Syracuse Scholar (1979-1991)

Volume 4
Issue 2 *Syracuse Scholar Fall 1983*

Article 2

1983

Mental Representations

Noam Chomsky

Mental Representations

Noam Chomsky

Noam Chomsky is Institute Professor in the Department of Linguistics and Philosophy at MIT. He is the author of many books and articles on linguistics, philosophy, intellectual history, and contemporary issues. In 1982 Professor Chomsky was the Jeanette K. Watson Distinguished Visiting Professor in the Humanities at Syracuse University. This article is an edited version of lectures given during that professorship.

My topic is cognitive psychology—that is, the study of the nature of human thought and action. Cognitive psychology incorporates parts of psychology, of linguistics, and of artificial intelligence; and it is a study which takes up questions which have been of central concern in philosophy and hopes to link them more closely with the neurosciences and general biology. In the light of the work of the past twenty years, it is fair to define cognitive psychology as the study of mental representations—their nature, their origins, their systematic structures, and their role in human action. I want to discuss some questions about mental representations: The questions are of various kinds and arise at various levels, and what I say about them must be briefer and more superficial than the topic deserves.

In some areas, there has been substantial progress in the understanding of mental representations: the study of the human visual system, the human language faculty, motor coordination, and a few others. In other domains we have made little or no progress, and the traditional questions still arise in pretty much their traditional form. This disparity between progress in some domains and lack of progress in others may indicate something about the nature of human intelligence, which, after all, is not a universal system designed to solve any conceivable problem but an evolved biological system which may well be adapted to conceiving, interpreting, and even answering certain kinds of questions while remaining forever baffled by others.

The first questions that arise are, What are mental representations? What right do we have to postulate them? What is their so-called ontological status, their status as things that exist in the world? Specifically, are we committed to some kind of metaphysical dualism when we discuss mental representations?

Assuming that we are satisfied of the legitimacy of postulating mental representations, logically we next want to ask, Under what cir-

cumstances does it make sense to postulate them? We will not postulate mental representations in explaining digestion, but we may postulate them in connection with the use of language. What kind of distinction is at work here?

Again assuming that that is settled, we next ask, What is the nature of mental representations in the domains where it is reasonable to postulate them? This includes what is sometimes called the format problem: What is the syntax of mental representations? What are their elements, and how are they put together? That is one core question of the science of mental representation. There is also what might be called the system problem: What systems of mental representations are there? Is there simply a single, homogeneous system of mind of which all mental representations and rules for constituting and operating on them form a part, or is the mind analyzable into separate subcomponents, perhaps analogous, somehow, to the organs of the body?

As well as problems about the ontological status of mental representations, and about their nature, we have a problem about how mental representations relate to other aspects of the world. There are two facets to this problem: How do mental representations enter into human thought and action? And how do they arise on the basis of experience?

The ontological status of mental representations

Let us begin with the questions, What is the ontological status of mental representations? And is it legitimate to postulate them at all? We all know in a general way how Descartes answered this (and he was largely responsible for putting this whole problem area on the modern intellectual agenda). He postulated two substances, body and mind, held them to be distinct, and raised various questions about the nature of their interaction.

The significance of Descartes's distinction is often not appreciated, I think. The position he developed was more rational and more compelling than it is commonly assumed to be, though it is not ultimately tenable. The basic questions he was raising are ones which we still face and still cannot answer.

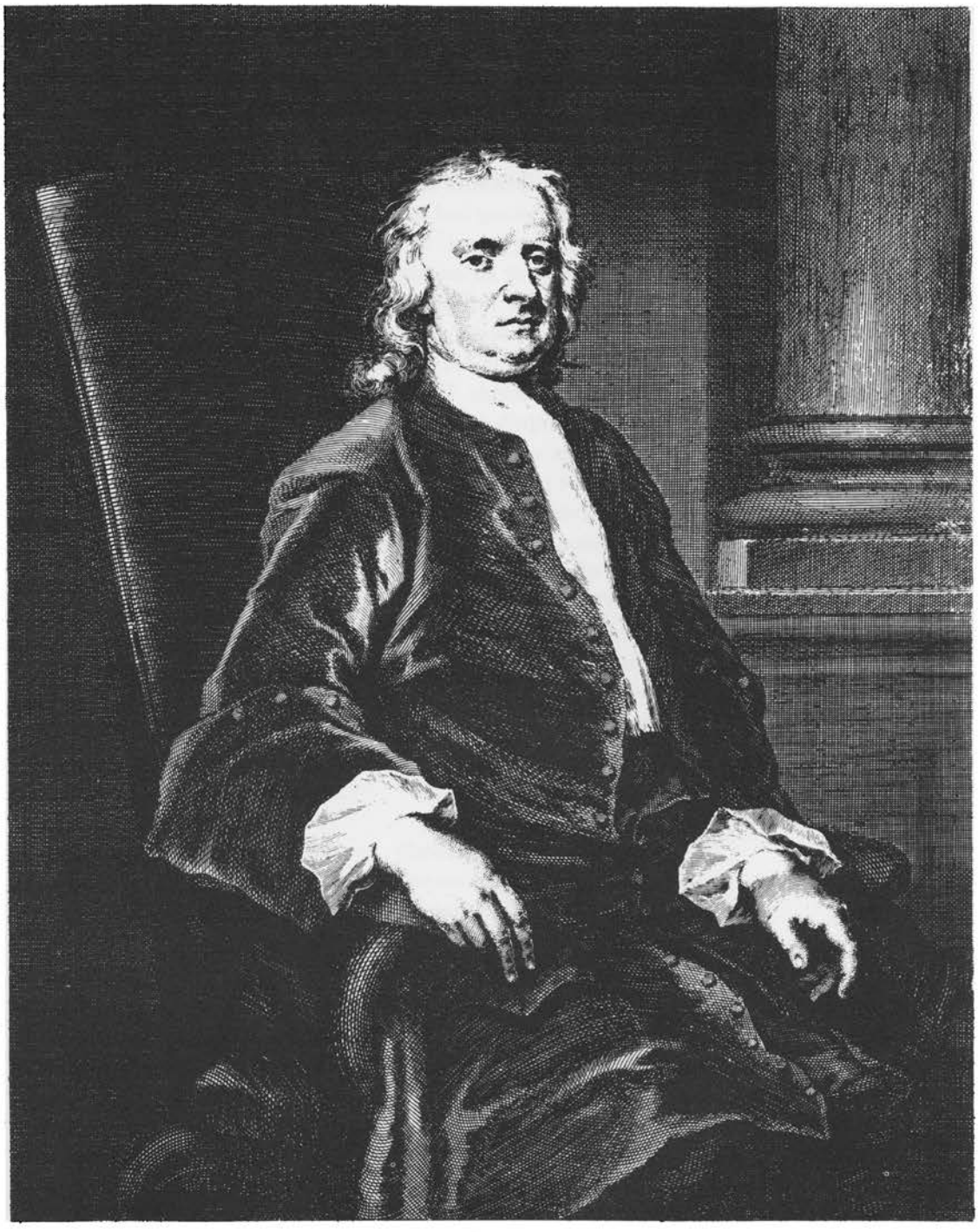
Let us consider one crucial way in which he developed his doctrines—one arising out of his scientific rather than his philosophical concerns, in our terms. As anyone should who is interested in the issue of dualism or monism, Descartes began by developing a concept of body. His specific concept of body was rooted in a theory of mechanics—a theory that reflects closely our intuitive ideas about how the world behaves. It was a mechanics of elements in contact, of how things push or pull or twist each other, and so on. Descartes believed that that kind of *contact mechanics* exhausted everything that happens in the physical world: He tried to use it to explain everything about inanimate objects, everything about animals, and quite a lot about humans as well, including aspects of sensation and the interpretation of experience and so on. But he thought he could show that some aspects of human thought and consciousness and expression could not be explained in terms of this contact mechanics. He and his followers thought that the use of language was a case in point. They argued that a particular

mechanism in a certain environment (or in receipt of certain stimuli) will behave in a determinate manner, whereas human speech and thought are not determined and not random either; instead they are novel, creative, appropriate to situations, and they evoke in other minds thoughts similar to those in the mind of the speaker. This collection of properties, Descartes argued, is beyond the bounds of mechanics. Part of his reasoning had to do with will and choice. As one of his expositors (La Forge) said, Descartes held that under a certain stimulation a machine or an animal is compelled to respond in a certain way, whereas a human being may be "incited and inclined" to behave in that way without being compelled to do so; the human is not compelled if he is conscious of his own actions, and if his body can respond to his mind's commands. The distinction between being incited or inclined and being compelled is what requires us to go beyond the first substance, body, to the second substance, mind.

It may be possible to develop good predictions of human behavior, perhaps even perfect ones; but this would not bother the Cartesians. Their point, which they took to be immediately obvious, is that when you do what you were inclined to do you often could have done otherwise. That possibility of doing otherwise—however confidently and rightly it was predicted that you would do what you did do—is the crucial residue unique to systems that are not automata. The *principle of mind* was introduced to account for the limitations of mechanism—it was supposed to be a new, creative principle standing alongside the mechanical principle—and a new kind of substance, mind, was needed as a basis for it.

Descartes, incidentally, goes on to ask why we are only inclined rather than compelled, and why we may choose to do something other than what we are inclined to do. He answers that this question lies beyond the domain of intelligence; there is no way to answer it. But although we do not have intelligence enough to determine how the soul can choose beyond what it is inclined to do, it would be absurd to deny what we directly perceive to be true, simply because we cannot explain it. For Descartes there is only mind, not "the human mind"; so he is saying not just that the human biological system cannot understand this matter but that it is inherently beyond the possibility of comprehension. Restricted to the human mind, a specific biological system, his position is by no means an unintelligible one, and there may be some degree of truth to it.

It is worth noting that in some respects this line of thought resembles Newton's. Newton also began his investigations with a system of contact mechanics and observed that this system could not explain the motions of the heavenly bodies. He then proceeded to invoke a new principle which both he and his opponents at first thought to be occult and mystical, namely, action at a distance, which he showed could account for planetary motions. The Newtonian enterprise and the Cartesian enterprise have had rather different subsequent histories, with one providing the central direction of modern science while the other was put to one side. But this is not because of any difference in logical structure (for logically the two lines of thought are strikingly similar) but because the Newtonian system worked while the Cartesian one did not.



ISAACUS NEWTON EQ. AUR. ÆT. 83.

I. Vanderbank pinxit 1725 Geo. Vertue Sculpsit 1726.

The principle which Newton himself thought to be occult turned out to have great power to explain phenomena, and its coming to be regarded in the next generation not as occult but as normal scientific

commonsense was due primarily to its sheer explanatory success. The property of action at a distance was incorporated into an extended concept of "body." In fact, in the subsequent centuries (and here is the real force of the Newtonian revolution) the concept of body itself was continually revised to incorporate electromagnetic forces and massless particles and quarks and whatever will be discovered a decade or a century from now.

In fact, the Newtonian revolution made the concept of body disappear. What replaced it was the concept of *the physical world,* which simply incorporates *whatever we understand.* If we postulate things acting on one another at a distance, or massless particles, or antimatter, or any other strange things physicists may devise, as long as they contribute to an intelligible picture of nature we assign them to the physical world, as our equivalent of body. That means that there is no longer a mind-body problem. A coherent mind-body problem requires a concept of body which is determinate and unchanging, not our post-Newtonian concept, which is historically evolving and which incorporates whatever becomes moderately well understood. So now the Cartesian body-mind problem gives place to a new challenge. There is a real difference between inclination and compulsion, and I think it is obvious that we do things which we could have refrained from doing; furthermore, Descartes seems to have been right about the creative properties of thought and language—their innovative character, their appropriateness to situation, their nonrandomness, and so on. The challenge is to devise principles which will give an intelligible explanation of these facts, enabling us to bring human thought and behavior within the scope of the natural sciences. If we succeed in that, it may be by introducing new principles which extend the concept of body, or it may be by applying principles we already have. Either way, no new philosophical problem arises that does not already arise in the normal pursuit of the natural sciences. So the issue about dualism can be set aside, while we search for an intelligible account of human thought and behavior, assuming that any principles that we may need in such an account will somehow—in ways to be discovered through further research—be related to facts about elements of the brain.

Well, assuming that it makes sense to postulate and to study systems of mental representations, we then turn to the next question: Under what circumstances is it appropriate to do so? Not—it is widely thought—in accounting for the capacity to ride a bicycle, or the exercise of that capacity; but it does seem appropriate to postulate such systems in explaining the use of language. Exactly what distinction is represented by this decision? We can see that there is a valid conceptual distinction here by considering some reasonably clear cases. Consider the problem of designing a missle which is to be sent to the moon, and consider two different design proposals. One was proposed during the Second World War by B. F. Skinner; in principle it ought to work, although no one has tried it. The idea is to have a collection of pigeons in the missile. The image of the moon is focused in front of them somehow, and each one is trained so that if the missile veers off course in that pigeon's direction the pigeon pecks something until the image of the moon is centered again. This is a kind of servomechanism (it doesn't need pigeons and could be done

mechanically), and it ought to reach the moon. Now consider a second missile system, one that has a computer which has built into it a theory of the heavenly bodies, which includes information about the position of the moon, its velocity, the initial position and velocity of the missile, and so on. This missile begins to go in some direction and then calculates where the moon is going to be on the basis of the physical theory represented in its "brain," on the basis of those calculations modifies its course, and continues to do this until it reaches the moon. If we were simply observing the behavior of these two systems we might not be able to distinguish them, but still they would be working in completely different fashions. One would be a reflex device and the other a computational device involving something like a system of knowledge in which a particular task is executed by considering the representation of some item of knowledge, predicting future states of affairs in terms of that representation, and adjusting behavior on the basis of these calculations. No inanimate system is a proper model of the human being, but this one comes close in some respects. Now, however difficult it may be to distinguish the two systems by their behavior, they are functioning in totally different ways, and only one of them—the second—is in any way analogous to a system that makes use of mental representations.

One way to study such a system would be through an abstract characterization of its functioning in terms of specific types of representations; our task is to discover what those representations are, what elements they are composed of, what principles they operate under, and so on. That would be a perfectly legitimate, scientific, and intelligible enterprise raising questions of fact. Suppose, for example, that we are studying the system whereby humans interpret visible motion. If someone observes several presentations of an object in motion, he can determine that it is a cube, say, moving through space, and the question is, Ought we in this case to postulate a system of processing involving mental representations and computations on them? Or take the case of language: Is knowledge of language simply a practical ability to do so and so, a kind of reflex system, or is it instead a system of representation of knowledge? These are questions of fact, just as it is a question of fact whether a missile is of the first or of the second type. It may be hard to answer, but it is a complex question of fact involving the truth or falsity of certain abstract statements about the nature of certain physical objects.

In the two cases I have mentioned—interpreting visible motion, and language—the weight of evidence seems to be strongly in favor of postulating systems of mental representations. The only theories so far that have any explanatory force and any empirical justification are of that type, to my knowledge; that is certainly true in the language case. While it is commonly said that knowledge of language is a practical ability, every effort to give substance to that proposal has totally failed. And the idea seems misguided from the start. Surely, for example, two people can share exactly the same knowledge but differ widely in their ability to put it to use.

Assuming that it is legitimate to postulate mental representations sometimes, and that we can discover in what cases it is legitimate to

postulate them, we next ask, What is their nature? On this point we really have made substantial progress in the past few years, so that the picture today is a very different one from that of previous centuries. Unfortunately, just because this is the area where there is something substantive and detailed to say, I will have to avoid it because it is too technical to be appropriate here. But for certain systems—visual processing and language and a few others—there *are* nontrivial theories with considerable explanatory scope that deal with the nature of mental representations, the rule systems that characterize them, and the processing systems that deal with them. In the case of vision there are even the beginnings of an attempt to link the systems of computation with the neural structures involved in the processing. This is where the question of modularity arises—whether the mind is a single uniform system or instead, like the physical body, a system of organs or subsystems each with its own specific internal structure, properties, and interaction with others.

Historically it has generally been assumed that the mind is uniform and homogeneous. In the case of Descartes, this is fairly clear; he held that the mind has no parts, that there are no mechanisms in it; it is just a single substance. Descartes's empiricist opponents, such as Hume, also believed in a homogeneous cognitive system without subsystems. Passing over the intervening history and coming to the present century, we find that the same view is widely held. If we consider the spectrum of opinion stretching from Skinner's radical behaviorism at one extreme to Piaget's developmental constructivism at the other, we find that all along the spectrum it is assumed that the mind is a single uniform system. This is obvious in the case of Skinner (though he would not speak of mind). In the case of Piaget, one of the main principles he insisted on, if I understand him correctly, is that at each stage of cognitive development the operative principles are uniform across domains, and that what is achieved in one domain—let us say, to be concrete, the domain of language—must reflect earlier stages of cognitive development in prelinguistic systems, sensorimotor systems, and so on. This is a more complex theory of mind than Hume's or Skinner's, but it still assumes uniformity and homogeneity.

It seems to me that although we do not know very much, everything we do know suggests this picture is totally false. In the case of the visual system, for instance, the mental processing seems to involve a principle of rigidity, according to which the mind will automatically interpret successive presentations of an object as if it were a rigid object in motion. So, for example, if under appropriate conditions I present a plane figure perpendicular to your line of sight and then rotate it, you will presumably see it as a rotating plane figure, although the visual presentation is consistent with its being a plane figure shrinking to a line. Language, of course, does not involve anything remotely like a rigidity principle, but it does involve other principles—for example, about where categories must appear, and about when two referential terms may refer to the same thing—these being rather abstract and general, and possessed of a good deal of explanatory force in accounting for very specific properties of utterances, and for how such utterances are understood in a variety of languages. In the study of motor coor-

dination one finds other specific principles, and so on. The systems about which we have any understanding and insight seem to be highly specific, to involve their own unique principles, to develop in their own fashion at their own rate. They do interact: Since we can speak about what we see, there is obviously interaction between the language faculty and the visual faculty, for instance. But that is a far cry from the quite incredible idea that there is a single homogeneous system of mental representation, with a fixed set of principles, developing in a uniform fashion across domains.

It could be claimed that homogeneity prevails wherever we do not know anything, and that by chance it is only where we know something that the system is highly modular. But that would plainly be a rather weak claim. It is much more plausible to suppose that the brain, which may be the most complex structure in the universe, is, like other complex biological systems, differentiated into highly specific subcomponents which have their own mode of functioning and systems of interaction and so on. That is not to say that they are physically separated. This is an abstract discussion of the structure of this complex system, and I am suggesting that it is a system of interacting mental organs with their own intrinsic properties, but not that these organs are realized in discrete, nonoverlapping regions of the brain.

This topic, the nature of mental representations, that I have just briefly sketched is the real heart of the subject. It is the place where there are real results, and as I skip over it quickly I want to emphasize that.

Mental representations and the outside world

Now let me turn to the other major set of issues that I raised at the beginning: How do systems of mental representations relate to other things in the world? (I say *other* things in the world because mental representations are themselves things in the world.) Two kinds of relations should be investigated. The first has to do with the so-called causal role of mental representations, that is, with how thinking affects action. The second is the relationship between experience and cognitive structure, that is, the ways in which stimulations enter into and contribute to the formation of systems of knowledge and belief.

At a superficial level it is possible to discuss the first kind of relation, and there is a fair amount of literature which shows how. Suppose that I believe it will rain, and so I take an umbrella. We could describe this as follows. I have a belief that it will rain; this is encoded in a mental representation; and that mental representation causes me to take an umbrella. While you can tell the story in that way, and perhaps it even clarifies some questions to do so, it is pretty evident that this does not get to the heart of the matter. It does not begin to answer the questions the Cartesians raised: What is the relation between what we are "incited and inclined" to do and what we choose to do? How does inclination differ from compulsion? What is that crucial difference, that residue, that involves choice and will?

The Cartesians regarded these problems as unsolvable, and it would be hard to argue that they were wrong. Anyway, no progress at all has

been made toward answering these questions, it seems to me; we are today exactly where the Cartesians were with regard to our understanding of this matter. If that is right, what explains it? One explanation could be that our situation is like that of physicists before Newton—it is just that nobody has yet come up with the smart idea. Or perhaps the explanation lies in some version of Descartes's view that we do not have intelligence enough to understand how the soul chooses among courses of action. While I do not think that it makes sense to pose the question exactly in his terms, we may be able to recapture something of the same insight in slightly different terms. Recognizing that we are part of the natural world and that our intelligence is simply a contingent biological system, just as our circulatory system is, we may ask whether our concepts and modes of thought and categories of explanation are adequate for all questions. Probably they are not. What understanding we have of how knowledge is acquired, and how convergence is achieved in selecting explanatory theories, indicates that our very success in certain areas makes it likely that we will fail in others. Any explanatory success involves fixing upon explanatory theories which go far beyond the evidence available, and these must be selected from a vastly larger set of theories all of which are consistent with the evidence. Our ability to make such selections in a narrowly limited if not entirely uniform fashion means that there is some built-in constraint on our capacity to form scientific theories. That constraint, which enables us to succeed in some domains, must condemn us to failure in others where the right answer lies beyond the limits of our theory-forming capacities. If the mind were quite unstructured and thus capable of everything, it would not really be capable of anything.

There is nothing strange about this. In fact, there is an obvious language analogy. We have the capacity to acquire knowledge of our own language on the basis of very fragmentary evidence; and we all do it in more or less the same way, which is why we can immediately understand new expressions and can be understood when we produce new expressions which have no resemblance to expressions that we or others have heard before. Our ability to develop in a uniform way a very rich system of knowledge, extending vastly beyond any experience we have had, is explained by our being specially designed to create certain systems of knowledge and not others. But that very special design means that there are other possible languages which we could never learn and would always find baffling if we encountered them, because our special design excludes their grammars. This relationship between scope and limits is an inescapable, logical one; the richer the systems of knowledge we can have in certain domains, the narrower the limits of those domains. And it could turn out to be true that for many of the questions which interest us—about the relations between human thought and action, for example—our minds are not designed to let us understand the answers. Some Martian with a differently organized intelligence, discovering how we have been posing these questions for several thousand years, might see immediately, that we are posing them in the wrong way; and this wrongness may be forced upon us by the very intellectual structures which give us our successes in other domains.

With regard to the second kind of relation—that between evidence

and knowledge—there has been a good deal of progress. It has come from proposing very general principles concerning the kind of knowledge that can be attained—like the rigidity principle in visual processing, or the principles of universal grammar—and then trying to show that the fragmentary evidence available suffices to set these principles into operation so that a particular system of knowledge results—about the behavior of objects in the external world in one case, about the properties of expressions of some language in the other.

Mental representation and language

I now turn to some questions about the structure of one particular system of knowledge, namely, our knowledge of language. This was the topic of the seminars and discussions which I conducted while at Syracuse University, but those treatments are too technical to be presented here. However, there are some things to be said about language, and knowledge of it, at the level of generality of my discussion so far.

The very recent period has been one of significant progress in the study of language. During the past quarter century new areas of inquiry have been opened up to serious investigation and many classical questions have been reformulated from a novel point of view, which has led to a much better understanding of the nature of human language. The study of language has a long and rich history, extending over 2,000 years, but the current period is in many ways a new era. The current rate of change is rapid, even by the standards of the past generation, so that the linguistics of a generation from now may prove to be as unrecognizable to us as the linguistics of today would be to the linguists of the not too distant past.

In part, these changes reflect a shift to a representational theory of mind, and to a mentalist or conceptualist interpretation of the study of language. The significance of this shift was not really clear when it happened, about twenty-five years ago. I think the scope is much broader than has been appreciated.

In my opinion the shift toward a mentalist conception of language is a shift toward the point of view of the natural sciences, offering the prospect of eventually assimilating the study of language to the mainstream of the natural sciences, in contrast to the operationalism and behaviorism which dominated the investigation of language and related areas of psychology about a generation ago. In my view, operationalism and behaviorism reflected a serious misunderstanding of the nature of rational inquiry and amounted to little more than a series of dogmatic constraints on legitimate theory construction; and they have collapsed largely for this reason. Without saying more about that, I shall consider a different point of departure from which we can gain some understanding of the changes that have been taking place in the study of language.

I will use the term "generative grammar" or simply "grammar" to refer to the system of rules and principles that constitutes a person's knowledge of language and that forms the various mental representations that enter into the use and understanding of language. The new field of generative grammar took seriously a crucial insight that was expressed in the early nineteenth century by the great German linguist

and humanistic scholar Wilhelm von Humboldt, who observed that language involves "the infinite use of finite means." Humboldt is certainly right about this, but the technical tools needed to give his insight substantive content were not available in his day; now, to a certain extent, they are.

What does it mean to say that language involves the infinite use of finite means? Well, a person's knowledge of language is a state of the mind, a relatively stable component of transitory mental states. And it is coded somehow in a finite physical brain: It is a finite system, realized in some arrangement of physical mechanisms. This finite system is what we refer to as a "generative grammar," and theories about its nature are assertions—either true or false—about the nature of a certain system of mechanisms.

So much, for the moment at least, about the "finite means." What about the "infinite use"? People who know a language can produce and understand sentences that they have never heard and that do not closely resemble any they have heard, and this capacity has no bounds. Apart from limits of time and attention, there is no finite limit to the number of sentences that are assigned a specific and definite meaning and structure by the system of knowledge that each person possesses.

Traditional grammars, or the structuralist grammars of the early and mid-twentieth century, could not come to terms with this property of human language. A traditional grammar, however voluminous, relies crucially on the intelligence of the reader—a contribution that the grammar takes for granted without analyzing it. The grammar will present many examples, along with general observations about the language in question, and on this basis the reader is supposed to be able to determine what an arbitrary sentence of the language means and what its structure is and to construct properly formed sentences expressing his thoughts. Well, how does the intelligent reader of the grammar succeed in this? That question was never answered—indeed it was not raised or seen as a problem—in the traditional study of language. But it does constitute a problem, and structuralist grammars in our century began to confront it, in that they offered fairly explicit procedures in an effort to determine certain properties of certain aspects of linguistic structure from observation of sample data. But the approach that was developed was in many ways wrong, as is now quite clear; so the infinite use of language remained a mystery.

Quite independently of linguistics, logicians and philosophers, beginning with the pioneering work of Frege about a century ago, were developing ways of thinking about language-like systems, primarily systems of mathematics and logic, that make it possible to start giving substantive meaning to the Humboldtian formula. In the mid-1950s these two intellectual currents merged, as ideas adapted from the study of the formal systems came to be applied to the far more complex systems of natural language. In the new field that developed from this confluence of intellectual traditions, the basic concept was that of generative grammar—an explicit system of rules and principles that assigns a specific representation of sound, meaning, and structure to each of an infinite array of sentences. We assume the grammar to be neurally encoded: The linguist's grammar, then, is a theory of this internalized,

neurally coded object, which may properly be regarded as constituting a system of knowledge. It is a finite object. But its rules for using and understanding language are unbounded in scope, and so it makes possible the infinite use of its finite means. And the linguist's grammar specifies these rules perfectly explicitly, not surreptitiously relying on the reader to contribute something to make the grammar work. Of course, this is only an expression of a goal of linguistic research, which is far from having been achieved. But it was important to get the basic problem formulated in a clear way, and this has generated questions the study of which has led to definite though limited progress. Notice that this progress does not reach the Cartesian problem of the *intelligent* use of language.

One of the questions is, How does a person attain this knowledge? This is one instance of a very old problem, which could be called Plato's problem. Plato held that much of our knowledge is inborn, remembered from an earlier existence. Leibniz, too, thought that our knowledge of arithmetic, geometry, and the in-built principles of the sciences, as well as some practical knowledge, is innately possessed by us, though not in a clearly articulated form. He did, however, want this innateness "purged of the error of pre-existence." Within a different tradition, Hume spoke of those parts of our knowledge that are derived "from the original hand of nature." Like Leibniz, Hume regarded such innate knowledge as a sort of instinct. I think that these formulations are basically correct, but they have never been considered satisfactory. Thirty-five years ago Bertrand Russell again raised Plato's problem: "How comes it that human beings, whose contacts with the world are brief and personal and limited, are nevertheless able to know as much as they do know?"[1] The question cannot be brushed aside lightly.

1. Bertrand Russell, *Human Knowledge* (New York: Simon and Schuster, 1948), p. 5.

It arises in a clear and explicit form when we undertake the study of generative grammar, and much of the interest of this study lies in its bearing on the classical problem. It will not yield a comprehensive answer to the problem, but it may be enlightening with respect to certain central points. It is clear that the highly articulated and subtle system of knowledge and understanding that each person has attained is vastly underdetermined by the evidence available, while much of what is known appears to be based on no specific evidence at all. The only plausible idea that has been advanced to deal with this striking fact is that the human biological endowment includes a system of principles that determines the basic structure of the grammars of attainable languages, while experience serves merely to refine and to sharpen these principles. The human genetic endowment obviously does not determine whether a child will acquire English or Chinese, but it does determine that the languages that can be acquired in the normal way will be of a strictly limited type. It is because this endowment is so rich that we can know so much when our contacts with the world are so limited, and that is also what enables us to share so much knowledge and belief with others when our personal contacts with the world are so different from theirs.

One central area of investigation, then, will be what we may call "universal grammar," now giving a traditional phrase a rather new

sense: Universal grammar is a system of genetically determined principles, each of which incorporates certain parameters—certain possibilities of "fine tuning" which can be set one way or another on the basis of experience. The combination of innate principles and socially determined fixing of parameters results in the generative grammar of a particular language, the system of knowledge attained by each person under normal circumstances. The linguist's grammar for a particular language is just the specification of a particular set of values for the parameters of universal grammar. It is clear that the system of universal grammar is fairly intricate, so that each choice of a value for a parameter has complex effects which proliferate through the system. Change in a few parameters may yield languages that appear to be typologically quite different, though they are cast in the same mold. Something of this sort has to be correct, given the apparent diversity of the languages which each person is capable of learning on the basis of limited evidence. The dimensions and the character of this problem are only beginning to be grasped in recent work.

Are there general principles of learning that enter into the process of language acquisition and guide the setting of parameters? There are some plausible suggestions that have emerged in recent work. One such principle is what Robert Berwick at MIT has called the "subset principle." The basic idea is this. Suppose that a certain parameter has two values (call them + and −), and that the choice of + for the parameter yields all the sentences that − yields and some more as well. Thus, if we select + we get a "bigger language" than if we select −. By the subset principle, the child will set the value at − unless presented with evidence for + in the form of sentences that fall within the larger language. That is, children choose the smaller language unless they have evidence that what is used in their society is the larger one. It can be shown that the subset principle is needed if knowledge is to be obtainable on the basis of only positive evidence (i.e., evidence that something *is* a sentence) as distinct from negative evidence (i.e., evidence that something is not a sentence). And there is fairly good evidence that negative evidence is not needed for language acquisition. Children are sometimes corrected by their parents, but such corrections seem to play no crucial role in the acquisition of knowledge of the language.

Berwick has also shown that his subset principle lets him explain in a unitary way a variety of data which had previously been dealt with ad hoc and piecemeal. Here is a simple example. Consider the sentences (1) "John wants to win the race" and (2) "John wants Bill to win the race." Each has the form "John-wants-clause," but in (2) the clause is "Bill to win the race" while in (1) it is "to win the race" with an understood missing subject, a pronoun referring to John: John wants *John* to win the race. It is clear that (2) reflects the actual meaning and structure more closely than (1) does, because (1) has suppressed something which (2) leaves out in the open. Yet children appear to use sentences like (1) before they use the likes of (2); and across the languages of the world (2) seems to be far less common than (1)—you cannot say the equivalent of (2) in French or Spanish, for example—although it is (2) which seems more transparently to reflect the actual meaning being expressed. In technical terminology, we say that struc-

tures like (2) are more *marked:* They are less common in the languages of the world of this structural type, and in languages which do have them they are acquired late.

Those are the facts. What do they mean? They mean there is some parameter, *P,* that differentiates English from French or Spanish, and English is more *marked,* more unusual, than French or Spanish in this respect. If we set *P* at +, we derive a language like English in which infinitival clauses may have overt subjects; if we set *P* at −, we get a language like French in which infinitives may not have overt subjects. Obviously, the value + gives the larger language: It permits sentences which the value − disqualifies. By the subset principle, the child will automatically set the parameter at − unless and until presented with positive evidence to the contrary in the form of sentences like (2), "John wants Bill to win the race."

That is just one example of the application of the subset principle. We have here a plausible candidate for a principle of general learning theory, and I do not know of any other plausible candidates for that title. For many years I assumed that there would never be any such thing as a theory of learning, involving significant general principles bearing on the acquisition of rich cognitive systems—but now I am not so sure. Notice, however, that the principle can do some real work only when embedded in a rich and articulated special theory of a specific cognitive domain—language, in this case. It is not a "generalized learning mechanism" of the sort that many people have believed to be operative across the board but is, instead, a modular principle of learning.

I have been speaking of the nature of acquired knowledge, and of how it is acquired, the latter being a variant of Plato's problem. A second major problem is that of understanding how the acquired system of knowledge is put to use. This breaks down into a number of subproblems, the first being that of determining how the grammar assigns a specific structure and meaning to a presented linguistic expression. This is sometimes called the parsing problem. (A second problem is to understand how we use language to express our thoughts, communicate with others, and so on. The parsing problem has proved amenable to serious inquiry, but this second problem resists our understanding in quite fundamental respects. As I mentioned earlier there has been little progress in answering—or even clearly formulating—questions that involve will and choice, of which this is one.)

The way of studying language that I have outlined had a different perspective from other contemporary approaches such as structuralist grammar and behaviorist psychology. These regarded language as an externalized object—a collection of behaviors, of actions, of sounds, of sounds paired with meanings, or whatever—and regarded a grammar as a collection of statements about the language, which is the real object of study. That would make universal grammar a collection of statements that are true of many or all languages. According to the approach of generative grammar, on the other hand, the objects of study are *grammar,* the system of knowledge represented in the mind and the brain in the eventual steady state, and *universal grammar,* the system of principles represented in the genetically determined initial states.

Grammar and universal grammar are real objects, part of the physical world. In contrast, the concept "language" as an external object is derivative and in fact has no very clear or definite meaning. The system of knowledge—the grammar—attained by a particular person assigns a certain status to every relevant sound wave. Some are assigned no phonetic representation at all; some are assigned a phonetic representation but no linguistic structure (i.e., are recognized to belong to some human language but not mine); some are assigned a full linguistic representation with literal meanings, or figurative meanings, or are recognized to be deviant though still intelligible, and so on. There are many such possibilities. How one draws the boundaries of "language" is not a very significant question, because a language in this sense is not a real thing in the world, unlike particular grammars and universal grammar.

The technical concept of language as an externalized object treats language as an artificial abstract construct. The intuitive concept of language, as it occurs in informal pretheoretical usage, is much closer to the technical concept "grammar" than it is to the technical concept "language" considered as an externalized object. When we speak informally of a person as knowing a language, we do not mean that he knows a set of sentences, or of behaviors, or of sound-meaning pairs, but that he knows what it is that makes sounds and meanings pair up in a certain fashion. That is just to say that he knows a grammar. So the shift of perspective from the technical concept language to the technical concept grammar—taken now as the object of inquiry is a shift toward realism in two important respects: It is a shift toward the study (*a*) of a real object rather than an artificial construct, and (*b*) of what we really mean by "a language" or "knowledge of a language" in informal usage.

Some people see a paradox in the claim that the shift from viewing language as an externalized object to viewing it as a mentally represented object is a shift toward realism, and that this is a shift toward incorporating the study of language within the natural sciences. I want to comment on this. There are two questions here. Is the concept of language as a mentally represented object something that could in principle be incorporated within the natural sciences? And is the concept of a set of sentences an abstraction that is more remote from physical mechanisms than the concept of a mentally represented grammar? We should answer yes to both questions. Obviously there is some physical encoding of knowledge of language in the brain, and when we formulate the rules and the principles of grammar we are trying to describe these actual mechanisms at a suitable level of abstraction. And the concept "set of sentences" (or other versions of language as an externalized object) is at best derivative from the concept of grammar and at worst not legitimate at all.

So the shift toward a mentalist picture, in which the object of study is the grammar that is really there in your brain and mine, really is a shift toward bringing the study of this cognitive system within reach of the natural sciences. Outside of that shift, there would be no hope of doing so. There is, then, no paradox.

Syracuse Scholar (1979-1991), Vol. 4, Iss. 2 [1983], Art. 2

In this regard, the study of formal language has been rather misleading, and I think that I played some role in helping to mislead myself and others. A formal language such as that of arithmetic is generally viewed in the following way. There is a specified finite notation, and of the arrays of items in that notation an infinite subset are the sentences of the language. This subset is generated by a set of formation rules, and one set of rules is as good as another so long as they yield the same sentences. Now, as I mentioned earlier, the modern study of generative grammar developed from a confluence of two intellectual traditions—one concerned with natural language, the other with formal languages. It was very natural to take over from the latter the idea that a language is just an infinite set of sentences, or of sentence-meaning pairs, so that the grammar is just something which *somehow* picks out this set of objects. Natural as this was, however, it was entirely misguided.

Let me return to Humboldt's aphorism that language involves the infinite use of finite means. The problem raised by this idea did not dominate subsequent research, as it deserved to do, but it was occasionally recognized and discussed loosely. One of the discussions appears in the important book *Philosophy of Grammar* by Otto Jespersen, written about sixty years ago. Jespersen observes that each speaker is able to abstract from presented sentences "some notion of their structure which is definite enough to guide him in framing sentences of his own,"[2] these being "free expressions" rather than frozen forms like the words that we learn. The statement is unquestionably correct, and it is a version of what Humboldt said. But the question was still not squarely addressed, as noted earlier. In the even more influential Saussurean structuralism, the whole question of free expressions was placed outside the scope of the study of language structure, what Saussure called *langue*. I think this is the fundamental defect of the structuralist tradition. Even in very recent years it has been argued, for example, by Charles Hockett at Cornell, that the construction of free expressions is simply a matter of analogy. This is not false, but it is vacuous until the concept analogy is given some sense.

2. Otto Jespersen, *The Philosophy of Grammar* (London: George Allen & Unwin, 1924), p. 19.

From another point of view, W. V. O. Quine has taken exception to the idea, expressed by Jespersen and others, that the speaker is "guided" by an unconscious "notion of structure" in forming or interpreting free expressions. Quine argues that this is "an enigmatic doctrine" and perhaps pure "folly," and that we may legitimately speak of "guiding" only when rules are consciously applied to "cause" behavior, which is not what happens in the ordinary use of language. Although we do not have guiding, he says, we have "fitting": We can speak of behavior as fitting one or another system of rules but must refrain from imputing any kind of psychological reality to any such system.[3]

3. W. V. O. Quine, "Methodological Reflections on Current Linguistic Theory," in *Semantics of Natural Language*, ed. Donald Davidson and Gilbert Harman (New York: Humanities Press, 1972), pp. 442–55.

Similar attitudes are revealed in some recent approaches to the theory of meaning, such as those inspired by the work of Donald Davidson. For example, Michael Dummett describes Davidson's approach as holding that "the proper method" for the study of meaning "is to ask, for any given language, what body of knowledge would be required for someone to be able, in virtue of his explicit possession of that knowledge, to speak and understand the language. Here it is not maintained that any actual speaker really has such a body of knowledge,

however tacitly or implicitly." What the speaker does "fits" the theory, but we do not go on to say that the speaker actually has the body of knowledge expressed in the theory. Dummett concedes that this "is somewhat roundabout unless ability to speak a language actually does involve having such knowledge."[4] So he appears to hold that it would not be "folly" to attribute possession of such knowledge to the speaker, but that somehow it is illegitimate, presumably because something—some kind of relevant evidence—is lacking. This is a curious construction. There seems to be a kind of miracle or mystery surrounding a body of knowledge which has the property that *if* you possessed it you *would* then speak and understand the language, although you actually don't possess it or we cannot properly ask whether you possess it because some relevant kind of evidence is lacking. Philosophers will recognize in this a very powerful strain in the current tradition. It is, in my view, just another reflection of the pernicious behaviorism that we seem unable to escape.

4. Michael Dummett, "Objections to Chomsky," *London Review of Books*, 3-16 September 1981, 5-6.

I think that the qualms of Dummett, Quine, and others are misplaced. Dummett asks us to seek a formulation of the body of knowledge that would be required to understand "any given language." But what does it mean to refer to "a given language"? An infinite class of expressions, or expressions paired with meanings, or use conditions, or whatever you like, is never "given." We can play this game when talking about the language of arithmetic, because we think we have some objective concept of what that system is. But we do not in the case of English. There is no sense in which the set of sentences of English or any other language, or indeed any infinite object whatsoever, can be "given." What is "given" is some finite object—for example, a finite amount of experience, or a grammar, that is, a finite representation of what is in the mind and brain. So there can be a "given language" only if a language is understood to be not an infinite set of sentences but rather a grammar. And a grammar is something that can be known.

What about the idea that some needed kind of evidence is lacking that might entitle us to accept some particular theory about what this knowledge is? I can make nothing of that. I cannot imagine what kind of evidence could be lacking such that if it were available it would enable us to make this otherwise forbidden leap to attributing the knowledge to the speaker. Of course, we will always want more evidence, and of more varied types, but there is no defect in principle in the kinds of evidence that we have.

As for Quine's conclusion, he assumes that if behavior is not "guided" by consciously adopted rules, then at best it merely "fits" rules stated by a theorist about the behavior. But why should we accept this doctrine? Behavior is guided, so it appears, by the rules and principles of the system of knowledge, and these are mostly not accessible to conscious awareness. This conclusion seems perfectly intelligible and is currently the only one that seems to be at all warranted by the known facts. Of course our behavior is not "caused" by our knowledge, or by the rules and principles that constitute it. In fact we do not know how our behavior is caused, or even if it is caused, but that is another matter entirely.