What has language to do with perception? Some speculations on the Lingua Mentis.

Zenon W. Pylyshyn

partments of Psychology and Computer Science
University of Western Ontario
London, Canada

1. Introduction.

The topic under consideration in this conference session (viz. Language and Perception) is not the one to which the greatest amount of attention has been devoted in philosophy of mind and philosophy of language. There a major concern has been the relation between language and thought. As everyone knows there has been a long standing dispute regarding whether or not it makes sense to view thought as being carried out in the medium of natural language or whether some other form of representation is involved. There has not been a comparable dispute over the relationship between perception and language. For one thing no one to my knowledge has proposed that perception occurs through the medium of natural language (though some early behaviorist writings come close, especially in respect to memory for perceptual events). What I propose to consider in this brief note are some respects in which the language-thought relationship is similar to the language-perception relationship.

2. Language and Thought.

At least since Aristotle there has been speculation and argument concerning the form (or language) of thought. Many contemporary philosophers (e.g., Quine, Sellars, Harman) as well as some past students of language (e.g., Whorf, Humbolt) believe that we think in our "outer" natural language: that knowing a language is being able to think in it. Harman (1975) takes a sophisticated approach to this position. He argues that in thinking in one's spoken language one need not parse or disambiguage it--since that would get us into the vicious circle of having to parse the thought into something which itself would be a thought and hence in need of further analysis. In Harman's view our thoughts are carried by "sentences under analysis" or by ambiguity-free already analysed sentence structures (e.g., P-markers). One problem with this view is that it denies the possibility of thought in animals and pre-verbal children. Other difficulties were recognized by psychologists. In the beginning of this century the Wurzburg school was able to argue that much of our thinking was unconscious. A more modern view (e.g., Paivio, 1975) takes the conscious experience of thoughts as occurring in language or in imagery as its starting point and demonstrates by operational means that at least two distinct modes of thought need to be postulated. This "dual code" approach is quite widely held in psychology although it is not precisely clear what intrinsic properties are being claimed for the imagistics mode of thought. But more on this later.

My own view, which I have been espousing for some half dozen years, is that an adequate account of the process underlying thought will show it as occurring in a symbolic mode which has few of the properties we would normally ascribe to either natural language or to images. For example, the vehicle of thought does not require words (but only concepts) nor does it have such intrinsic properties as size or shape. Rather it consists, as do all computations, of the transformation of formal symbolic expressions whose terms are given an intentional interpretation by the theoretican. In other words, thought is a symbol manipulation process. Because the data structures representing thoughts have an implicit syntax and because its terms and composite expressions are interpreted, one can think of them as expressions of an internal language-or lingua mentis--call it "mentalese".

While the particular arguments and examples I have presented in support of this position have varied over the years the thrust of the arguments has always been a two-pronged one. On the one hand I maintain that criteria of explanatory adequacy require one to give an account of certain specifically cognitive phenomena in a manner which neither presupposes certain crucial properties which themselves require a cognitive explanation, nor avoids a complete process explanation (involving a reduction to primitive mental operations) by attributing certain phenomena to intrinsic features of the brain. On the other hand, the argument has always appealed to empirical evidence. It is the dual requirement of meeting explanatory criteria and empirical evidence that has, for me, been the basis of my rejection of specific imagistic models such as those of Paivio, Kosslyn, and Shepard.

This is obviously the wrong forum in which to continue this debate especially since many of the details are peripheral to our present concerns. However, I do want to elaborate very briefly on what I referred to above as criteria of explanatory adequacy since I believe that this is the real crux of the debate, not only over imagery accounts of thought but also over some of the

issues about language and perception I want to raise later. Further details can be found in Pylyshyn (1978a).

The issue about explanatory adequacy is the following. Positivist doctrine not withstanding, an explanation of a phenomenon has to do more than predict or duplicate aspects of the phenomenon. It must also explicitly characterize the properties of the system by virtue of which the observed (or predicted) behavior occurs. Since some of these properties are adventitious or ad hoc while others are principled, such a characterization is essential. Furthermore, the account must separate *properties which are fixed and universal from those* which vary from task to task. To use an analogy from logic, it must separate the contribution of the notation, the logical axioms and inference rules from the particular premises used in deriving entailments. In the case of a process theory it is not sufficient to simply provide a procedure which generates behavior similar to that observed in humans. We must, in addition, explicitly isolate those properties and mechanisms which will remain fixed over all cognitive processes (the underlying system architecture), those which can vary gradually with learning or accomodation but whose component parts and intermediate states are not available to the whole system (the compiled skills), and those which represent particular methods adopted for particular tasks or which represent particular knowledge which the system possesses (and thus which can change freely). Furthermore, this parametrization or attribution of behavior to separate sources must be individually empirically justified--e.g., we must show that it is reasonable to postulate such properties of the architecture as we do by appealing to empirical evidence. If we can in this way isolate the fixed properties and show how these can be combined to produce the observed behavior, then we would have an account of the behavior which refers it to both fixed universal properties and to particular task specific ones. Such an account would not only capture cross-task generalizations but it is the best we can do from a cognitive or functional point of view. Further explication would involve describing, for example, how the fixed properties are realized in neural tissue or how and why the variable aspects got to be the way they are given the nature of the environment-organism interactions. An account partitioned this way would provide a means of deducing current behavior from fixed universal properties of mind and hence would provide a basis for explanation.

My main objection to such notions as analogues and to such hypothesized mental operations as scanning and rotation (to cite just two) is that the empirical evidence does not support the position that these are primitive properties of the mental architecture. I have argued that in all the proposals I am aware of which postulate analogues or analogue-like operations on images, there is independent evidence that the phenomenon in question must be attributed, at least part, to tacit knowledge which the system or person) possesses or to more articulated and piece-meal processes than those claimed. In other words these analogue operations cannot be taken as explanatory primitive operations in the mental architecture. Consequently to explain the experimental findings that these terms were introduced to account for we are forced to show how they could be carried out in an architecture in which scanning and rotation are not primitive operations. In such an architecture the processes might be quite different (e.g., while there might be a subroutine that accomplishes scanning or rotation, these particular terms would only be descriptive and not explanatory since the functions implied by them would in turn have to be explained in terms of more detailed computations using other more primitive and independently justified operations). The exact form of the argument against the hypothesis that scanning or rotation are primitive operations in the fixed mental architecture can be found in Pylyshyn (1978a, 1978b). Essentially they depend on showing that certain empirical facts (e.g., that rate of rotation depends on properties of the figure, the probe, and the task in general) require for their explanation that we specify more detailed processes which carry out the function described as rotation or scanning, thus demonstrating that the function was not a primitive.

The general conclusion I draw from these arguments is not that talk of analogues or other non-symbolic systems is incoherent or logically ruled out, but only that none of the phenomena which people typically appeal to have been shown to require them--and even if they were admitted they would, at least in these instances, not be explanatory in the required sense, though they might well be predictive (but then so would a multiple regression equation). Within the information processing paradigm (i.e., excluding phenomenological or purely neurophysiological explanations for reasons which we cannot go into here) the only remaining candidate paradigm for explaining the nature of thought is computation, in the sense of transformations on symbolic expressions. Of course within this alternative we may still posit different symbols, and even different composite data structures for different areas of cognition. What I am saying, however, is that this most basic level of symbolic representation is the modality independent medium of thought, the "mentalese" in which goals, beliefs, hypotheses, knowledge, and other cognitive states are expressed.

What makes this point of view on the relation between language and thought relevant to the perception-language discussion is that mentalese is not only taken to be the form in which thoughts are carried. it is also proposed as the appropriate representation of percepts.

3. Language and Perception.

Before discussing the similarities between the language-perception relation and the language-thought relation it may be useful to consider why one might be motivated to ask about the relation between language and perception in the first place. An obvious connection between the two is the fact that we can talk about what we perceive. But that tells us little about how the two are related. We get hints that the relation may be more intimate from the widespread use of perceptual terms (especially spatial relation and movement or transfer terms) to refer to abstract relations in general. The experiments on imagery by people like Shepard, Kosslyn, Moyer, Paivio, and others show,

if nothing else, that perception and thought are closely related. Thus the issues raised in discussing language and thought become relevant here too.

But perhaps one of the main reasons why language and perception are inextricably related is that the perceptual system is the primary means through which language acquires a semantics. A system which contained a body of data and a language processor might conceivably be able to carry on a coherent dialogue. But without a perceptual component it would, in an important sense, not know what it was talking about. We could, in principle, change the ASCII coded strings in its lexicon and it might conduct an equally intelligent conversation on an entirely different topic without anything (other than the external tokens) having changed. This is possible because the only constraints in the system are intra-linguistic ones and hence only linguistic and data-base consistency can be detected. In such a system there is no correspondence between internal symbols and things and hence the system makes no reference to the world.[1] This argument is made with painful force by Fodor (in press). It would be more obvious to people in A.I. that this is indeed the case if they heeded McDermitt's (1976) suggestion and refrained from using English words and phrases inside their programs and only employed nonsensical atomic symbols (GENSYMS). In that case it would be clear that only the programmer (and not the program) knew what it was talking about.

There is in fact a general and largely ignored problem of the distribution of explanatory burden between program and programmer that needs to be explicitly acknowledged in discussions in which programs are presented as theories. I have come to realize over the years that any crackpot theory can be implemented on a computer in some sense or other simply by assigning the appropriate names to various things in the program (e.g., call this buffer "consciousness", that data structure an "image" and this procedure "the mind's eye"). Elsewhere (Pylyshyn, in preparation) I have suggested a number of ways in which some of the arbitrariness can be taken out of this enterprise. They include independent validation of the "fixed mechanisms" that are to serve as the primitive components out of which cognitive processes are constructed (what I called the mental architecture) and the independent provision of at least a partial intrinsic semantics for symbols in the system by relating them to perceptual and motor subsystems. A further step might also be to provide the system with a learning component (in the very general sense of a history-dependent relationship with an environment) which would also serve to constrain the interpretation of symbols by connecting it to the physical world through a historical causal chain (c.f., Kripke's, 1972, causal theory of reference).

Now if we accept that in order for a system to have a semantics, as opposed to merely a complex intra-verbal deductive system operating on uninterpreted symbols (or "logical forms"), it must at least have a perceptual component, a number of fundamental questions arise. Though the whole issue of semantics is fraught with difficulty I will take advantage of the invitation to speculate by rushing in where many have been lost. The questions I shall in a sketchy way comment on concern the nature of the perception-language correspondence, the way in which this correspondence might be represented, and how such a correspondence could arise in the first place.

3.1 The nature of the language-perception correspondence.

Since the set of perceptual patterns and the set of definite descriptions are both unbounded, the correspondence between the two cannot be through existing associative links. The mapping can only be given by a recursive procedure which associates subpatterns of the language with subparts of the percept--in other words the correspondence is between some analysis of both descriptions and percepts. We are of course no more aware of the conceptual analysis of percepts than we are aware of the analysis of linguistic inputs. Given the necessity of an analysis of both, the most parsimonious story of how this occurs is one which assumes that both are analysed into a structure in the same interlingua--viz., mentalese. Contrary to some of my critics on this point (Kosslyn & Pomerantz, 1977, Anderson, 1978) such a view is neither inconsistent nor unnecessarily complex. Independent arguments suggest that at least this much analysis or translation is necessary and there is, to my knowledge, no convincing argument that more than one form of interlingua is needed. Though this latter possibility is not ruled out, the relatively weak constraints placed on the formal properties of the representing medium at present (viz., that it consist of sumbol structures) make this possibility seem unlikely. Furthermore, the freedom we have in thinking about information received through all modalities and the readiness with which we forget (outside of experimental settings) how we came to know something argues that at least memories and thoughts might appropriately be viewed as being amodal.

Another question that arises in connection with the nature of the language-perception correspondence is whether the formal properties of the two are independent or whether one might be able to explain linguistic properties in terms of perceptual or general cognitive ones and vice versa. Such a possibility is most attractive since it would increase the explanatory power of the resulting theories. On the other hand, there is no a priori necessity that such an explanatory link exist. As Chomsky (1975) has frequently pointed out we do not expect to be able to explain why humans have certain physical characteristics (e.g , why they have 10 as opposed to 8 toes, etc.) so why should we expect to explain why the noun-verb dichotomy appears to be a linguistic universal. Still one might be permitted to hope for some economy of explanatory principles by unifying over cognitive domains.

There is already reason to believe that at least some of the lexicon can be explained in terms of universal properties of perception Perhaps the clearest and most familiar example is the case of color terms. Berlin and Kay (1969) have demonstrated that color terms in various cultures form a strict hierarchy so that languages with more color terms invariabley include the terms used by languages with few color terms. In this example, however, it has been

possible to go further and demonstrate universal color perception properties paralleling the linguistic findings and even to relate these to visual physiology. Denny (in press) has cautiously suggested the possibility of a similar hierarchy across cultures of lexical systems for spatial deixis. For example, compared to English's two terms "here" and "there", Kikuyu has 8 spatial deictic terms and Eskimo has 88, all forming an inclusive hierarchy.

It is not inconceivable that the structure of the lexicon will exhibit many such points of contact with perception--at least for concrete descriptive terms. Is there any reason to believe that this parallel might also hold for other parts of language--specifically for grammar? There have been suggestions that syntactic classes such as noun or verb or even adjective correspond to conceptual categories--to ways of conceptualizing the named entitities. There have even been occasional suggestions that grammatical rules are a reflection of how people conceptualize what they perceive.

We must be quite clear about what such claims can mean There is a sense in which these claims are very likely (but perhaps not too interestingly) true. For example, when I choose to say "that's a red ball" as opposed to "the color of that ball is red" it seems reasonable that I select a part of speech and grammatical form which highlights certain aspects of what I intend to assert. Grammar provides many options on how essentially the same propositional content can be asserted. These alternatives may differ in respect to which items are treated as figure and ground (or topic and comment). Which option we take on a particular occasion no doubt depends, at least in part on how we conceptualize the situation This, however, is very different from the claim that grammatical categories represent conceptual categories. Even less does it suggest that syntactic rules can be expressed in terms of conceptual properties In spite of considerable effort devoted to the problem no one has, to my knowledge, provided even a glimmer of hope that any particular grammatical rule of language bears anything but a conventional relation to things in the perceptual field. It is as though syntactic structure provides a sort of system of codes which can be exploited to carry conceptual distinctions even though the system of codes itself is independent of what it can be used to express. In fact the linguistic code is rather severely constrained by properties of the communication channel into which it encodes ideas, for example by the serial nature (i.e., low bandwidth) of our speech and hearing apparatus in contrast with the richness of our conceptualizations an our perception in general.

Since, however, language is in all likelihood a function of the same cognitive apparatus as is available for other cognitive domains, we might expect an influence to be apparent at some level--even if not at the level of rule structures. For example, if figure-ground organization was a primary mode of structuring perception and thought one might expect syntactic features of some kind to be used consistently to reflect this organization--even though the code could in principle also be used to represent quite a different type of conceptualization or the same conceptualization in a different way. Thus, it is entirely conceivable that some predicate-argument type of characteristic might be found in grammar, whether represented as a surface taxonomy or some less obvious way. Whether or not this is the case is an empirical question in respect to which I don't believe there is wide agreement at present.

When it comes to more abstract properties of language, such as some of the putative linguistic universals, I believe the possibility of showing parallels between language and other areas of cognition may be more hopeful. My rather tentative view on this is based on the belief that whereas the form of grammar may well be an unexplainable consequence of some properties of brain structure together with properties of channels of communication, sentence comprehension must be implemented on a system with the same architecture as that used in other areas of cognition. Consequently, there may be some very general processing constraints that might show up as linguistic universals. In any case, if they appear in linguistic data at all the effects of system architecture will be seen in abstract universals rather that particular language specific syntactic rules.

For example, one very general universal property which Chomsky (1975) has cited as evidence for the innateness of Universal Grammar is that of "structure dependent rule" Rather than infer the apparently simplest rule (or the rule whose features are most evident on the surface of the set of samples) the child infers more complex structure-dependent ones. For instance, rather than infer that declaratives and questions are related by virtue of a certain pattern of permutation of substrings of the sentences, the child learns that the permutation applies over an analysis of the sentence into abstract phrases. Thus, while the simple rule accounts for the relation between "The man is tall" and "Is the man tall?", this would produce the incorrect transformation of "The man who is tall is in the room" as "Is the man who tall is in the room". Yet children never make the latter error, thus suggesting that their hypothesis formulation capacity is constrained in ways characterized by Universal Grammar.

But structure-dependence is not only a phenomenon of language, it is also ubiquitous in perception. Even a casual examination of what is involved in visual tasks, such as the solution of geometrical analogy problems, makes it clear that the rules employed must be sensitive to various level of abstract structure as opposed to more superficial features of the figure. In fact it is characteristic of all of perception that the structuring of the perceptual field must be hierarchical. If we were to describe what a child learns in learning to perceive its world we would come to the same conclusions about vision as Chomsky does with language--viz., that the way in which the regularities of the visual field are captured is constained by innate mechanisms in a way which would be described as "structure dependent".

There have also been attempts to explain more specific linguistic universals--such as the Specified Subject or Subjacency constraints--in terms of general properties of the processor (e.g., Marcus, 1977). Such studies are only beginning but I have no doubt that some linguistic properties will eventually be attributable to architectual or strategy properties unique to the human cognitive

system. How much will be explainable this way remains an open question.

3.2 Representing semantics.

The much misused term "semantics" refers to the interpretation of a symbol system (in this case language) into some other domain. In a computer without a perceptual component the only symbols which strictly speaking have a semantics are ones which are either directly executable by the hardware or are translated into other symbols which are executable.[2] All other symbol structures which are referred to as semantic are really supports for the deductive apparatus. They simplify the process of deducing new expressions from old ones in such way as to maintain the truth of the expressions under a consistent interpretation. This interpretation, however, is provided by the user, not the system.

Often what is referred to as the semantic representation has some of the properties of a model. For example, it provides a set of objects which can be used to evaluate expressions, the way models are used in mathematics. In a sense then, these models form a domain of interpretation. They are not, of course, the ultimate intended domain of interpretation. Expressions are typically intended to refer, for example, to beliefs about objects in the real world, not to other symbols. But this formal model can itself be taken to represent such cognitive objects and so provides a formal semantics for the symbolic expressions which hopefully is valid in the intended domain. The design of such formal models is a major concern in A.I. and the computational version of such systems are typically hybrid mixtures of models and inference schemes. I will have very little to say about them here.

In a system which does contain a perceptual component there has to be some facility for translating between the perceptual analysis and the linguistic analysis. In order to deal with the "semantic content" of sentences and percepts we must provide the potential for cross-modality and extra-linguistic correspondence. I have suggested that the most parsimonious view of how this occurs is that the end products of both perceptual and linguistic analyses are conceptual structures, or expressions in a single symbol system which we call mentalese. Other alternatives are occasionally proposed. We shall very briefly examine one below.

There have sometimes been objections to the view that percepts are conceptually analysed into articulated symbol systems. Some people feel that this loses the holistic and continuous aspect which seems intuitively to characterize percepts. It is hard to know what to make of such intuitions. They seem to suggest to people something more than that we see distributed features (e.g., roundness) or continuous properties and therefore that the percept must represent such properties. Rather these intuitions seem further to suggest that the percept must <u>have</u> such properties--i.e., it must not only represent the property of continuity but it must actually be continuous. This is a dangerous direction to pursue, however, since it could lead one to also claim that percepts actually are large, blue, warm, heavy, etc., running us right into Leibniz's problem.

The only proposals I have seen for dealing with the holism concern are ones which propose unanalysed objects such as templates or holograms as perceptual representations. These are not only atomic wholes but are clearly relatable to the proximal stimulus, at least in the case of vision.

I have discussed such proposals elsewhere (Pylyshyn, 1974; 1978b). Their inadequacy stems from several sources. One is that by considering the percept to be holistic in this sense one loses the ability to attend selectively to parts or aspects of it or to notice the respects in which two such representations differ. Of course, one can gain this facility back by positing a process of comparison or analysis which yields the more detailed features--but this is just to postpone the translation into mentalese. Alternatively one might posit that the comparison itself is done by a non-symbolic holistic process like that used in matching holograms. But here we run into trouble with the sheer empirical facts concerning the cognitive structure of percepts. The type and degree of perceived similarity among stimuli cannot be matched by a uniform interpretation-independent process like the hologram one. To what extent and in what respect two things are perceived to be different depends entirely on what we perceive those things to be. In other words similarity must be defined over an already interpreted --and hence conceptual, nonuniformly detailed, preanalysed, and articulated--representation.

Even a compromise in which the representation is an articulated structure with something like "imagoids" or pieces of templates at its nodes will not help. For if those template pieces need in some cases to be further analysed then we are back with the problems sketched above. If, on the other hand, they do not need to be analysed then there is no distinction between this proposal and one in which the templates are replaced by atomic symbols--i.e., terms in the mentalese vocabulary. Recall that mentalese terms appear in the output from the perceptual system and thus can arise from such perceptual properties as "large", "round", "red" or ones for which there is no single word in English, such as "sand-like texture" or ones best displayed graphically. What mentalese terms there are--i.e., what well-formed perceptual categories exist--is an empirical question.

Whatever merits the proposals for imagistic or analogue representations may have they clearly do not help the language-perception interface problem since sooner or later the representation must be analysed in such a way as to be commensurable with natural language terms. Whether this is done at the time of perception, or postponed by storing an unanalysed proximal stimulus so that it must be done at the time of sentence generation, does not affect the basic problem. Other independent considerations, discussed in Pylyshyn (1973, 1974, 1978b), argue against the view that unanalysed stimulation is stored in memory.

3.3 The genesis of the language-perception correspondence.

In an earlier paper I noted three major preconditions for learning a language (Pylyshyn, 1977).

1. Sensory experience must be structured. The "blooming, buzzing confusion" of William James must be susceptible to segmentation, analysis, and re-

construction. Some aspects must be foregrounded relative to others so that the environment becomes articulated or differentially noticed in some fashion.

2. Communication codes (both verbal and nonverbal) must likewise be structured The stream of vocal or gestural behavior must be perceived as segmented and a distinction between signifying and nonsignifying variation must be made (in generation and/or perception).

3. The occurrence of a speech act must be recognized. This is perhaps the most important but most neglected aspect of preconditions for language acquisition Not only must a child attend to the appropriate aspects of his environment, but he must do it within the context of what Merleau-Ponty would call (loosely) an "intention to mean".

In this section I wish to deal primarily with the first of these preconditions and with what has to happen in order for a simple naming or describing correspondence to occur. I will not dwell on the other two preconditions except to note that, as the third precondition suggests, a simple associative pairing will not make one perceived pattern (e.g., a word) refer to another. The pairing must be conceptualized and subsequently treated as a particular kind of asymmetrical irreflexive relation called naming or reference. This in turn means that one pattern (e.g., a word) is not simply an indicator that, say, the other pattern is about to appear but rather becomes a symbolic surrogate for its referent. It can then be used in arbitrary cognitive combinations with other such surrogates. It can be used not only instrumentally to anticipate or to ask for objects, but also to think about, hope for, question, assert something about, plan for, and vicariously play with the designated object.

What I would like to consider in a general way is how a linguistic sign or word can come to refer to something in the perceptual field Take the simple example of naming by ostention. A child is shown a dog and the word "dog" is uttered. Suppose the preconditions are fulfilled. The first problem to be faced is the well known difficulty of how the child is to know that what is being pointed at is the object rather than any of its properties. Alternatively, how is the child to know whether the word refers to that very object lying on the carpet with a collar around its neck and a bone in its mouth or any member of the Cocker Spaniel family or any canine or mammal or living creature, and so on.

First of all it is clear that what the speaker is referring to must be a conceptually integral unit for him--something he can conceptually detach from his cognitive or phenomenal field Secondly, if the hearer is to have any chance of acquiring the same referent for that word he will also have to to have conceptualized the field in such a way as to individuate the same entity as the speaker. Given the unlimited number of in-principle possible ways of analyzing the entire ostention situation, nothing short of a miracle could ensure that the same analysis was given by both participants. Nothing, that is, except a highly constraining universal innate mechanism that severely limits the set of alternatives which are humanly conceivable.[3] What this in turn comes to is the claim that the terms of mentalese are innate This outrageous claim, which is argued for in considerable detail by Fodor (1975), is also pressed on us by other considerations which we take up below. Thirdly, the listener must use both his perception of the physical situation and his understanding of the social context to infer the intentions of the speaker. This gives definition-by-ostention a problem-solving character.

John Macnamara (1972) has revived interest in the view, often associated with St. Augustine, that "...infants learn their language by first determining, independent of language, the meaning which the speaker intends to convey to them, and by then working out the relationship between the meaning and the language (p. 1)." In other words the child has various sources of evidence concerning such things as what objects, classes and properties are in his environment and what the adult intends to convey, say, by pointing and speaking a word. His task is then to make the inference to the best hypothesis concerning the correspondence between these events But the question arises, how is the hypothesis formulated? Clearly this view assumes that the relevant aspects of thought and perception (my first precondition) are present prior to language learning. This in turn presupposes that the terms of mentalese are also available prior to language learning since the hypothesis must be expressed in mentalese. But how then is mentalese acquired?

The answer is that if it is "acquired" at all no one has the slightest idea how this could possibly occur. The only notion around (as Fodor, 1975, has argued) regarding how a new concept (or term of mentalese) could be learned is one which says that what people learn is the relation of the new concept to some relational structure of already known concepts. But this precludes the learning of any concepts that are not definitional composites of old ones, and therefore strictly eliminable. Unfortunately, this appears to include most natural concepts Like the theoretical terms in science, most natural concepts cannot be given a context-free definition but rather depend on the entire system of concepts for their meaning (which is why dictionary definitions are invariably circular). While one can speak of the accomodation of linguistic usage (e.g , the referents of words can vary as we discover new empirical facts--such as that both steam and ice are really just forms of water), the accomodation of the mental concepts, in terms of which the linguistic terms can be understood, remains a mystery. The mystery is not lessened, moreover, by talk of motor schemata or "equilibration" as Piaget does. In each case of putative conceptual change the process either depends on assimilating new concepts into arrangements (or schemata) made up of old concepts, thus severely limiting the type of conceptual change possible, or it is left unexplained There is no explanation, nor even the beginnings of an approach, for dealing with the accomodation of schemata or conceptual structures into ones not expressable as definitional composites of existing ones. There is, in other words, no inkling as to how a completely new non-eliminable concept can come into being.

This is in fact an extremely deep problem about which very little sense has been made. People are sometimes mislead by certain computational metaphors into believing that the problem can be dispensed with by something like compilation. But however attractive that

notion is, as a way of talking about how new procedures can come into being which are themselves expressed in terms of new operations, it does not generalize to concepts in general Such a notion works in the case of procedures because the set of computable functions is closed and reduceable to elementary (Turing machine) operations in a way that the set of conceptualizations of the world is not.

It seems to me that there are two general avenues open for dealing with this dilemma, both of which simply raise more questions than they answer. In both cases what we are doing is opting for a different locus for the mystery, rather than resolving it.

The first approach is to simply accept what seems an inevitable conclusion and see what it entails. This is the approach taken by Fodor (1975) who simply accepts that mentalese is innate This means accepting that virtually all unitary concepts of which we are capable are genetically determined. Compound concepts (such as circular red object) can also be constructed as well as definitional composites, but these constitute a minority of our mentalese vocabulary. Of course, there need not be (and in fact certainly will not be) a one-one correspondence between concepts and words in the spoken language. It is quite likely that most words do correspond to concepts, though there have been suggestions that some words are represented by compositions of more primitive concepts (e.g., kill = do something to cause to die; never = not ever). So far few, if any, of these suggestions have withstood empirical tests (c.f., Fodor, Fodor, and Garrett, 1975). Clearly, however, not all mentalese terms correspond to words. Not only do societies differ in their basic vocabulary but the view of mentalese we have been discussing requires terms for stable perceptual features which are not encoded in our language, at least not as single words

While the notion of all our concepts being innate is repugnant to the contemporary Zeitgeist, part of this attitude may be due to the connotations of this way of speaking. If we thought of the innate mentalese vocabulary as corresponding to the fixed structural properties of the computational system, together with the input-output transducers, this might not seem as distasteful. Even the simplest modern computer has a considerable amount of fixed hardware (i.e., innate) structure--including a facility for discriminating an unlimited number of formal atomic symbols. If each of these symbols had predetermined potential referents (say by virtue of the way they were wired to mechanisms which were eventually connected to transducers), they could be considered innate concepts. Of course this is not the whole story since it is hard to see how many of the required concepts (e.g., Kant's transcendental categories such as space, time and cause) could be thought of as wired to transducers. The problem here is that it is still not very clear what the force of the claim is when we say that concepts, qua interpreted symbols, are innate Conceivably it could mean little more than that the constraints on the system of symbols is so great that the class of possible interpretations (like the class of realizeable grammars) is extremely limited. In fact one way that the class of possible interpretations could be characterized might be to formulate them in terms of the requirement that the only concepts the organism can hold are ones expressable in terms of a certain "innate vocabulary". In that case, "innate vocabulary" has the same status as "universal grammar"--viz., they both somehow characterize the endowed cognitive capacity of the organism.

This approach to the innateness dilemma places the puzzle of conceptual development on a different mechanism from the usual one of concept learning. Now the problem becomes, given that most of the concepts are innate why do they only emerge as effective after certain perceptual and cognitive experience and at various levels of maturation?

Another approach to this dilemma is to locate the puzzle in yet another quarter. We think of the "innate concepts" as being the representational capacity of the fixed hardware architecture--so that mentalese becomes identified with machine language. The innate concepts are thus not truly concepts but, as suggested above, symptoms of the interpretive constraints imposed by the computational architecture on the system of available symbols. Now the symbols do have to be exploited in representing the world, and for any particular machine architecture their interpretability is constrained in certain ways For example, if a certain subset of available atomic symbols is treated in a certain way by the motor transducer (e.g., cause the hand to open or the arm to reach out) then they cannot consistently be interpreted as, say, referring to phonemes

Now the problem we had was to explain how new concepts can develop which are not definable in terms of old ones. This is the essense of radical conceptual change or accomodation. The paradox arose because the only formal mechanism which seemed to be available was symbolic composition (or definition). A whole new realm of possibilities opens up however if we allow non-symbolic changes to occur--i.e., if we allow the actual hardwared connections or architecture to change Concepts can then drift or mutate insofar as the constraints on symbols can change in novel ways.

The trouble with this proposal, of course, is that it is nothing more than a burying of the problem into hardware. So long as the relation between hardware and symbolic levels is not systematically understood--so that, for instance, we had some formal rules for how the underlying architecture could change in response to programmed instructions --then this proposal is not a real alternative. It does, however, contain one recurring suggestion which seems to surface in many different contexts and for many different reasons (most, in my view, are invalid)--viz., that there are some cognitive functions whose realization will require that we transcend the symbolic mode and deal with physical (or, at any rate, a quite different set of symbolic) processes. Maybe that's what Kant had in mind when he spoke of "transcendental reasoning".

Footnotes

1. The fact that a system without instrinsic semantics could conceivably still pass the Turing test and meet Newell's criterion for

understanding (viz., "S understands knowledge K if S uses K whenever appropriate") suggests that such criteria may show that there is a difference among (a) achieving "understandin" (b) knowing what things, properties, etc. in the world are being referred to, and (c) explaining what such understanding consists in, or what it means to comprehend on utterance. As noted earlier, criteria of performance are distinct from criteria of explanation.

2. Even numerals are not interpreted by the machine. The transformations of numerals into numerals carried out by what are called arithmetic commands are just formal operations on symbols. The user typically interprets the symbols as designating numbers and the operations as designating the usual arithmetic operations but he could just as well interpret the symbols as, say, propositions and the operations as deductions (though the interpretation function might be quite complex)--or any other interpretation which happens to maintain its coherence.

3. It is understandably not easy to provide an example of a humanly inconceivable unitary concept. Goodman's "Grue" and "Bleen", introduced to highlight certain problems of induction, may be such examples. Grue is the unitary concept which in English corresponds to the color description "Has a green color up to time t and a blue color after". Thus in the new system green would be the name given to that strange color which is Grue up to time t and Bleen afterwards So far as anyone knows, concepts like Grue and Bleen never occur in human cultures. However we must not be too presumptive about what concepts actually can exist. Exotic societies frequently provide examples of what are for us inconceivable ways of carving up experience. For example Foucault (1972, xv) quotes Borges' citation of an ancient Chinese encyclopedia which has the following strange taxonomy. "Animals are divided into (a) belonging to the Emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies." If very strange concepts do exist we might find it very hard to decipher them, given our constrained schemata

References

Anderson, J. R. The status of arguments concerning representations for mental imagery. Psych. Review, in press.

Berlin, B., & Kay, P. Basic color terms. Berkeley: Univ. of California Press, 1969.

Chomsky, N Reflections on language. New York. Pantheon, 1975.

Denny, J. P. Locating the universals in lexical systems for spatial deixis. Papers from the 14th regional meeting of the Chicago Linguistics Society, 1978, in press.

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

Fodor, J. A., Tom Swift and his procedural grandmother. Cognition, in press.

Fodor, J. D., Fodor, J. A., and Garrett, M. F. The psychological unreality of semantic representations. Linguistic Inquiry, 1975, 6. 515-531.

Foucault, M. The order of things. London: Tavistock publication, 1970.

Harman, G. Thought. Princeton, N.J.: Princeton Univ. Press, 1973.

Kosslyn, S. M., & Pomerantz, J. R. Imagery propositions and the form of internal representations. Cognitive Psychology, 1977, 9. 52-76.

Kripke, S. A. Naming and necessity. In D. Davidson and G. Harman (eds.), Semantics of natural language Dordrecht D. Reidel, 1972.

Macnamara, J. Cognitive basis of language learning in infants. Psych. Review, 1972, 79, 1-13.

Marcus, M. P. Theory of syntactic recognition for natural language. M.I.T. Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, 1977.

McDermitt, D. Artificial intelligence meets natural stupidity. Newsletter of the ACM Special Interest Group on Artificial Intelligence (SIGART), 1976, 57, 4-9.

Paivio, A. V Neomentalism. Canadian Journal of Psychology, 1975, 29, 263-291.

Pylyshyn, Z. W What the mind's eye tells the mind's brain. a critique of mental imagery. Psych. Bulletin, 1973, 80, 1-24.

Pylyshyn, Z. W. The symbolic nature of mental representations. Paper presented at a conference on Objectives and Methodologies in Artificial Intelligence, Canberra, Australia, May, 1974 (mimeo)

Pylyshyn, Z. W. What does it take to bootstrap a language. In J. Macnamara (ed.) Language learning and thought. New York: Academic Press, 1977.

Pylyshyn, Z. W. The explantory adequacy of cognitive process models. Paper presented at a workshop on mental representation, M.I.T., January, 1978a (mimeo).

Pylyshyn, Z. W. Imagery and artificial intelligence In C. Wade Savage (ed.). Perception and Cognition: Issues in the Foundation of Psychology, (Vol. IX of Minnesota Studies in the philosophy of science). Minneapolis, Minn. University of Minnesota Press, 1978.

Pylyshyn, Z. W. Towards foundations for Cognitive Science. Book manuscript, in preparation.