

abstract sort. And it is not easy to see how any neurological mechanism that might account for the relevant behavior and that could be interpreted as an instantiation of one of the descriptions could not also be interpreted as an instantiation of the other. Therefore, I see no reason not to accept the incorporation view.⁶¹

⁶¹ I discuss these issues further in *Thought* (Princeton, N.J.: Princeton University Press, 1973).

Knowledge of Language

There are a number of different questions that I would like to touch upon in these lectures, questions that arise at various levels of generality and that grow out of different, though not totally unrelated concerns. I want to present a certain framework within which, I believe, the study of language can be undertaken in a very fruitful way — not the only framework, to be sure, and one better suited to certain problems than to other, equally legitimate ones. Within this framework I would like to discuss some technical questions that are at or near the borders of research. At this level of discussion, I will be presenting some material that is internal to the theory of transformational generative grammar. But I would also like to suggest that this rather technical material is potentially of quite general interest, that by studying it we can hope to learn some important things about the nature of human intelligence and the products of human intelligence, and the specific mechanisms that enable us to acquire knowledge from experience, specific mechanisms that, furthermore, provide a certain structure and organization for, and no doubt certain limits and constraints on, human knowledge and systems of belief. I think that the work I will describe at least hints at a concept of man that is rather different, in interesting respects, from others that have been implicit in much modern thinking about these matters, and would like to elaborate on this question as well. I have in mind, then, a large enterprise, of which only a small part can be carried out with a satisfactory degree of clarity and precision. Still, I think it is useful to consider this small part against the background of what might ultimately be achieved.

The study of language, as I will be considering it here, can be regarded as a part of human psychology. It forms a part of the general study

AUTHOR'S NOTE: This essay is the first of six John Locke Lectures delivered at Oxford University in May–June 1969. Parts appeared in the *London Times Literary Supplement*, May 15, 1969. The other lectures have not yet been published.

of human cognitive structures, how they relate to behavior, and how they are acquired. This point of view is not at all characteristic of study of language in the modern period. In fact "psychologism" has been often stigmatized as one of the most grievous sins, a danger that linguists must scrupulously avoid. No doubt one can avoid this sin and still do quite useful work. But I do not see the force of the injunction. I believe, and will try to show, that by disregarding it, by developing the study of language frankly and openly as a branch of theoretical psychology, we can considerably advance our understanding of language and arrive at interesting conclusions with regard to human intelligence and the innate structures of mind that permit us to act as free and creative beings, undetermined, even probabilistically, by stimulus conditions, yet at the same time functioning within a certain system of rules and principles that are in part a product of intelligence, and in part a fixed and immutable framework within which such products of intelligence are constructed. These notions are rather vague. I will try to give them a sharper form in the course of the discussion.

I have tried, a number of times, to relate these considerations to some long-standing issues in the history of Western thought — not too persuasively, to judge by recent commentary. However, I think a strong case can be made that this interpretation, both of history and of current work, is quite legitimate.

There has been a fair amount of controversy and debate over these matters in the past few years, much of it in my opinion beside the point because of faulty and misleading construction of the issues. I don't propose to spend much time reviewing these debates. Let me illustrate with just one example, perhaps extreme in the degree of misunderstanding but not, I am afraid, totally untypical. A recent book contains several essays on the issues that I want to raise here.¹ In one contribution entitled "Innate Knowledge,"² Rulon Wells offers a refutation of what he takes to be my basic assumption, namely, that "the difference between automaton and brute, on the one hand, and man on the other, is just the difference between finite and infinite power." He argues at length that "an automaton (and presumably a brute) can be regarded as incorporating an infinite power, and . . . to finish the picture, a hu-

man being, equally with an automaton and a brute, will have a 'part two' that imposes a finitude limitation. The capacity of man may be ever so much greater than the capacity of automaton or of brute; but the difference will be the difference between one finite magnitude and another . . ."

It is surely true that automata, animals, and man can be regarded as incorporating an infinite power as well as a "part two" that imposes a finitude limitation." Wells's mistake is double: first, to suppose that this has ever been an issue, and second, to maintain that "the capacity of man may be ever so much greater than the capacity of automaton or of brute," where capacity is measured in numerical terms, in terms of range of possible responses. In fact suppose that we abstract away from "part two," the finitude limitation imposed by finiteness of life or shortness of breath. This abstraction is surely legitimate and enlightening; it enables us to study the inherent capacity, the "generative capacity," of the mechanisms underlying behavior in the case of automaton, animal, or man. If we do not carry out this abstraction, if we consider generative capacity together with the finitude limitation, then Wells is surely incorrect in his claim that human capacity is greater in scope than that of animal or automaton. But if we make the abstraction, as we must to move to a significant further investigation of capacity, then he is wrong for more interesting reasons. It was obvious even to the Cartesians that "because there may be an innumerable variety in the impressions made by the objects upon the sense, there may also be an innumerable variety in the determination of the Spirits to flow into the Muscles, and by consequence, an infinite variety in the Motions of Animals." (François Bayle, 1669.)

In short an animal can operate on the principle of the speedometer, producing a potentially infinite, in fact in principle continuous, set of signals as output in response to a continuous range of stimuli — the signaling system can be infinite, in fact continuous, in the only sense in which any physical signaling system can be regarded as continuous. The range of behavior, the scope of the signaling system, is thus in principle greater than that of human language — which is a discrete communication system — not by a difference of degree, but by a difference of kind.

Examples can easily be found. Von Frisch's beautiful work on the dance language of the bees is one — but considerable doubt has been cast on it in some recent studies (Adrian Wenner, Patrick Wells, Den-

¹ *Language and Philosophy*, ed. Sidney Hook (New York: New York University Press, 1969).

² *Ibid.*, pp. 99–119; the quotation is on p. 118.

nis Johnson, *Science*, 4 April 1969). Consider therefore a case described by W. H. Thorpe, the case of a bird song in which the rate of fluctuation between high and low pitch signals the intensity with which territory will be defended. This is in principle a continuous system — again, in the only sense in which a physical signaling system can be regarded as continuous. Its scope is therefore greater in kind, not degree, than that of discrete human language. (In passing, I should note that there is a gestural element in human language that is also continuous in scope, but this is not to the point here.) This signaling system differs from human language not only in that it is far greater — not numerically more limited — in scope, but also, more importantly, in that it is directly associated to external stimulus conditions; it is, in short, a signaling system, like a speedometer, and not a language in the human sense, a system that is available for free expression of thought precisely because it is not tied directly to external stimuli. The point was emphasized, again, by the Cartesian linguist-psychologists. Human speech is infinitely varied in scope and appropriate to situations, but is not controlled by external stimuli in the sense of animal signaling systems of the sort just described. This fact poses interesting problems for inquiry; the Cartesians took it as one demonstration that man escapes the inherent limits of mechanical explanation, as they understood the latter notion. Whatever the explanation and analysis may be, the point is that human language differs qualitatively in this respect from animal signaling systems, not in that its scope is numerically greater in finite magnitude, as Wells erroneously believes, nor in that it is infinite as compared to the finiteness of animal behavior, in accordance with the quite ridiculous view that Wells refutes and that no one has ever proposed, but rather in that it is in principle an infinite discrete system rather than a continuous system, and that it is related to external stimuli not by the mechanism of stimulus control, but by the much more obscure relation of appropriateness.

When we consider automata, the matter is still more clear. There is, in fact, an interesting literature on the generative capacity of restricted infinite automata, and there has been much effort — so far inconclusive — to place human languages in a hierarchy of restricted infinite automata, in terms of generative capacity. In the light of this work, Wells's misunderstanding is remarkable. In any event it completely misconceives the issue.

There is, in fact, an interesting issue. Every animal communication system that is known operates on one of two principles: either the principle of the speedometer, as described a moment ago, or else a principle of strict finiteness; that is, the system consists of a finite number of signals, each produced under a fixed range of stimulus conditions. Human gestural systems are not well understood, but it is reasonable to suppose that they too observe these limitations. Human language, however, is entirely different. A person who knows a language has mastered a set of rules and principles that determine an infinite, discrete set of sentences, each of which has a fixed form and a fixed meaning or meaning-potential. Even at the lowest levels of intelligence, the characteristic use of this knowledge is free and creative in the sense that utterances are not controlled by external stimuli but are appropriate to situations, and in that one can instantaneously interpret an indefinitely large range of utterances, with no feeling of unfamiliarity or strangeness — and of course no possibility of introspecting into the processes by which the interpretation of these utterances, or the free and creative use of language, takes place. If this is correct, then it is quite pointless to speculate about the evolution of human language from animal communication systems, as pointless as it would be to speculate about the evolution of language from gesture. It is an interesting question whether properties of human language are shared by other cognitive systems. But no dogmatic assumptions are in order — that much seems clear.

The set of rules and principles that determine the normal use of language I will refer to as the 'generative grammar', or simply the 'grammar' of the language. There is an ambiguity in the usage of this term that should be noted; that is, the term 'grammar' is also used for the explicit theory, constructed by the linguist, which purports to be a theory of the rules and principles, the grammar in the first sense, that has been mastered by the person who knows the language. No confusion should arise if the distinction is kept in mind. The linguist's grammar is a theory, true or false, partial or complete, of the grammar of the speaker-hearer, the person who knows the language. The latter is the object of the linguist's study. He cannot, of course, observe it directly and can only attempt to construct a theory of the speaker-hearer's grammar, making use of whatever evidence he can obtain by observation or introspection, all such evidence, of course, being fallible and subject to correction and revision.

The linguist's grammar is, thus construed, a psychological theory. It is an attempt to account for evidence of behavior and introspection by ascribing to the language-user a certain system of rules and principles that he applies in language use, as a speaker and hearer. It is postulated, then, that a person who knows a human language has internalized, has developed a mental representation of a grammar, a set of rules and principles, of which the linguist attempts to construct a precise and explicit model. How this abstract representation is realized is another question, concerning which we have no serious information at the moment. Conceivably such evidence might be forthcoming, but nothing is known today of much significance.

To allay misunderstanding, let me make clear that I am not proposing that this conclusion is one of logical necessity. Obviously not. It is an empirical hypothesis, to be judged in terms of its success in explaining and accounting for certain phenomena, observations that can provide evidence for or against certain explicit assumptions about this grammar which, it is postulated, has been internalized by the language-user. What is postulated is that to know a language is to have a certain mental constitution which is characterized by the linguist's grammar. There is nothing mystical about this approach, contrary to what is sometimes believed. It is precisely the approach that would be taken by a scientist or engineer who is presented with a black box that behaves in a certain fashion, that evidences a certain input-output relation, let us say. The scientist will try to construct a theory of the internal structure of this device, using what observations he can as evidence to confirm his theory. If he is unable to investigate the physical structure of the device, he will not hesitate to ascribe to the device a certain abstract structure, perhaps a certain system of rules and principles, if this turns out to be the most successful theoretical approach. There is no reason to adopt some different standpoint when the object under investigation is the human being.

In the case of the black box, the scientist may be mistaken in attributing to it an abstract grammar as its internal structure. He may be wrong in postulating that a certain specific set of rules and principles has been internalized by the device and constitutes its 'mental state', if one wishes to use this terminology. He may be wrong in that he has selected the wrong rules and principles, the wrong grammar. Or he may be wrong, more deeply, in that he has taken entirely the wrong approach to under-

standing the device. To make the analogy closer, let us use the terminology that Wells suggests and distinguish, in our theory of the device, two components: part one, a grammar that has a certain generative capacity, and part two, a set of conditions that impose a finitude restriction. The latter can be elaborated in the case of study of mind; we can distinguish certain conditions that have to do with organization of memory and perceptual strategies, perhaps certain belief systems that are extraneous to language, and so on. Abstracting away from these elements of the total theory of mind, we can focus attention on the grammar, the set of rules and principles that determine the form and meaning of an infinite number of sentences. The abstraction is legitimate. It may be misguided, as may any theoretical construction, but I do not think that it is, in this case. My belief that it is the proper abstraction, that the theory of mind is best conceived in these terms, is grounded in two sorts of considerations: First, its success in explaining a variety of phenomena. Second, the difficulty of constructing a coherent alternative. I will discuss the first kind of grounds in subsequent lectures. Let me briefly turn to the second.

One of the very few attempts to construct an alternative to the approach I've just briefly outlined is by Gilbert Harman.³ He argues that the goal of linguistic description should be not a grammar, abstracted as described above, but rather a performance model, which he outlines (p. 80) as a device with a perceptual and phonetic input, and with an output consisting of bodily movements and sentences. The internal structure of his performance model includes a representation of beliefs, of plans, goals and intentions, and a component labeled "phonetic, syntactic, semantic rules of the grammar." It contains a central "thinker and decider" which makes use of perceptual input, plans, goals and intentions, beliefs, and rules of grammar, in producing its output. Harman proposes this as an alternative to what I described, say, in *Aspects of the Theory of Syntax*. On the contrary, it is virtually identical to what is described in *Aspects*, and — the crucial point here — incorporates an abstract component containing the phonetic, syntactic, and semantic rules of grammar exactly as in the model I described. In attempting to construct an alternative, Harman ends by simply presenting the model

³ "Psychological Aspects of the Theory of Syntax," *Journal of Philosophy*, 64, no. 2 (1967):75-87.

to which he objects, the model which, he believes (mistakenly, in my opinion), embodies all sorts of confusions and problems.

Harman argues that from the fact that the speaker relates sound and meaning, it does not follow logically that he uses a representation of the rules of grammar to carry out this act. To assume the contrary would lead to an infinite regress, by a familiar argument used, for example, by Ryle in a similar connection. The argument is in part correct. The evidence of language use does not logically imply that the language-user has internalized the rules of a grammar. Rather, the latter conclusion is an instance of what Harman has elsewhere termed induction to the best available explanation. Furthermore, as noted, his alternative is not different from the concepts that he criticizes, in that his model of mind also incorporates a generative grammar as one component. This component of his model of mind contains a representation of the phonetic, syntactic, and semantic rules. It is only a shift of terminology to go on to say that the mind, so modeled, is postulated to contain an internal representation of a grammar. If we want to go on to study acquisition of language, we will have to consider the question how this grammar is constructed on the basis of experience, and what kind of structure must be attributed to the mind to account for this achievement. When misunderstandings and misinterpretations are cleared away, I see no issue raised in Harman's critique. Harman's several papers are important, in this connection, because they represent the only serious attempt, to my knowledge, to construct an alternative to the conception outlined earlier, taking into account the actual empirical problems.

On the surface the behaviorist account of language use proposed by many philosophers, psychologists, and linguists appears to be a genuine alternative approach. However, the behaviorist alternative, as actually formulated, contains so many escape hatches that it ultimately has no empirical content whatsoever, so far as I can see. The matter is worth a few moments' discussion, since the development of the behavioral approach in psychology and the social sciences has been heralded as a major breakthrough. My personal opinion is rather different. I think that it is an intellectual and social calamity. However this may be in general, the behaviorist position to language collapses when the issue is pressed.

I think that Quine's recent writings are quite informative in this regard. Quine has been the leading and certainly the clearest exponent of

a behaviorist position with regard to human language, its use and acquisition. He has frequently indicated that he sees himself as developing a view rather like Skinner's. The latter proposition, incidentally, seems to me without content. Skinner, so far as I can see, has no position at all with regard to human language. He has made a variety of terminological proposals; in particular he insists that the words 'stimulus', 'response', 'reinforcement', and several neologisms be used in describing language use, but he proceeds to deprive these terms of any content. For example, his notion of 'reinforcing stimulus' includes as a special case stimuli that do not impinge on the organism at all, but are merely hoped for or imagined. Quine too uses the term 'reinforcement' in a purely ritual fashion. Thus he suggests that in the case of language learning the child's reinforcement may be the "corroborative usage" of the speech community. Anyone would agree that corroborative usage, that is, data, is required in language learning. Quine in fact insists that his behaviorism is virtually empty. In his discussion of this matter in *Language and Philosophy*, he rejects the restriction of "behaviorism" to the theory of conditioning and says this: "When I dismiss a definition of behaviorism that limits it to conditioned response, am I simply extending the term to cover everyone? Well, I do think of it as covering all reasonable men. What matters, as I see it, is just the insistence upon couching all criteria in observation terms." All conjectures, he says, must "eventually be made sense of in terms of external observation."⁴ This is, to be sure, a sense of "behaviorism" that would cover all reasonable men.

Quine states explicitly that "conditioning is insufficient to explain language-learning." In fact, he states that this was in essence the content of his doctrine of indeterminacy of translation. That is, he interprets this doctrine as asserting that language learning cannot be explained in terms of conditioning. I find it difficult to read this interpretation into his exposition in *Word and Object*, for example, but instead of pursuing the matter of "indeterminacy of translation," let us consider rather the notion 'conditioning'. In *Word and Object* Quine states that a theory, in particular a language, can be characterized as "a fabric of sentences variously associated to one another and to non-verbal stimuli by the mechanism of conditioned response" (p. 11). On the face

⁴ "Linguistics and Philosophy," p. 97; emphasis added.

of it, this definition seems inconsistent with his assertion, in *Language and Philosophy*, that "conditioning is insufficient to explain language-learning." If the latter is true, then a language will not be a fabric of sentences and stimuli associated by the mechanism of conditioned response.

I think that the solution lies in the fact that Quine has not only virtually abandoned behaviorism and the concept 'reinforcement', but also the notion 'conditioned response'. We can see this by considering his account of language-learning, what he calls "learning of sentences." In *Word and Object* he specifies three mechanisms by which sentences can be learned, three mechanisms for language-learning. The first is association of sentences with sentences; the second, association of sentences with stimuli. These two methods would, it is true, lead to a fabric of associated sentences and stimuli. But there is a third method that is left rather obscure in *Word and Object*, namely, learning of sentences by what he calls "analogic synthesis." He gives only one example, which I quote in full:

"It is evident how new sentences may be built from old materials and volunteered on appropriate occasions simply by virtue of analogies. Having been directly conditioned to the appropriate use of 'Foot' (or 'This is my foot') as a sentence, and 'Hand' likewise, and 'My hand hurts' as a whole, the child might conceivably utter 'My hand hurts' on an appropriate occasion, though unaided by previous experience with that actual sentence."⁵

In Quine's terminology, the sentence "My hand hurts" might be learned by this method, the method of analogic synthesis. Clearly this is a curious use of the word 'learning'. Putting that matter aside, however, consider the consequences for his theory of language. Suppose that the sentence "My hand hurts" is 'learned' in this manner, and consider now the assumption that a language is a fabric of sentences associated by the mechanism of conditioned response. Then the sentence "My hand hurts," in the given example, is associated to the complex containing "This is my foot," "foot," "My foot hurts," and "hand" by the mechanism of conditioned response. To say this is to deprive the notion "conditioned response" of its strict meaning, or anything resembling this meaning. The responses and stimuli entering into the relationship of "conditioning" need not even appear together. Having deprived the no-

tion of any content, we are free, without fear of contradiction, to describe a language as a fabric of sentences and stimuli associated by the mechanism of conditioned response.

In a response in *Synthese*⁶ Quine emphasizes that one cannot regard sentences as associated with one another and "learned" as "unstructured wholes." Rather, they are associated and learned by various modes, such as "analogic synthesis," not merely as unstructured wholes. It must be stressed that this remark is, in its entirety, his positive theory of language learning. Again, we have a sense of "behaviorism" with which no one could disagree. What has happened, clearly, is that the terms "conditioned response" and "association" have joined "reinforcement" and "behaviorism" as terms with a merely ritual function, virtually deprived of substantive content, so far as I can see.

In *Word and Object* Quine introduced the notion of an innate quality space with a distance measure, to explain induction. In commenting on this, I noted (in the issue of *Synthese* just mentioned) that the examples Quine gives associate the qualities of the innate quality space with dimensions that have some simple physical correlates such as hue or brightness, with distance defined in terms of these physical correlates. I suggested that one could develop a substantive doctrine by making this association explicit. Quine, however, rejects this interpretation and rightly so, for it would make the doctrine false. He holds the "denizens of the quality spaces" to be "stimulations, any and all, with no prior imposition of dimensions." No further conditions are given. The concept is therefore quite vacuous. There is, for example, no objection to a quality space with a dimensional structure so abstract that any two sentences of English are "closer" than a sentence of English and a sentence of any other language, so that a person innately endowed with this quality space could learn all of English, by induction, from a presentation of a single sentence. He could, that is, generalize properly from such a presentation to full knowledge of all sentences of English with the situations in which they are appropriate. I take it that with this conclusion the notion "quality space" joins "reinforcement," "conditioned response," and "behaviorism."

This conclusion may be premature, however, for in *Language and Philosophy* (p. 97) Quine asserts that the quality space, thought not restricted to dimensions with simple physical correlates, nevertheless

⁵ W. V. Quine, *Word and Object* (Cambridge, Mass.: M.I.T. Press, 1960), p. 9.

⁶ 19(1968):274-283.

still permits only what he calls "induction," and not the move to the "analytical hypotheses" which, he holds, must be developed somehow by the language-learner. I quote: "The as yet unknown innate structures, additional to mere quality space, that are needed in language-learning, are needed specifically to get the child over this great hump that lies beyond ostension, or induction." That is, they are needed to get the child to "analytical hypotheses." Quine also insists (in the *Synthese* article) that "generative grammar is what mainly distinguishes language from subhuman communication systems" — correctly, I am sure. Perhaps it is proper, then, to regard the principles and rules of the internalized generative grammar as among the "analytical hypotheses" that are arrived at by the "as yet unknown innate structures, additional to mere quality space, that are needed in language-learning . . . to get the child over this great hump that lies beyond ostension, or induction." Personally, I would find this interpretation of Quine's position quite congenial. The task of the linguist, in this interpretation, is to arrive at the internalized generative grammar using the evidence of language use. He will call the rules and principles of this grammar "analytical hypotheses," insofar as they are not determined through induction by the mechanism provided by the quality space of unknown dimensions. He will then seek to determine the nature of the quality space and the unknown innate structures, additional to the quality space, that are needed to account for the construction of this generative grammar by the language-learner. I believe that this is a fair rendition of Quine's most recent formulations. I think it is fair to describe this as an abandonment of behaviorism. I would only suggest that we now also abandon the terms "association," "conditioning," "reinforcement," and "behaviorism," now that they have been deprived of whatever content they have in the psychological literature, and now that all of the characteristic assumptions of behaviorism have been abandoned.

One last remark on Quine's behaviorist theory of language. In *Word and Object* Quine defines a language as a "complex of present dispositions to verbal behavior, in which speakers of the same language have perforce come to resemble one another." Surely he must also abandon this view. Presumably, a complex of dispositions is representable as a set of probabilities for utterances (responses) in certain definable circumstances or situations. Suppose that knowledge of language is represented by such a description. Then assuming 'circumstances' and 'situations' to

be defined in terms of objective criteria, as Quine insists, it is surely the case that almost all entries in the situation-response matrix are null. That is, in any objectively definable situation, the probability of my producing any given sentence of English is zero, if probabilities are assessed on empirical grounds; in any event, it is not detectably different from the probability of my producing some sentence of, say, Japanese, if probabilities are assessed on empirical grounds. Thus knowledge of English is not differentiable from knowledge of Japanese. Clearly the whole approach is untenable and should be simply abandoned. If it is a generative grammar that mainly distinguishes language from subhuman communication systems, as Quine holds, then a language cannot be defined as a complex of dispositions to respond, since a generative grammar cannot be characterized in these terms.

To summarize: as a first approximation it is fair to assume, not as a matter of logical necessity, but as a plausible hypothesis, that the 'state of mind' of a person who knows a language is characterized by a generative grammar, a system of rules and principles that determines a sound-meaning connection for an infinite set of sentences. A deeper problem will be to investigate the 'state of mind' that is innate to the organism, that makes it possible for the grammar to be constructed in the specific way it is on the basis of experience. Thus far there appears to be near agreement between Quine, Harman, and myself. Agreement does not prove correctness, of course. To show that this approach is a valid one, one must demonstrate empirical successes in accounting for some interesting phenomena. But agreement does indicate, I think, the difficulty of conceiving a coherent alternative.

Let me consider a number of other apparent alternatives to this approach. In a brief essay in *Language and Philosophy*,⁷ Nelson Goodman suggests that even if certain principles are discovered that characterize what might be called "the workings of the mind" in such a way as to meet the empirical conditions of description and explanation, there is no necessity to regard these principles as "in the mind" except in the sense that they can be inferred from what the mind does. That is, these principles may be nothing "more than descriptions, by an observer, of the resulting organization." They "need no more be in the minds in question than the theory of gravitation need be in bodies." (I should

⁷ "The Emperor's New Ideas," pp. 138-142.

add, for clarification, that Goodman dislikes this whole terminology and accepts it only for purposes of discussion.) Presumably this argument holds both of the 'final state of mind' and the 'initial state of mind' of the language-learner, both of the mind of one who knows the language, and the mind of one equipped to learn it. In either case, following Goodman, we may say that the linguist's theory of the two states says nothing more about the organization of mind than the theory of gravitation says about the internal structure of bodies. That is, there is no reason to say anything other than that minds, in their initial and final states, are governed by these rules and principles. Just as the laws of gravitation are not represented internally in bodies governed by these laws, so the rules and principles of grammars are not represented in the minds governed by these rules and principles.

The argument is a bit unfair to Newton, who, after all, was not quite unconcerned with the question of how the behavior of a body under the laws of gravitation was affected by its form and internal structure. But putting that aside, suppose we were to apply Goodman's argument to a scientist who is presented with some device, say an automobile engine, which he cannot take apart and can investigate only by studying its behavior and 'input-output' relations. Suppose that the scientist cleverly hit upon a theory that worked quite well, namely, the theory that makes use of components with the properties of cylinders, spark plugs, and so on. He might hesitate to speculate about the precise physical realization of these concepts, and leave it as an abstract theory of the device, with various components, a certain kind of interaction between them, and so on. Observing Goodman's scruples, he should say only that the device behaves according to principles enunciated in his theory, that these principles describe its behavior only in the sense that the laws of gravitation describe its behavior. If the scientist were to propose that the device is actually constructed in accordance with the principles outlined, that the hypothesized components are embodied in the device in some fashion or other, he would be going beyond what Goodman regards as proper. Goodman's strictures are a bit vague, but if I understand him, he is saying that the scientist may say *that* the device obeys certain principles, but he may not ask *why* it does so; he may not speculate on the internal structure and organization that lead to its behaving in accordance with these principles. If this is his intention, the reference to the theory of gravitation is incorrect, since, as noted, Newton

was quite concerned to explain why a complex object would behave like a point mass. But in any event, the scruples, so interpreted, simply amount to a lack of curiosity, an unwillingness to pursue questions beyond a certain arbitrary point.

Consider the remark that the principles may be nothing "more than descriptions, by an observer, of the resulting organization." To say that certain principles are descriptions of the resulting organization does not seem to me any different from saying that the system is organized in the way described by the principles in question. In this case Goodman's strictures appear to be merely terminological.

In any event I fail to see that he has offered any objection to the conclusion that a generative grammar is somehow internally represented in the mind of the person who knows the language, and that innate structures to which we are led in the study of language acquisition are internally represented in the mind of the language-learner; or, if you like, that the final state of mind of the language-learner is correctly described by a generative grammar, and the initial state by the postulated innate structures and principles. There is no difference, so far as I can see, between these terminologies.

Goodman does, however, have a different view with respect to language-learning. He argues that acquisition of language is a case of second-language learning, and that it is facilitated by the possession of a language of gestures and perceptual signs that can be used perhaps as analogues, or perhaps in explanation and instruction. This is pure hand-waving. In answer to the question how the specific structures of natural language are derived from systems of gestures and perceptual signs, he says nothing; or, to be more precise, he offers another analogy, namely, that just as tools can be used to make a clock, so prelinguistic symbolic systems can be used to explain — let us say — the principle of cyclic application of grammatical transformations. But this clearly tells us nothing about how the latter can be derived as an analogue to the former.

Instead of trying to deal with these problems, Goodman offers still another analogy. He points out that "the features that identify a picture as by a certain artist or of a certain school or period are in some sense deep (or obscure). Yet we learn with rather few examples to make some of the latter rather subtle distinctions. Must the mind therefore have been endowed at birth with a 'schematism' of artistic styles?" Apparently he regards it as self-evident that the answer to this rhetorical question is

negative. The analogy is entirely without force. To explain how a person learns to make subtle distinctions with few examples, we must attribute to the mind an innate structure rich enough to achieve this result, and not so rich as to be falsified by the actual range of such 'input-output' relations. If a postulated 'schematism of artistic styles' will meet these empirical conditions, there is no reason to be at all surprised, or to regard this as an objectionable hypothesis. Goodman seems amazed at the idea that one should seek to study human mentality in exactly the way one would approach any organism, or an inanimate device of unknown properties that modifies its state through time. The recent literature on this question contains many other rhetorical questions like the one Goodman asks, as if it were somehow absurd, *prima facie*, to construct a theory of the structure of mind that accounts for empirical facts, attributing to it as much structure as is necessary to do so. If we are forced to attribute to the mind separate 'faculties' in order to account for the empirical facts, then so be it. There is no reason for any *a priori* attitude as to the rightness or wrongness of such a result. If we find generalizations governing several such 'faculties', or more general structures that can account for many types of learning, well and good. Again, there is no reason for dogmatism in this matter.

When we try to characterize the state of mind of a person who knows a language, taking account of his ability to use and understand an indefinite range of sentences, each with its phonetic form and meaning-potential determined in a specific way, we are led to certain empirical hypotheses, specifically, to the construction of a generative grammar, a system of rules and principles that establishes a certain sound-meaning relationship. We may then say that this theory describes the organization of mind or that the mind is organized in accordance with this description. I see no difference. In either case we can then go on to ask how this organization developed through an interaction of experience and innate structure, and can seek to determine the specific innate endowment that makes this achievement possible.

So far I have said nothing of 'knowledge of language'. I think that it would be quite reasonable to suggest a characterization of 'knowledge of language' in the terms just given; to say, that is, that to know a language is to have internalized a generative grammar (equivalently, to have developed a state of mental organization as described by a generative grammar). It might be argued that this proposal does violence

to the concept 'knowledge'. The latter concept seems to me sufficiently obscure so that I do not know whether or not this criticism is just. It seems to me that our concept of knowledge fades into obscurity at the point where we consider what Leibniz referred to as the principles that "enter into our thoughts, of which they form the soul and the connection," principles as necessary to thought "as the muscles and sinews are for walking, although we do not at all think of them." In the past I have tried to avoid, or perhaps evade the problem of explicating the notion 'knowledge of language' by using an invented technical term, namely, the term 'competence' in place of 'knowledge'. However, the term 'competence' suggests 'ability', 'skill', and so on, through a chain of association that leads direct to much new confusion. I do not think that the concepts of ordinary language suffice for the purpose at hand; they must either be sharpened, perhaps somewhat arbitrarily, or replaced by a new technical terminology. Either approach has a familiar disutility.

Suppose that we were to propose that to know a language is to have constructed, to be sure unconsciously, a specific generative grammar. A familiar argument against this proposal is that I can tell whether someone knows English, but I know nothing of the internal workings of his mind. The argument seems to me weak, because it begs a question rather like the one under discussion. If we are prepared to admit that the mind can incorporate unconscious theories, systems of principles and rules that we might describe as unconscious knowledge, then it is at least conceivable that I have an unconscious theory that attributes to other persons minds of a certain character, and that assigns to them mental states by virtue of certain actions that they perform. Suppose, furthermore, that this theory relates to my unconscious theory of English in such a way that I believe someone to know English when I attribute to him the mental state described by (or incorporating) the rules and principles of English grammar, arriving at this conclusion on the basis of his behavior. I see no incoherence in this formulation, which would support the conclusion that my concept 'knowledge of a language' is directly related to the concept 'internalization of the rules of grammar'.

Perhaps one might approach the analysis of "knowledge of a language" in a simpler and more direct way, supposing that to know a language is to know how to speak and understand, the latter being a dispositional concept of some sort. This leads us nowhere, so far as I

can see. The problem immediately arises of characterizing the relevant dispositions, determining how they are related, how this complex is organized and developed, why certain dispositions are excluded from it and others not. Furthermore, we must face the fact that two people may have very different dispositions, may be inclined to say very different things in given circumstances and to interpret what is said quite differently, and yet we may attribute to them exactly the same knowledge of language. We are back where we were, so far as I can see.

There are difficulties, perhaps not insuperable but worth mention nevertheless, in attempting to account for knowledge of language in terms of 'knowing how'. To know English is, let us say, to know how to talk grammatically, how to understand what is said to us in English, and so on. To say that someone knows how to do such things is to say that when the relevant acts are performed, they are performed intelligently. To act intelligently, as Ryle put it long ago, is not merely to satisfy certain criteria but also to apply them. But what does it mean "to apply criteria"; how do we know which criteria to apply, how are these criteria related and organized. Again, we run into a barrier that can be overcome, I think, only by introducing some concept like generative grammar, and mental representation of generative grammar.

In fact, the discussion of 'knowing how' often takes a different, and I think misleading course, at exactly this point. It is said that to know how, to be able to act intelligently, is to have a skill, a disposition or complex of dispositions to act in a certain way. Consider the standard example, knowing how to play chess. This is sometimes described as a skill acquired by training, a complex of dispositions to make certain appropriate moves, a skill that improves with practice. A person is said to know how to play chess if he normally makes the permitted moves, avoids the wrong moves, and so on (Ryle). But there is a possible source of confusion here. Consider the difference between the two questions: "Does John know how to play chess?"; "How well does John (know how to) play chess?" It is the latter, not the former, that asks about a skill and how well it is exercised. The former asks about possession of knowledge, which can be partial or improved only in the way in which knowledge of facts can be partial or improved.

Knowing how to ride a bicycle may be a matter of having certain habits and reflexes, and knowing how to drive a car may be a matter of having certain skills. Similarly, being able to play chess well may be a

matter of having certain skills, acting thoughtfully, and so on. But knowing how to play chess seems to me more like knowing how to get from New York to Chicago than like knowing how to ride a bicycle or drive a car. If a person does not know some specific rule of chess, say the rule of castling, he may still play chess very well; he may (in principle) be world champion, although it would still be correct to say that he does not know how to play chess in the way the rankest amateur who loses every game may very well know how to play chess. We may say that this chess champion has partial knowledge, but his defect of knowledge is very different from the defects in knowing how to play chess on the part of the poor player who knows all the rules (perhaps unconsciously, without ability to formulate them). Discussions of skills, dispositions, thoughtful action, and so on, seems to me appropriate in relation to ability to play well, but not in relation to knowing how to play; the latter concept has an irreducible intellectual component.

It is true that we say that a person knows how to play if he normally makes the right moves, etc. (as Ryle puts it); that is, his actions provide the evidence that leads us to attribute to him the knowledge how to play chess — perhaps erroneously. But to attribute to him this knowledge is not the same thing at all as attributing to him the ability to play well, perhaps on the same evidence. The concepts are entirely different, though the evidence for applying them may overlap.

Knowledge of a language, knowledge of classical mathematics, knowledge of chess, seem to me to be instances of 'knowing how', if at all, only in the sense in which knowing how to get from New York to Chicago is an instance of 'knowing how'. The latter implies no particular ability to get from New York to Chicago. That requires other talents apart from knowing how to get from New York to Chicago. And, as in the case of playing chess, the person most successful in getting from New York to Chicago, repeatedly and characteristically, may not know how to get from New York to Chicago as completely, as fully, as perfectly, as someone who repeatedly fails. Similarly, two people may know English in exactly the same way, they may have exactly the same knowledge of language, and yet differ in their dispositions, in how well they speak and understand the language that both know; as they might in principle match perfectly in these dispositions, and yet differ in their knowledge of language. Ability to speak and understand involves not

only knowledge of language, in the strict sense, but much else as well. The concepts should be kept quite separate.

There is a further problem that should be mentioned. Knowledge of language can be partial only insofar as there is an external standard. Let us assume, for simplicity, a perfectly homogeneous adult speech community. Customarily, one speaks of a child as having partial knowledge of a language, taking as the external standard the grammar of the adult language, of the speech community to which he belongs. In another sense the child has, by definition, a perfect knowledge of his own language. There is no contradiction here; rather, there are two notions that must be distinguished. The study of language should be concerned, in the first place, with the speaker's perfect knowledge of his own language. The notion of 'language' as a common property of a speech community is a construct, perfectly legitimate, but a higher order construct. In the real world there are no homogeneous speech communities, and no doubt every speaker in fact controls several grammars, in the strict sense in which a grammar is a formal system meeting certain fixed conditions. A person who is capable of learning a language in the real world of complex and overlapping dialects has a certain ability which would enable him to learn the language of a homogeneous speech community. It is this ability that we must seek to capture and understand, abstracting away from much real world complexity. Given some understanding of this ability, we may try to come to grips with the concrete problem of 'knowing how to speak and understand', of acquiring and using knowledge of language, in the real world of heterogeneous and overlapping styles and dialects.

One last problem. Suppose that one is willing to accept the characterization of knowledge of a language in terms of possession, internal representation of a generative grammar. Clearly the rules and principles of this grammar are not accessible to consciousness in general, though some undoubtedly are. I think that what we discover, empirically, is that those principles and rules that are accessible to consciousness are interspersed in some obscure and apparently chaotic way among others that are not, the whole complex of rules and principles constituting a system of a very tight and intricate design, and meeting stringent and restrictive general conditions of a sort that I will try to illustrate in subsequent lectures. Suppose that one is prepared to apply the notion 'knowledge' in this

case — that is, to accept the locution "so-and-so knows the grammar of his language, its rules and principles," including those that lie beyond awareness. Suppose it is true, as I believe, that further investigation leads us to the conclusion that this knowledge is acquired on the basis of certain innate principles of what might be called 'universal grammar'. I mean by 'universal grammar' a certain fixed language-independent schematism that determines what counts as linguistic experience and what knowledge is acquired, what grammars are constructed, on the basis of this experience. Would we want to say, as well, that the child 'knows' the principles of universal grammar?

It seems to me that very little turns on the answer given to this question. It is also unclear to me whether the concept 'knowledge' is sufficiently clear to guide us in making a decision. It would be quite useful to have a concept *X* such that a person who knows English, and thus knows certain facts of English, also *X*'s these facts; that furthermore, he *X*'s the principles and rules of his internalized grammar, both those that can be brought to awareness and those that are forever hidden from consciousness; and that he *X*'s the principles that underlie the acquisition of language in the first place. We have no clear concept such as *X*. My guess is that Leibniz would have been happy to extend the concept 'know' to have the meaning of *X*. It seems a natural enough step to me, though I do not want to press the point. If we were to adopt this way of speaking, extending familiar terminology in what seems to me a natural direction, we would then speak of knowledge, unconscious knowledge, and innate knowledge. A person who knows English would no doubt have conscious knowledge of certain facts, say that bachelors are unmarried or that "is" goes with singular and not plural subjects. He would have unconscious knowledge of the fact that the passive transformation is ordered in a certain way with respect to the transformation that inverts direct and indirect object. He would have innate knowledge of the fact that transformations apply in a cyclic ordering. We might refer to his innate knowledge as 'knowledge of universal grammar'. Alternatively, we may use the term 'know' in a narrower way, restricting it to conscious 'knowledge that', and to knowing how, why, who, and so on. 'Knowledge of' in the sense of 'knowledge of language' will then be explicated in terms of a new technical terminology: 'internal representation of a grammar', '. . . of an innate schematism,' and so on. In

this usage what is 'known' will be a somewhat ill-defined and, perhaps, a rather scattered and chaotic subpart of the coherent and important structures and systems that are described in terms of the new, more forbidding terminology. As long as we are clear about what we are doing, either approach seems quite all right.

Language, Rules, and Complex Behavior

Prominent among recent additions to psychology's store of theoretical terms are the terms "rule," "tacit knowledge of rules," and "rule guidance." Language, Noam Chomsky tells us,¹ is a rule system and what every speaker of a language has mastered and internalized is the system of rules that comprise his language. What speakers do when they speak and understand the language is done in accordance with the rules. In speaking and understanding a language, we interpret, execute, and follow rules.

Looking beyond language and verbal behavior to other forms of intelligent behavior, J. A. Fodor has maintained that "the paradigmatic psychological theory is a list of instructions for producing behavior."² What the rule theorist advances is what Fodor calls an "intellectualist" theory of psychological explanation. For the rule theorist, performing some bit of complex behavior involves employing rules. The explication of these rules is tantamount to a specification of how to perform the behavior.

In saying that performance involves rules or that the behavior is guided by rules, the rule theorist means to say more than that the rules describe regularities in the behavior. There are many alternative ways of describing complex behavior and one way is to have the description take the form of a rule or instruction that indicates how to behave from moment to moment. However, the claim that a system employs rules or that some bit of its behavior is guided by a rule presumably implies more than that the rule describes the behavior. There are regularities in the behavior of falling bodies and in the behavior of gases, but, of course,

AUTHOR'S NOTE: Portions of this paper were presented at the Third Annual Conference on Structural Learning, Philadelphia, March 1972, and at the Pacific Division Meetings of the American Philosophical Association, San Francisco, March 1973.

¹ Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, Mass.: M.I.T. Press, 1965).

² J. A. Fodor, "The Appeal to Tacit Knowledge in Psychological Explanation," *Journal of Philosophy*, 65, no. 20:630.