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.

MICHAEL D. ROOT

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.

almost no one wants to say that the behavior of bodies in free fall and of gases under increased pressure are rule-governed activities. Similarly there are regularities in the light emitted by a firefly during a brief flight, regularities in the feeding reactions of thrush nestlings, and in the zigzag dances of male sticklebacks. In man there are regularities in salivating and hand-out-of-the-fire responses, and there is a regularity in the upward kick of a crossed leg when a mallet is applied briskly to a spot under the kneecap. Yet we would not be prepared to say in all of these cases that the response or the behavior is guided by a rule.

Presumably when a rule theorist argues that speaking a language and doing arithmetic involve rules, he means that the rules play a role in the etiology of that behavior. The rules somehow cause or enable us to do what we do. Rules like beliefs can be the determinants of behavior.

However, when the rule theorist says that rules determine a subject's behavior, he doesn't mean that the subject is always or ever aware of these rules. The rule theorist denies that everyone who employs a rule for summing arithmetic series is able to answer the question, "How does one find the sum of an arithmetic series?" That is, a rule may guide a subject's behavior and the subject may not be conscious of the rule. In fact the rules that guide our behavior are often rules of which we are not aware.

What the rule theorist doesn't tell us, however, is how to justify talk about guidance by rules. That the rules describe the behavior is obviously not enough. What is needed is some sharper analysis of the notion of someone's behavior being guided by a rule. What is the difference between a rule's fitting someone's behavior and its guiding that behavior? As George Miller has said: "In spite of our increasing reliance on rules as explanations of thought and behavior, I do not know of any clear account of what rules are and how they function."[3]

In sections II and III of this essay I take up the question, "How does one justify a claim that what an organism does it does because it is guided by a rule?" Rule-guided behavior, I argue, displays a pattern of organization that is characteristic of language. The behavioral events that make up a rule-guided sequence of behavior are not selected in serial order. There are dependencies between distant events similar to the dependencies between distant words in a sentence. Most important:

[3] George A. Miller, "Four Philosophical Problems of Psycholinguistics," *Philosophy of Science*, 37, no. 2:191.

if there is no limit on the distance between dependent events, the etiology of the sequence must include a rule and we say that the behavior is rule guided. Here we have a sufficient condition for ascription of "rule guidance."

Language behavior is most often offered as the clearest example of rule-guided behavior. However, one result of our discussions of rule guidance is that the ability to speak and understand a language turns out to be more like our other abilities than first thought. Rules guide speech but they guide many of our other performances as well.

In section IV, I consider Chomsky's claims that man's ability to speak and understand a natural language represents a unique type of intellectual organization that manifests itself in what he refers to as the "creative aspect of language use." Language use is creative, Chomsky tells us, because language is unbounded in scope and stimulus-free. But the features that mark creative use are only coarsely drawn in these discussions. Once they are drawn more clearly, we see that, despite Chomsky's claims, language is not unique in having a creative use. Speaking a language again turns out to be a less distinctive activity than first thought.

I

Psychologists assume that behavior can be analyzed and that fixed, elementary units can be singled out. Descriptions of behavior are descriptions on this vocabulary of elementary units. However, two important questions remain, and of late these questions have received considerable attention: (a) what are the elementary units, and (b) on what principles of concatenation are complex behaviors built? Many of the interesting results that shed some light on (b) have come from linguistics.

Linguists have shown that there is a structure to most sentences that is independent of their linear structure. That is, two linguistic expressions may contain the same morphemes and the morphemes may occur in the same linear order and yet the expressions may differ in syntactically significant ways, for example, the expressions "*old men* and women" and " old *men* and *women*." Thus sentences like

(1) John gave ten dollars to Harry and Phil
(2) Intriguing women can be a drag

are ambiguous and the ambiguity does not arise because the meanings of some of the words are ambiguous, but because different syntactically significant structures can be associated with the same strings.

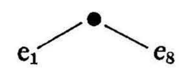
What has not been as often noticed is that these linguistic results can be easily generalized and quite naturally applied to nonlinguistic behavior as well. Most forms of complex nonverbal behavior have a nonlinear, inner structure of their own and the significance of the behavior for the organism depends on this structure. Among the first American psychologists to make this point was Karl Lashley.

> From such considerations, it is certain that any theory of grammatical form which ascribes it to direct associative linkage of the words of the sentence overlooks the essential structure of speech. The individual items of the temporal series do not in themselves have a temporal "valence" in their associative connections with other elements . . .
>
> This is true not only of language, but of all skilled movements or successions of movement. In the gaits of a horse, trotting, pacing, and single footing involve essentially the same pattern. . . . The order in which the fingers of the musician fall on the keys or fingerboard is determined by the signature of the composition; this gives a set which is not inherent in the associations of the individual movements.[4]

Nonverbal behavior, then, may be ambiguous in much the same way that sentences are ambiguous. That is, two superficially identical bits of behavior may be derived from two different underlying cognitive configurations. Two bits of complex behavior may include the same behavioral events and the events may occur in the same temporal order and yet the bits of behavior may differ in cognitively significant ways. Consider, for example, the following sequences of behavioral events: *B* walks out the front door of his house, walks to the end of the block, notices a newsstand, buys an evening paper, stops and speaks to a passerby, turns, walks back down the block, and enters his house through the front door. This description may equally well fit the following three cases: (a) *B* makes it a practice always to return home by the same door from which he leaves; (b) *B* was out for a stroll and bought a newspaper as a second thought; and (c) *B* went out in order to buy a newspaper. If (a) or (c), then there is an order to the previously specified sequences of behavioral events that is independent of the linear order. There are significant relations between events that are not temporally adjacent. In order to give a more complete and unambiguous description of *B*'s activities, we need to take account of underlying structure. A description sufficient to distinguish (a) from (c) will need to include a description of hierarchical relations between the events. In short, *B*'s walk to buy a newspaper is hierarchically ordered in the way in which the sentences of his language are hierarchically ordered, and while we need to appeal to that ordering in order to disambiguate many of those sentences, so we need to appeal to underlying structure and hierarchies in order to disambiguate *B*'s nonverbal behavior, his walk and purchase of a paper.

[4] K. S. Lashley, "The Problem of Serial Order in Behavior," in *Cerebral Mechanisms in Behavior*, ed. L. A. Jeffress (New York: Wiley, 1951), p. 116.

We can draw the linguistic analogy even farther, for we can represent behavioral organization in the same manner in which we represent the constituent structure of linguistic expressions, for example, tree diagrams or labeled bracketings. In an unambiguous description of *B*'s behavior, we need to indicate which events depend on the character of other events in the sequences, and we can do this for nonadjacent events by connecting the dependent events to a node that lies outside the linear sequence itself. The event of *B*'s walking out the front door of his house, e_1, and the event of *B*'s walking back in by way of the front door of his house, e_8, are dominated by a common node



and are thus coordinated even though they are a considerable distance from one another in the temporal chain of behavioral events. What a tree diagram for a temporally ordered sequence of behaviors will indicate is that there are psychologically significant relations between events that are independent of their temporal order. What the node represents is an integrative function that is independent of the associations between temporally adjacent events. As the example illustrates, a subject's selection of one event will often determine or restrict the choice of proceeding events that are far removed in space and time. Again the parallel with the structure of natural language is obvious. In English choice of verb determines or restricts choice of complement and, similarly, it often determines or restricts the choice of subject. We say (3) but not (4) and we say (5) but not (6).

(3) John broke the bottle

(4) *John broke the water
(5) John who loves dogs also love cats
(6) *John who love dogs also love cats

In *Syntactic Structures* Chomsky noted that dependencies between morphemes, between elements in a terminal string, often stretch across a number of other words and that there is no upper limit on that number. There are, for example, English sentences of the form

(7) If S_1, then S_2

and these sentences include a dependency between "if" and "then," and yet S_1 and S_2 may in turn have other sentences within them and those in turn sentences within them and so on.

A similar case can be made out for most forms of complex nonverbal behavior. Complexity, we noted, frequently takes the form of a hierarchy. The behavior involved in leaving one's house in order to purchase a newspaper is hierarchically ordered. Moreover, there is no upper limit on the number of behavioral events across which these motor coordinations or behavioral dependencies can stretch.

Let me illustrate this last by way of another example. You learn to knit and then undertake to knit your husband or wife a sweater. Your plan is first to knit the left sleeve, next to knit the front and back panels, and then to knit the right sleeve. You proceed to execute your plan and knit the left sleeve to some desired length *n* and proceed to complete the center panels. Finally, you turn to the right sleeve. How you knit this last sleeve will depend on how you knitted the first. You want the patterns, the stitches, the colors, and length to be the same for each. Although the sequence of events that constituted the knitting of the first sleeve is far removed from the events that need to be included in the knitting of the last, they depend upon one another. You want to coordinate what you do now with what you did earlier. Of course, you could return to the left sleeve, rehearse your earlier behavior, and follow each "operation" on the first sleeve with an appropriate bit of knitting behavior on the second. However, this is a needlessly onerous and time-consuming strategy. It is likely that you can coordinate the knitting of the two sleeves with very little explicit cross-reference. If indeed you are following a rule, it is the rule that provides the basis for the coordination and not the individual units of behavior. In any case, knitting the sweater includes coordination of events that are discontinuous and

the events that are distant need to be coordinated. Further, there is no upper limit on that distance. You can interrupt knitting the sleeve in order to complete the front and back panels, but can also take time out to write a book about knitting, brush your teeth, prepare dinner, or sail to Calais. There is no longest interruption.

There are then nonverbal behaviors of the form

(8) $X_1 + X_2 + X_3$

and these behaviors include dependencies between events in the sequence X_1 and events in the sequence X_3. More importantly, the dependencies cut across the sequences of events X_2 and yet this last sequence of events may itself include behavior of the form (8) and so on. Thus knitting behavior can include an unlimited number of what linguists call "nested dependencies" and a structural tree for the behavior involved in

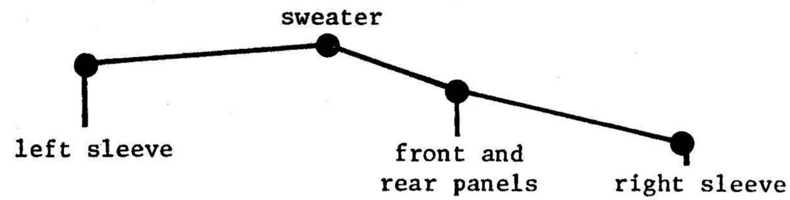


Figure 1

knitting a sweater might look like the diagram in Figure 1, and this, of course, resembles the configuration assigned to sentences of the form of (7).

II

It is possible to make the parallels between the structure of language and the structure of many forms of complex behavior more precise. Specifically, we can define a relation on hierarchically ordered behavioral events that corresponds to a relation that Chomsky[5] has defined on elements in terminal strings.

Suppose we have a natural and antecedently agreed-upon taxonomy of behavioral events. For example, we might count drawing as a type of behavior and each of the following as (1) a type of behavior and (2) a proper part of the type of behavior that precedes it: drawing a face, John's drawing a face, John's drawing a face with prominent features.

5 Noam Chomsky, "Three Models for the Description of Language," *I.R.E. Transactions on Information Theory*, IT-2, no. 3, pp. 113–124.

Or consider: entering and leaving, John's entering and leaving, John's entering and leaving by the same door.

Let K be a class of behavioral sequences each of which (a) a subject S is able to perform and (b) counts as an instance of the same type of behavioral event. Let $Y_1, Y_2, \ldots, Y_n = X$ and X be a sequence in K. We say that X has an (i,j)-dependency with respect to K if and only if:

(i) $1 \leqq i$ and $i + m < j \leqq n$ where $m \geqq 1$,

(ii) there are behavioral events Z_i, Z_j which S is able to perform and $Z_i \neq Y_i$ and $Z_j \neq Y_j$,

(iii) Z_i, Z_j have the property that X' is not while X'' is in K where X' is formed from X by replacing the i^{th} behavioral event in X (namely, Y_i) by Z_i and X'' is formed from X' by replacing the j^{th} behavioral event in X' (namely, Y_j) by Z_j.

In other words, a behavioral sequence X has an (i,j)-dependency with respect to K if and only if replacement of the i^{th} event Y_i of X by Z_i ($Z_i \neq Y_i$) requires a corresponding replacement of the j^{th} event Y_j of X by Z_j ($Z_j \neq Y_j$) for the resulting sequence to belong to K. Now let us assume that sequence X and every other sequence included in K is (i,j)-dependent with respect to K. For simplicity we might assume that K consists of X and all and only those sequences that result from X by means of the following steps: we select those pairs of events that are not identical with Y_i, Y_j but whose substitution for this last pair will preserve membership in K; we replace Y_i, Y_j with each pair of these events. I assume that there are only finitely many such pairs.

I wish to consider a subject, S, who has the ability to perform every sequence in K and who has the ability to perform them in circumstances in which it is possible for S to perform sequences like X'. That is, I wish to rule out cases in which S is able to perform all of the sequences in K but only because performance of any other sequence, e.g., X', is excluded by external circumstances. An example here would be helpful. Consider those behavioral sequences that we would call "entering and leaving by the same door." K consists of all and only the sequences that satisfy this description. Each of these sequences will be (i,j)-dependent with respect to K. We will let X here be a sequence in which one enters by the front door and leaves by the front door. Event Y_i is entering by the front door and Y_j is leaving by the front door. Events Z_i and Z_j can be entering by the back door and leaving by the back door. Sequence

X' will not be in K while X'' will. The subject I wish to consider is one who is always able to enter and leave by the same door but one who is able to do this in quarters that contain more than one door, in quarters that allow him to enter and leave by different doors.

In these circumstances, where the subject has a choice in his enterings and leavings, the ability to perform all of the sequences in K requires that the subject's choice of an i^{th} event determine his choice of the j^{th} event. When S goes to leave he cannot leave by just any door and of course he cannot decide willy-nilly to leave by the window. Our subject has to leave as he entered and this requires an agreement between i^{th} and j^{th} events.

Finally let us assume that there is no limit on m in the expression "$i + m < j \leqq n$" and that there is no longest sequence in K. We will assume that K is an infinite class of (i,j)-dependent behavioral sequences and S can perform each of the sequences in K. To return to our example of entering and leaving by the same door: to say now that S can perform all of the sequences in this class is to say that he can enter by the front door, walk up and down a flight of stairs, and still manage to leave by the same door. To say here that there is no limit on m is to say that in principle there is no limit on the number of times that S can repeat the walk-up-and-down-the-stairs cycle. Indefinitely many trips up and down the stairs may intervene between the i^{th} and j^{th} events, between entering by the front door and leaving by the front door. Nevertheless S in being able to perform all of the sequences in K is always able to get his enterings and leavings to agree. The recursion of trips up and down the stairs results in infinitely many sequences that begin with an entrance and end with a departure. S is able to perform them all. This is no simple task. It requires the employment of what I want to call "a rule."

III

We return now to the question of rule guidance. How do we justify a claim that a rule is guiding an organism's behavior? What is the cash value of such a claim? The notions introduced in the last section provide the basis for an answer; they provide us with a principle by which inferences to rule guidance can be licensed. The principle goes thus: if an organism is capable of producing a certain type of behavior and this behavior consists of (i,j)-dependent sequences and there is no limit on the distance between i^{th} and j^{th} events, then in producing these sequences the organism is being guided by a rule.[6] Thus if we know that an organism's ability is an ability to produce a class of behavioral sequences like the sequences in K, then by this principle we can conclude that in the production of each sequence the organism is being guided by a rule.

There are, I believe, a number of good reasons for accepting this principle and each reason depends on features built into class K, a class of (i,j)-dependent sequences with no limit on m, no limit on the distance between i^{th} and j^{th} events. The behavioral sequences in this class have the property that they cannot all be produced from left to right in a single pass. The processes that are involved in their production have to be hierarchical processes. With hierarchical processes the linear order of behavioral events may not reflect the order in which the events are actually produced. Control over production may shift forward and backward across the sequence. An initial event may determine the character of an event that occurs only much later in the sequence. In the case of our (i,j)-dependent sequences not every k^{th} event $(1 \leq k \leq n)$ in the sequence can be selected on the basis of the $k\text{-}1^{th}$ event alone. Selection of the j^{th} event is determined by the i^{th} event and $i \neq j\text{-}1$. What we require for the production of all of the sequences in K is a special pattern of control: at various times control over the subject's behavior has to shift from the preceding event to a state or structure outside of the sequence that stores the location of the i^{th} event and that selects the appropriate event for j.

Class K shares a feature with languages that include sentences of the form of (7) where there is no limit on the length of S_1. The sentences of these languages and the sequences in K cannot be produced by a Markov process, by a process that moves only from left to right and in which the occurrence of each element that the process produces is determined by the immediately preceding element or series of elements. In sentences like (7) choice of a word is determined by a word that occurs much earlier in the sentence. While in sequences like those in K choice of a behavioral event is determined by an event that occurs much earlier in the sequence. A process that produces these sentences and a process that produces the sequences in K must employ elements or structures that never appear in any of the sequences or sentences. In the case of

[6] We again assume that our subject can perform sequences like X′, that the circumstances are such, for example, that he can enter and leave by different doors.

language these elements are grammatical constructions or categories that label the nodes on the grammarian's tree. In the nonlinguistic case they have a quite similar role. Their role in both the production of sentences and in the production of the sequences in K is essentially integrative. They impose the order on the sequences of skilled movements of which Lashley spoke. They provide a "generalized schemata of action which determine the sequence of specific acts, acts which in themselves or in their associations seem to have no temporal valence."[7] Given their independence of motor events and integrative function, an apt description of these structures is "rules." That is, (1) having an integrative function and (2) being independent of the elements they integrate are the two properties that we intuitively assign to rules. A rule, I suggest, is any structure that is somehow realized in an organism or machine, that controls the order in which a sequence of operations or events is to be performed, and yet that is independent of those operations or events. A rule is guiding some bit of an organism's behavior when it is the rule that determined the sequence of operations the organism is carrying out.

Consider again a subject B who goes out to buy a newspaper and makes it a practice to depart and return by the same door. $X = e_1, e_2, \ldots, e_8$ where e_1 is B's walking out the front door, e_2 is his walking to the end of the block, e_3 is his noticing a newspaper stand, e_4 is his buying an evening newspaper, e_5 is his stopping and speaking to a passerby, e_6 is his turning, e_7 is his walking back down the block, and e_8 is his entering by the front door. X is an instance of leaving-and-entering-by-the-same-door behavior and B performs X. Moreover, *ex hypothesi*, there is an (i,j)-dependency between events e_1 and e_8. These events must "agree" with one another, the door he leaves and enters by must be the same door. If B were to choose to depart by a different door, by the back door, his choice somewhat later on, his choice of the 8^{th} event would have to be of a different door as well. Moreover, B can always lengthen the sequence by iterating walks up and down the block. He may always walk up and down another block before he reenters his home. So long as these sequences are hierarchically ordered this poses no special problem. Even in the short sequences agreement between the first and last events, between e_1 and e_n, was in no way dependent on the $e_{n-1}{}^{th}$ event.

Figure 2 reveals the cognitive relations and structure that underlie X.

[7] K. S. Lashley, "The Problem of Serial Order in Behavior," p. 122.

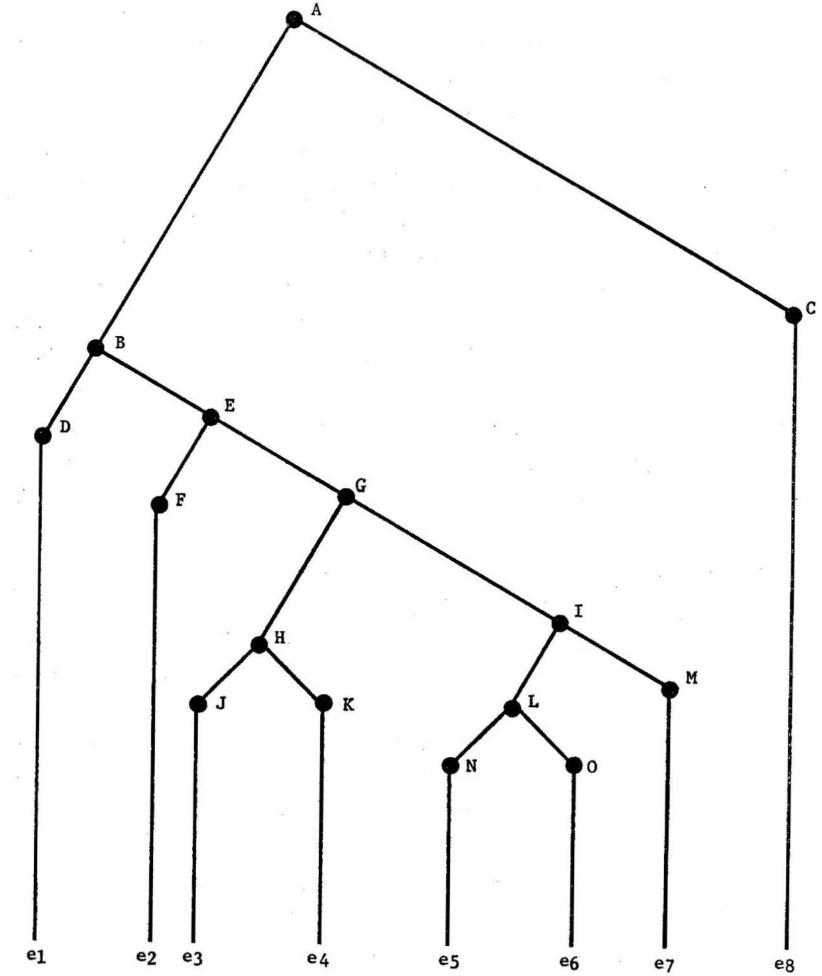


Figure 2

The entire sequence is dominated by the node marked "A." Each node, i.e., *A* to *O*, is associated with a class of behavioral events. For example, node *A* is associated with all and only the events that fit the description "leaving and entering by the same door." The lines that connect the nodes stand for relations between the sets. The lines that connect nodes *B* and *C* to *A* indicate that the set associated with *A* is the Cartesian product of *B* and *C*. All of the sequences of events that belong to *A*, all leaving-and-entering-by-the-same-door events, consist of an ordered pair

of sequences; the first member of the pair is a sequence drawn from *B*; the second member is a sequence drawn from *C*. The lines that connect the nodes to the "terminal" symbols, to e_1 to e_8, stand for the membership relation. Event e_1 is a member of the set of events associated with *D*, e.g., the set of all departures through a door. The relation between events e_1 and e_8, between our two (i,j)-dependent events, is represented by the lines that connect e_1 to node *D*, e_8 to *C*, *D* to *B*, *B* to *A*, and *C* to *A*.

Looked at from the point of view of production and control, each node plays a special role in the production of the final sequence. Node *A*, for example, constrains the sequence by requiring that it consist of sequences drawn from *B* and *C*. In addition sequences from *B* and *C* are so selected that the final sequence of events, e.g., events $e_1 \ldots, e_8$, belongs to the set associated with *A*. Looked at this way each node has the function of a rule.

My claim has been this: for an organism to have some behavioral capacities, for an organism to have the capacity to perform each sequence in a class *K* of behavioral sequences, is for him to be guided by a rule. By "rule" I mean something that (1) is independent of the individual events that make up the sequences in *K* and (2) still serves to integrate the events and to preserve various sorts of agreement between them. But I need to add a note of caution here. The term that I have been discussing is the term "rule guidance." This term is one of a family of terms that have been introduced into talk about language and into talk about complex behavior. What I have said about rule guidance suggests that the process of being guided by a rule is neither a distinctively human nor a particularly mysterious process. In fact a general purpose computer is on my analysis a paradigm case of a system whose behavior is guided by a rule. But I have said nothing about some of the more exotic members of this favored circle of terms. I have said nothing about knowledge of rules or the expression "tacit knowledge of rules," the expression that Noam Chomsky so often uses when talking about language. On my analysis the expression "guided by a rule" has broad application. It applies to the behavior of systems to which we are not inclined to attribute knowledge or beliefs. On my analysis of "guided by a rule" the following is not a valid inference: *S*'s behavior is guided by a rule; therefore *S* has knowledge (albeit tacit) of a rule. *S*'s behavior may be guided by a rule and *S* may have no knowledge at all.

On the other hand, a subject may know a rule and his behavior may not be guided by it. In fact here lies one of the paradoxes of Chomsky's distinction between language performance and language competence. Chomsky attributes knowledge of a grammar to a speaker but denies that the speaker puts the grammar to use in the perception and production of speech.[8] The point seems to be that while the speaker knows these rules (in some sense of "knows") he is not guided by them when he actually speaks and understands the sentences of his language. This of course makes the speaker's possession of the rules much like a wheel that doesn't turn anything, and it makes Chomsky's talk about knowledge of these rules quite mysterious. An investigation of this mystery, however, will have to await another day. (See my paper "Language, Rules and Psychological Explanation," *Biosciences Communications* [in press].)

IV

Chomsky believes that every speaker of a natural language has mastered and internalized a system of rules that express his ability to speak and understand his language. He also believes that any fully adequate theory of a language must assign to each of infinitely many sentences a structural description that indicates how the sentence is understood by the speakers of the language. In sections II and III, I gave reasons for believing that you and I have other abilities that are somewhat similar. We have the ability to perform sequences of complex nonverbal behavior, and we have seen reasons for saying that this behavior is rule guided. Thus in mastering these sequences we master and internalize a system of rules. These rules subsequently guide our behavior and describe or express our ability to perform each of infinitely many sequences of complex nonverbal behavior. Moreover, a fully adequate theory of this behavior must assign to each of the sequences a structural description that indicates the significance of the sequence for the subject.

Among the claims that Chomsky makes about our ability to speak and understand a natural language is that it is species-specific and consists of a unique type of intellectual organization that manifests itself in the "creative" aspect of ordinary language use: "man has a species-specific capacity, a unique type of intellectual organization which can-

[8] Noam Chomsky, *Aspects of the Theory of Syntax*, p. 9.

not be attributed to peripheral organs or related to general intelligence and which manifests itself in what we may refer to as the 'creative aspect' of ordinary language use — its property being both unbounded in scope and stimulus-free."[9] According to Chomsky the properties of being unbounded in scope and being stimulus-free are independent. A system may have either one of these properties and not the other. A tape recorder has the ability to record speech. In principle there is no limit on the number of sentences it can record. Given another reel of tape and neglecting the mortality of the machine, the tape recorder can always record one more sentence. Since there is no last sentence that the machine can record, we can say that the recorder's ability to record speech is unbounded in scope. However, the recorder's ability to record speech is not stimulus-free. There is a 1–1 mapping from what we say to what it records.

On the other hand, a machine that has only two responses that are produced randomly is a machine whose responses are stimulus-free but whose ability to respond is obviously not unbounded. In short, devices that are capable of producing infinitely many responses manifest unboundedness. Devices whose responses are not determined by any stimulus manifest stimulus-freedom.

One way of reading Chomsky on the issue of "creativity" might be as follows: while many abilities may manifest either one or the other of these properties, namely, unboundedness and freedom from stimuli, only the ability to speak and understand a natural language manifests both: "Animal behavior is typically regarded by the Cartesians as unbounded, but not stimulus-free, and hence not 'creative' in the sense of human speech."[10] Despite the textual evidence, this is not the way to read Chomsky. The ability to use a natural language is obviously not the only ability that manifests both unboundedness and stimulus-freedom. There are machines with simple nonlinguistic abilities that manifest both of these properties. A machine that randomly prints out the integers manifests both. The machine can print any integer, but which integer it prints at any moment is selected randomly. That is, each integer has an equal probability of being selected and printed at each moment in time as any other. If the only properties that mark "creative"

[9] Noam Chomsky, *Cartesian Linguistics* (New York: Harper and Row, 1966), pp. 4–5.
[10] *Ibid.*, p. 77.

use are stimulus-freedom and unboundedness, a machine that randomly prints the integers is creative.

However, these are not the only properties that mark the "creative" aspect of ordinary language use. As Chomsky points out, language not only provides a means for expressing indefinitely many thoughts, i.e., speaking and understanding an unbounded number of sentences, but also a means for reacting appropriately in an indefinite range of new situations. Our linguistic responses are not only unbounded and stimulus-free, but are also appropriate to the contexts in which we make them. While the contexts do not determine the responses, the responses are appropriate to the contexts: "The essential difference between man and animal is exhibited most clearly by human language, in particular, by man's ability to form new statements which express new thoughts and which are appropriate to new situations."[11] The sentences (responses) a fluent English speaker produces are not determined by the context in which they are uttered. The speaker could have always produced a different sentence in the same context. Nevertheless the sentences are all appropriate to the context.

Consider, again, the machine that randomly generates the integers. The machine can print infinitely many integers but, since each integer has an equal probability of being selected and printed at each moment, the machine's responses cannot properly be said to be appropriate to the contexts in which they are made. While the machine can produce an unlimited number of responses and the English speaker can produce an infinite number of sentences, the speaker's sentences are appropriate to the contexts in which they are produced and the machine's are not. Our ability to speak English involves more than unboundedness and stimulus-freedom, and this something more, namely, appropriateness to the context of use, is apparently not a property shared by our machine.

But perhaps we are selling our machine a bit short. There are contexts in which printing out any integer would be quite appropriate. I could ask the following questions: "What is an example of an integer?" or "What is an example of a whole number?" or "Can you give me an example of a number?" In these contexts printing out any integer would be to produce an appropriate response. That is to say, there are some contexts in which the machine is able to produce infinitely many responses each of which is appropriate to, but undetermined by, the context. On the other hand, there are some contexts in which an English speaker is not able to produce more than a small finite number of appropriate responses. If I ask you to tell me in six words or less why you read and study philosophy, to reply with a sentence of ten words would, presumably, not be to produce an appropriate response. But if the distinction between man's language abilities and the machine's abilities is not that the one's responses are always appropriate and the other's never appropriate, how is the distinction to be made?

[11] *Ibid.*, p. 3.

If this objection is at all appealing, it is only because of some unclarity over the notion of one's responses being appropriate to the context of use. We can provide the machine that prints the numbers with some contexts in which each of its responses may appear appropriate. But we provide the machine with the contexts. We bring about the appropriateness. We select the context to fit the response. If we change the context, there will be no appropriate change in the machine's response. If we ask "What is an example of an integer between 4 and 6?" we can expect the same response as if we were to ask "Can you give me an example of an integer?" If the machine's responses are appropriate to the context, it is presumably because of something we do. If the machine's responses fit, it is due to our abilities and our sapience rather than the machine's. The case is quite different with language. A speaker of English is able to produce sentences that are appropriate to novel contexts. Moreover, if the response is appropriate to the context, it is because of something the speaker is able to do. The speaker brings it about that his response fits the context. Part of what the speaker of English is able to do when he is able to speak and understand English is to choose sentences that are appropriate to the context of use. On the other hand, this is not part of what a machine is able to do when it is able to randomly print the integers.

While I think this reply is apt, more needs to be said about a sentence's being appropriate to the context of use. As G. P. Grice has pointed out,[12] the use of a linguistic expression may be inappropriate for any number of reasons. An expression may be inappropriate because it fails to correspond to the world in some favored way, because it is pointless, or because it violates some general principles governing communica-

[12] G. P. Grice, "Logic and Conversation." William James Lectures, Harvard University, 1967–1968. Unpublished.

tion or rational behavior. Chomsky's point that an essential property of language is that it provides the means for producing novel but appropriate sentences and for producing sentences that are appropriate to novel contexts suffers, because he says so little about the conditions under which sentences can be said to be appropriate and inappropriate to the contexts of their use.

Grice's work is helpful here. Most linguistic behavior is purposive and Grice has formulated a rough general principle that speakers of a language expect one another to observe and that sets some conditions on what counts as appropriate: make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk-exchange in which you are engaged. One purpose that underlies the use of some linguistic expressions is to produce an effective exchange of information. Given this purpose, users of these expressions share some expectations and presumptions concerning whether their use is appropriate in various contexts. If a speaker uses a sentence in a context that does not satisfy the purpose of the talk-exchange, that sentence is inappropriate.

It is clear from this that the responses of the number generator (the machine that randomly prints out the integers) are neither inappropriate nor appropriate to the contexts in which they are produced. The behavior of the machine is not purposive. Accordingly, it does not have a set of expectations or presumptions concerning how it ought to respond in order to achieve its ends and its purposes, and, accordingly, it cannot meet or fail to meet some shared expectations. The ability of the machine, then, is quite different from the abilities of language speakers and the machine does not manifest all the properties that Chomsky identifies with the creative aspect of language use. Where I take issue with Chomsky is over whether other human intellectual abilities have the properties that he associates with the creative aspects of ordinary language use. I believe that man has many nonlinguistic abilities that are "creative" in the sense that linguistic processes and abilities are.

The rules that express a speaker's ability to speak and understand his language can iterate and thereby generate an indefinitely large number of linguistic structures. In sections II and III, I argued that there are rules that express man's ability to perform nonlinguistic behaviors. We saw that these rules can similarly iterate to generate an indefinitely large number of structures. So man has the ability to perform an unbounded

number of nonlinguistic behaviors. But we also saw that these behaviors are determined not by stimuli or by previous responses but by a generalized schemata of action, i.e., rules. These rules are independent of the acts that they determine but serve to determine the order in which these acts are carried out. It remains to show that these abilities can include the ability to perform acts that are appropriate to the contexts in which they are performed and which are appropriate to new and novel situations.

Grice believes that language use, talking, is just a special case of purposive or rational behavior. Further he believes that some of the purposes that underlie the use of language underlie the use of nonlinguistic behaviors as well. One end for which language is used is for the exchange of information. For example, we can demonstrate our displeasure with another's conduct by saying, "What you have done displeases me." But there are also exchanges that don't involve talk that accomplish this same end. Participants in these exchanges share common purposes. Like those who participate in talk-exchanges, they also share some expectations and presumptions concerning whether some bit of behavior is appropriate to the context in which it is performed. In this way the terms "appropriate" and "inappropriate" can be applied to sequences of nonverbal behavior. If you are assisting me to fix my automobile and I need a wrench, I don't expect you to hand me a feather duster. If you share with me the end that I get the wrench and, nevertheless, hand me the duster, then, on its face, your behavior is inappropriate.

Earlier I argued that there are classes of nonlinguistic behaviors which each of us is able to perform and which require rules for their production. We are capable of performing any one of indefinitely many sequences of behavior each of which require some agreement between nonadjacent events, and there is no limit on how far removed these events may be from one another. All of these behaviors are purposive in at least one sense: changes in the j^{th} event in a sequence are for the purpose of maintaining agreement with th i^{th} event. The subject makes it a practice or girds himself to enter the same door from which he leaves. Under this description, his entering by the back door is purposive. The subject enters by the back door in order to bring it about that he enters and leaves by the same door.

Nonlinguistic behavior, then, like linguistic behavior can be purposive. Both can be purposive because we can perform either kind of behavior in order to bring about some desired result. We knit a sweater in order to please a friend. We solve a mathematics problem in order to determine the cost of a new driveway. In addition there are some rough general conditions that persons who engage in purposive behavior expect each other to observe: make your performance such as is required, at the stage at which it occurs, by the accepted purpose or direction of the activity in which you are engaged. And a performance or some bit of behavior is appropriate to a context just in case it is what is required in the context by whatever is the accepted purpose or direction of the activity.

Someone who makes it a practice to enter and leave by the same door, who coordinates his entering and leaving regardless of how removed these events may be, has an ability that can only be expressed by a rule. However, given this practice, a subject's entering by the back door can be said to be appropriate to the context in which it occurs just in case he first leaves by that door. If the subject makes it a practice to enter and leave his home by the same door, it would be inappropriate for him to leave by the front door and enter by the back door. If he resolves or makes it a practice to enter and leave his home by the same door but not before walking around the block four times, it would be inappropriate for him to enter after having walked around the block twice. A subject who is able to coordinate his arrivals and departures around an unlimited number of trips around the block and who is able to carry out this coordination in order to satisfy some resolve, has the means for fitting his behavior to the context of its use. However, not only is the subject able to make his performances appropriate to their contexts, he is able to form new behavioral sequences that are appropriate to new and novel situations. He has the means for performing an unlimited number of behaviors and for reacting appropriately in an unlimited number of new or novel situations. In our present example, for the subject to act appropriately is for him to enter and leave by the same door. But a subject who can coordinate his enterings and leavings around an unlimited number of other behavioral elements can react appropriately in new or novel contexts. Specifically, he can react appropriately in a situation in which, in order to enter and leave by the same door, he must walk n times around the block for any integer value of n. In other words, walks around the block interrupt his leaving and entering. Since there are infinitely many such situations, i.e., one situation for each

value of n and infinitely many values for n, the subject can act appropriately in an unlimited number of new or novel situations.

In short, the ability to coordinate one's leavings and enterings has the following properties: it is unbounded in scope, i.e., there is no longest sequence of entering-and-leaving-by-the-same-door behavior that a subject who has this ability is able to perform; the performances that result when the subject exercises this ability are stimulus-free, i.e., the sequences of behaviors that he performs are mediated by rules; and finally the ability to coordinate one's leavings and enterings includes the ability to make each of the sequences appropriate to the contexts in which they are performed even when the contexts are new or novel, i.e., a person who is able to coordinate his leavings and enterings can enter and leave by the same door when to do so requires the performance of a sequence of behavior unlike those which he has previously performed. If I am right, then there are nonlinguistic abilities and processes that are "creative" in the same sense that, according to Noam Chomsky, linguistic processes are. The type of intellectual organization that manifests itself in language use and in the ability to speak and understand a natural language is not unique. This same type of intellectual organization manifests itself in nonlinguistic performances and in the ability to perform many types of nonlinguistic behavior of which the ability to coordinate one's leavings and enterings is but a simple example.

V

I have been arguing that some of the notions and results that have arisen from the study of natural languages can be generalized and the generalizations applied to nonlinguistic behavior. In sections II and III, I pointed out that a good deal more behavior than verbal behavior is hierarchically ordered and that this order is significant for both the organism and the specification of his behavior. In section IV, I argued that if there is a creative aspect of ordinary language use, there is a creative aspect in the performance of some nonlinguistic behavior as well. However, the analogies between verbal and nonverbal abilities can be pushed too far, and I wish to turn my final comments to some of the differences.

The system of rules that express a speaker's abilities to speak and understand his language can be analyzed (at least on Chomsky's view) into three major components: syntactic, phonological, and semantic rules. The syntactic component, in turn, includes a base subcomponent and a transformational subcomponent. The former consists of rules that generate a set of basic strings each with an associated structural description. The latter consists of rules that map strings and their associated structural descriptions onto strings with associated structural descriptions. There is, to my knowledge, nothing corresponding to each of these components in the nonlinguistic case. Though there is some evidence that the notion of grammatical transformation can be generalized and applied to phenomena in visual perception, there is presently no good reason to believe that for each of the phenomena the linguist describes or for each element in his description there corresponds a similar phenomena or element in some nonlinguistic dimension.

However, I think I can make a more interesting point. Where there are (i,j)-dependencies between elements in linguistic strings, in strings of linguistic formatives, there are also always selectional restrictions or restrictions of co-occurrence that constrain the choice of elements that can be nested within i^{th} and j^{th} elements. A speaker of English can speak and understand the sentence

(9) Blonds are abundant in Sweden

and infinitely many other sentences in which the subject of the sentence is "blonds" and "are abundant in Sweden" its predicate. In each sentence there will be an (i,j)-dependency between "blonds" and "are abundant." There is no longest such sentence because we can iterate relative clauses between the i^{th} and j^{th} elements. For example,

(10) Blonds who love children are abundant in Sweden,

(11) Blonds who love children who love dogs are abundant in Sweden,

but not

(12) *Blonds which boil at 212 degrees Fahrenheit are abundant in Sweden.

What the examples show us is that there is no limit on the number of relative clauses that can be nested between subject and predicate, but the relative clauses cannot be freely selected. Within the language there are restrictions on what clauses can modify what nouns. Consequently, there are restrictions on what we can iterate between i^{th} and j^{th} elements. The clause "who love children" is permitted, but the clause "which boils at 212 degrees Fahrenheit" is not.

On the other hand, in at least some of the nonlinguistic cases I considered in sections II and III, there were no such restrictions. That is, there were no restrictions on the behavioral elements that we could nest between i^{th} and j^{th} behavioral events. Where there are (i,j)-dependencies between elements in nonlinguistic strings, e.g., entering and leaving by the same door, there need be nothing that corresponds to selectional or co-occurrence restrictions on the embedded elements. One can walk up and down the block, write a book about knitting, brush his teeth, prepare dinner, or sail to Calais. In this nonlinguistic case one's choice is unrestricted by the rules that express the ability to perform the sequences in K, e.g., to enter and leave by the same door. That is, the intervening behavior may have nothing to do with the i^{th} and j^{th} events.

In short, linguists have erred in overestimating the distinctiveness of language. Walking and talking are more alike than the Cartesians think. On the other hand, it is easy to err on the side of parity. It would be a mistake to allow the similarities between sweater-making and English-speaking to mask the interesting differences. Restrictions on occurrence may be one difference. I leave it to the linguist to remind us of the rest.