footnote 3) is that the 'intelligent robot' of circa 1970 does <u>not</u> simply <u>classify</u> partial aspects of its environment - rather it continually seeks to update a model of that environment relevant to its activity. Thus utterances like "I see a man approaching" express mere aspects of its total state, and are appropriate as responses to a question like "Who is approaching?" But many different aspects of its current 'thoughts' could have been elicited by different questions.

Once we agree that the 'utterance' is but a sampling of the 'mental state', we must admit that it shares with all samples the danger of being unrepresentative.

In short, I regard it as a contingent fact that some reports are sufficiently reliable to delude some philosophers into believing that reports of mental states are infallible! It follows from any reasonable theory of the evolution of language that certain types of report will be highly reliable; it follows from the complexity of our mental states (which is lost in theories which hold that any simple report can capture their full content) that reports must be but samples, and hence may be unreliable.

Dennett's version of a perceiving machine has a visual analyser feeding into a speech centre. I would rather have a visual analyser in two-way interaction with a model generator which in turn directs interaction with the world. A speech centre, possibly subject to other inputs such as other's questions, samples the state of the model generator. The output of the speech centre may or may not be uttered. Dennett's point (110) seems to be that speech centre output is an infallible report of speech centre input. My point is that, even if this were true, it would be an infallible report of but a sampling of the true mental state. I should further add that it is not clear what is meant by infallibility for even if I have a system that reliably transcribes input A into the output 'I am in state B' this does not mean that the report is infallible unless input A results from a state of the organism which on criteria of behavioural equivalence outside observers will label as being akin to B. Thus where Dennett says:

> Using the notion of content ascription, and staying firmly on the subpersonal level of explanation, we can say that a sentence uttered is not a description of a cerebral event, but rather the expression of the event's content, which, after all, may be itself a description - of the visual field, for example (111).

we can agree with the general point, but must substitute "a partial (and thus possibly misleading) expression of the event's content" for the phrase "the

expression of the event's content". In other words, even if we agree to express a mental state in intensional terms, we must acknowledge - as our study of robots (cf. Footnote 3) makes clear - that the level of detail required in the intensional description for scientific purposes is far greater than that involved in our usual reports; and that our usual reports may often omit that which is most significant.

With this critique of the relationship between 'mental states' and 'utterances' we may turn to Dennett's analysis of 'awareness', by which he means to encompass all and only the intensional senses of common usage of awareness and consciousness. He then distinguishes two notions of awareness:

> What one can report directly, infallibly, and without speculation or inference is one thing, and what serves, or is relied upon, to direct behavioral responses is another. This can best be brought out by coining [two] artificial terms...:
>
> (1) A is aware$_1$ that p at time t if and only if p is the content of the input state of A's 'speech centre' at time t
>
> (2) A is aware$_2$ that p at time t if and only if p is the content of an internal event in A at time t that is effective in directing current behaviour (118-119).

In view of our critique, we might suggest that (1) be rephrased as "A is aware$_1$ that p at time t if and only if simple questioning could almost immediately cause p to be the content of the input state of A's 'speech centre' at time t" and resist the implication that if A is aware$_1$ of p, then p must be an infallible report of the state of A at time t.

In view of our general thesis that the internal model is an assemblage of representations which have <u>potential</u> command of action, we might change "that is effective" in (2) to "that is potentially effective", i.e. given an appropriate turn in environmental events - which perhaps destroys some of the force of the adjective 'current' in (2).

A door with a photocell-driven opening mechanism is thus $aware_2$ of the changes in light intensity that we know as shadows, while a tape-recorder is $aware_1$ of the pattern of magnetization under its reading-head. The very trivality of these examples (which Dennett would reject on just this ground) suggests that something crucial is missing from these definitions. It is, perhaps, the insertion of complexity in the form of an internal model of the world (with complexity then being in some sense a measure of level of awareness) that better approximates awareness. I would suggest that an animal may focus on specific aspects of its internal model and that this constitutes the essence of $awareness_1$ in a way for which the speech criterion is irrelevant. The real question is "What is it about a projection of the mental state that tempts us to describe it as being appropriate as an input to a speech centre?" rather than the existence of a speech centre. Perhaps we should make a new definition:

(1.5) A is $aware_{1.5}$ that p at time t if and only if p is a projection of the content of the mental state of A which expresses the concentration of A's attention at time t.

The point I am trying to make is that, in the evolution of language, $awareness_{1.5}$ must have been a necessary intermediate, and that the real challenge here is to understand the role $awareness_{1.5}$ could have played in an animal's behaviour before it evolved language. $Awareness_1$ is then a cheaply definable subset of the really interesting $awareness_{1.5}$. Thus I cannot agree with Dennett "that there is no important residue in the ordinary concept of awareness that is not subsumed under either $awareness_1$ or $awareness_2$(121)", though I do agree that "there is no room...for a concept of $awareness_3$ which would apply only to people and rule out all imaginable machines". Given our

example of the aware$_1$ tape-recorder we cannot even agree that awareness$_1$ is the more important concept. Rather, one has to analyze the role of language in articulating an internal model before one has the important sense of aware$_{1.5}$, which builds aware$_1$ on top of redundancy of potential command from aware$_2$. One then distinguishes humans from current machines by noting the relative paucity of the internal models held by current machines.

By emphasising awareness$_{1.5}$, we insist that dumb animals are aware$_{1.5}$ almost the way people are, and suggest that, in terms of the mental state of humans, awareness$_1$ is derivative and so relatively uninteresting - which is not to deny that it is extremely interesting when our study shifts from mental states to communication. Thus we must understand what distinguishes language from other communication systems (including action) and ask if the distinction between awareness$_{1.5}$ and awareness$_1$ is really crucial for our discussion of awareness. Dennett states that

> It is not that one must be aware$_1$ in order to recognise, but that one must be aware$_1$ in order to say that one recognises. To the extent that our ordinary concept of awareness leads to such confusions it is a poor concept in spite of its ordinariness (126).

To the extent that awareness$_1$ inhibits explicit discussion of awareness$_{1.5}$, it is a poor concept in spite of its novelty.

III.

We must not be too literal in our interpretation of the word "model" in the phrase "an organism's internal model of its environment". None of us would be so naive as to imagine that if we were to peer through the layers of cerebral cortex of the young child, we should see a global replica of the world, which under great magnification, would reveal cardboard replicas of her family and friends. Rather, we should simply imagine a neural network so changed as a result of experience, that if the pattern of spike trains entering it can in some sense encode a question about the world, then the pattern of spike trains leaving it can represent an answer about the world, without any actual inspection of the real world at that time being necessary to produce that answer. As times goes by, then, the computer in our head is so adjusted that our actions are better adapted to a whole range of properties of the world, in addition to those that confront our senses at the very moment. The word "model" in the phrase "internal model of the world" is thus to be used in this rather abstract sense, rather than in any pictorial sense. We may then distinguish <u>the</u> internal model of the world - the collection of properties of neural tissue which reflect past experience in a way which will help us compute our present behaviour - from <u>an</u> internal model that summarises what we believe is going on around us at present. We do not perceive just what we sense in front of our eyes. If we are in a room, we perceive that fact, and we have in some sense an internal model of that room with what is in it so that we may, for example, reach for an object previously seen behind us, without searching for it anew. Our perception thus does not involve independently processing a succession of "snapshots" of the room, but rather involves an initial comprehension of the room and the more salient of its contents, after which we need merely note discrepancies between our model and what is really out there to update this momentary model - as when we reach

for that object behind us only to find that someone has moved it. We repeat that this modelling and updating is all encoded in terms of the properties and activities of neurons and need have no resemblance to a photographic record.

Of course, it must not be thought that we have a different model for every different object of every different size or position in space. Rather the models must be flexible, in a way which may be suggested by a slide-box metaphor drawn from the making of movie cartoons. Drawing each frame individually is too inefficient. One might go for a whole minute of the cartoon without the background changing, so one could draw it just once. In the middle ground, there might be a tree, say, about which nothing particularly changes during a certain period of time except its position relative to the background. It could thus be drawn on a separate slide, which could then be displaced for succeeding frames. Finally, in the foreground, it may well be that one could draw most portions of the actors for repeated use, and then position the arms, facial expressions, etc., individually for each frame. The slides can then be photographed appropriately positioned in a slide-box for each frame, with only a few parameter changes and minimal redrawing required between each frame.

A similar strategy for obtaining a very economical description of what happens over a long period of time might be used in the brain, with long-term memory corresponding to a "slide file" and short-term memory corresponding to a "slide-box". The act of perception might then be compared to using sensory information to retrieve appropriate slides from the file to replace or augment those already in the slide-box, experimenting to decide whether a newly retrieved slide fits sensory input "better" than one currently in

the slide-box. Also, part of the action of the organism in changing its relationship with the environment might be viewed as designed to obtain input which will help update the short term memory by deciding between "competing" slides, as well as helping update the long term memory, by "redrawing" and "editing" the slides.

"Slide" is, of course, a bad name. The "slide-box" is not a box into which static slides are inserted. Rather, it is a mass of neural tissue lying athwart the channels which link the sensory and motor systems. "Putting in a slide" corresponds to activating this network, thus initiating transient wave-forms which change autonomously. Lying athwart the lines of communication, the "slide-box" fills the whole "postural-effector frame of reference", and so a slide does not contain information from any single modality. Our brain may use cues from one or more senses, or feedback from motor activity, to "address" the activation of a wealth of multi-modal action-oriented information. In using the slide-box metaphor we must not fall into the trap of letting terms such as "slide" force too rigid a view of neural activity, but should continue to emphasise that fine perceptual acts take place against a background; that we cannot recognise small details in a void; and that present "slides" strongly color the choice of each addition.

Let us now use this metaphor to extend our feel for the advantages and disadvantages of the "verbal mediation" approach. We first consider a case of programming a computer to answer questions and contrast this with a hypothetical machine which seems to solve the problem in a more biological way.

How might a computer be programmed to answer the question, "Is President Nixon a man?" when it has stored in memory a data base which includes the encoding of various sentences about politics, including sentences about

President Nixon? One sentence might actually state that "President Nixon is a man," and then the computer could find it simply by scanning its data base, and then print out, "Yes, President Nixon is a man". If no such statement were stored, it would have to combine several pieces of information in memory to obtain an answer. One stored statement might be, "Nixon is President of the United States," and another might be, "All Presidents of the United States have been men." Finding these strings, it would manipulate them in a purely formal way to obtain the deduction, "Therefore President Nixon is a man," and print out the answer "Yes".

In contrast to the above formal approach (typical of those which currently dominate artificial intelligence research) one could design a machine which uses the question, "Is President Nixon a man?" not to take sentences related to politics and President Nixon and combine them to answer, but rather use the name "President Nixon" to retrieve a "slide" of him, and then use the question, "Is he a man?" to retrieve a pattern recognition routine which can determine of a "slide" whether or not it represents a man.

To store a "slide" as a 'raw' picture encoded as a two-dimensional array of words, with each word giving the light intensity at a corresponding point of the picture, and then use a present-day computer which only processes a few words at a time, might be a fairly inefficient approach for computerized question-answering. However, when we turn to the brain, we find a series of approximately spatial arrays, corresponding to spatially organized reprocessings of the visual input - and, overall active system that the brain is with all cells continually active (as against current computers in which most storage is completely passive), it seems likely that a human brain would indeed answer such questions by recreating something of the highly parallel

pattern of neural activation somewhere reasonably high up in the head between the input streams and the neurons carrying commands down to the muscles, and then compute upon this pattern of activation in just the same way as it would upon that caused by visual input. Further, one might expect the internal representation of an activity within the brain to then be correlated with the spatial extent of the object it represents.

Let us contrast this viewpoint with Dennett's comments on imagery:

> Images can be in two or three dimensions, can be manufactured or natural, permanent or fleeting, but they must resemble what they represent and not merely represent it by playing a role - symbolic, conventional or functional - in some system. ...If an image is to function as an element in perception, it will have to function as the raw material and not the end product, for if we suppose that the end product of the perceptual process is an image, we shall have to design a perceiver-analogue to sit in front of the image... and so forth ad infinitum...
>
> Once perceptual analysis has begun there will indeed be elements of the process that can be said to be representations, but only in virtue of being interrelated parts of an essentially arbitrary system... The upshot of this is that there is no room in the subpersonal explanation of the perceptual process...for images (133-135)

This point is a vital one. However, I do not agree with Dennett that imagining is more like depicting in words than like painting a picture simply because we can fail to mention a lot in the verbal description of a man, but must commit ourselves to the presence or absence of a hat when we draw a picture. If one does not insist that the "slides" be pictorial - and my insistence on the nature of the metaphor emphasised that they are not[6] - the "slide-box" has some of the crucial spatial relationships of the picture while still having the necessary incompleteness of the verbal description.

> Unlike a snapshot of a tiger, a description of a tiger need not go into the number of stripes at all; "numerous stripes" may be all the description says (136).

Note that this is as true of looking at a tiger when not concerned with the number of stripes as it is of imagining a tiger. What we see is not so much a verbal description but a picture painted with qualities, rather than colors - and unless we pay attention to detail, the tiger's body is seen as painted with the quality "striped". We do not perceive all details, nor do we imagine them. In both perception and imagination, additional "painting" can take place, in the first case constrained by sensory matching, in the latter not. Since much perception is of where as well as of what, the brain-state may have a correlate of spatial location without implying that our perceptions exist in some ghostly space outside the brain. I agree with Dennett that it is naive to believe that when we look at a scene all details are available to us "at a glance" and can then be derived by introspection.

> When we are looking at something, as soon as questions come up they are answered immediately by new information as a result of the inevitable shift in the focus and fixation point of our eyes... When this occurs one is not scanning some stable mental image... one is scanning the outside world. One can no more become interested in part of one's visual experience without bringing the relevant information to the fore than one can run away from one's shadow. For this reason it is tempting to suppose that everything one <u>can</u> know via the eyes is <u>always</u> present to consciousness in some stable picture... [T]he natural operation of our eyes...is to make a cursory scanning of the environment whenever it changes and as soon as it changes, and...the operation of short-term memory...holds the results of this scanning for a short period of time (139-140).

However, it is also naive to go too far the other way. We move our eyes not simply in response to a novelty in the scene, but also to build up an internal representation. Eye movements are thus directed not only by external movements, but also by "information calls" during hypothesis generation, and by "mismatch signals" when peripheral features are no longer consonant with

the current representation. In other words, even though our internal model is incomplete and is continually being elaborated by shifts of attention, it is by no means a quickly fading memory of the last few fixations, as Dennett seems to imply. I may not be $aware_1$ of the blackboard behind me, yet as soon as I need to write on the blackboard, since its location is already part of my current internal model of the environment, I can immediately become $aware_1$ of it. In short, one must abolish any theory which posits photographs from the brain, but it seems inappropriate to exclude all other spatially structured forms of representation as well.

From this viewpoint, it no longer puzzles us that "color as a quality is eminently spatial ('Everything colored is extended')(142)". 'Red' is no more magical than 'human' in describing a patch of visual space, but the extent of the space so occupied is in both cases an important parameter. Of course, our metaphor is consistent with Dennett's insistence that "having an idea of a color...[does not] involve the existence of anything mental that <u>has</u> the color (142)". We have to agree, then, with the dismissal of the question of how we know that we "really" see red rather than green when others see red

> if the pupil's agreement with his teacher's public use of 'red' is satisfactory to show that the pupil has learned the word, we can, as Wittgenstein said, 'divide through' by the private quality, which is superfluous to the analysis (143).

This 'dividing through' is just what the automata theorist does when he identifies states which exhibit the same input-output function.[7] As Dennett says:

> Colors are what might be called functional properties. A thing is red if and only if when it is viewed under normal conditions by normal human observers,it looks red to them (146)."

However, it is a little dangerous to say that their saying an object is red

> does not hinge on their perusal of an internal quality, but on their perception of the object, their becoming aware$_1$ that the thing is red (146)

if one insists on awareness$_1$, and, in denying perusal, further denies some range of internal representations which correlate with our descriptions of redness. None of our perception is of primary qualities - as soon as we consider states of second-order and higher-order neurons, primary qualities are far behind us. The slide-box metaphor, by insisting that spatial extent enters all our basic perceptual processes, demolishes the problem of the secondary qualities.

Summary

I have sought to establish the following main points:

I. A crucial aspect of mind is captured in the notion of an "internal model", and that such a model is not adequately representable in colloquial language. A more adequate (but still inadequate) representation would use a programming language which (even in describing a robot) uses intensional descriptions, but does so in a sufficiently detailed way to allow causal explanations which possess extensional translations.

II. 'Utterances' are but samples of the states of such an 'internal model', and thus share with all samples the risk of being representative. We then ask what aspects of 'mental states' are sufficiently germane to action to merit 'awareness', and suggest that the ability to verbally express them is then to be seen as a secondary property.

III. The slide-box metaphor is used to suggest that the internal representation of an activity within the brain is correlated with the spatial extent of the object it represents, and that this correlation demolishes the problem of the secondary qualities.

Footnotes

1 London; Routledge and Kegan Paul, 1969. All quotations in this paper are taken from Dennett's book, with the indicated paginations. Dennett and I have agreed to drop the distinction between Intention and intension which he tried to maintain in his book.

2 The need for "resolution of redundancy of potential command" is but one of many principles I have set forth in "The Metaphorical Brain", (New York: Wiley Interscience, 1972) as required for the analysis of mind, be it manifested in the brain of a human or the control-computer of a robot. Related principles are that our approach to perception must be action-oriented, that perception is not only of "what" but also of "where", and that an adaptive system must be able to correlate sensory data and actions in such a way as to update an internal model of its world. A critical role of perception, then, is to update this internal model, in addition to its role in controlling action.

3 In Chapter 4 of "The Metaphorical Brain", I indicate in some detail how a robot's control-computer may be so programmed that it can process input from a TV camera to represent its environment in terms of objects and their positions. It can then respond to a command such as "Push the red cube to the door" by computing a path which will take it to the red cube and then push the cube to the door without bumping into any obstacles, and then issuing the series of instructions to the drive wheels of the robot that causes it to follow the path. The point is that programmers find it convenient to initially describe the programs for the control of the robot in intensional language and then simply translate these high-level programs into machine language - to obtain the detailed programs that actually cause the robot to move efficiently in a wide variety of novel environments.

4 See "The Metaphorical Brain", Section 4.2, for further information on heuristic search.

5 The obscurity of the two previous sentences to any philosopher untrained in computer science may perhaps underscore the dangers of basing a philosophy of mind upon common language usage.

6 For further confirmation of this point, see the discussions of "output feature clusters" in "The Metaphorical Brain".

7 See, for example, Section 3.4 of my "Theories of Abstract Automata" (Englewood Cliffs, N.J., Prentice-Hall, 1969).