Memory & Cognition
1979, 7 (5), 390-400

Processing of text containing artificial inclusion relations

BARBEE T. MYNATT and KIRK H. SMITH
Bowling Green State University, Bowling Green, Ohio 43403

The present studies were based on the hypothesis that the majority of college students have available to them the appropriate schema for understanding set inclusion relations, but that various factors influence the likelihood that the schema is used in the processing of text containing artificial inclusion relations. Although group data did not support this hypothesis, the data of individual subjects could be readily interpreted as resulting from the selection of one of a small set of representational schemata. Among the factors shown to influence schema selection were the choice of sentence frame used to present each relation, the presence or absence of real-world contextual information, and the structure (simple vs. complex) of the underlying inclusion relation. In addition, one experiment showed that the processes used in constructing a mental representation of an inclusion relation and in retrieving information from the representation are similar to those used with linear orderings.

Recent studies of text processing have drawn attention to both the necessary role of inference in comprehension and the pervasive effect that inferences have on subsequent memory for texts (Bransford, Barclay, & Franks, 1972; Collins & Quillian, 1972; Harris & Monaco, 1978). In contrast, certain studies of set inclusion relations in meaningful text (Carroll & Kammann, 1977; Frase, 1969; Griggs, 1976; Potts, 1976) have focused on a situation in which subjects do not store in memory what seem to be obvious inferences and apparently even fail to draw these inferences. The purpose of the present paper is to demonstrate the extent to which earlier results with set inclusions have been overgeneralized. We suggest that the earlier conclusions were due in part to the text materials that were chosen and the logical interpretations they were given by the investigators.

The original studies of set inclusion texts by Frase (1969, 1970) relied on a pivotal assumption about the nature of text derived from two papers by Dawes (1964, 1966). Dawes argued that all declarative sentences assert set inclusion relations. In other words, the sentence, "The farmers are peaceloving," can be interpreted to mean that the class of entities (people, in this case) who are farmers is a proper subset of the class of entities who are peaceloving. Thus, studies of inferences with texts containing set inclusion relations could be viewed as representative of the processes operating in descriptive text of all kinds.

This work was supported by National Science Foundation Grant BNS-77-16860 to both authors. Experiments 2, 3, and 4 were presented at the annual meeting of the Psychonomic Society, San Antonio, Texas, November 1978. Appreciation is expressed to Ken Tohinaka for his patient discussions of logic. Requests for reprints should be sent to Barbee T. Mynatt, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43403.

Frase's (1969) texts are in fact rather unusual. An example of a paragraph from Frase (1969) that has been used extensively in subsequent research is given in Table 1. The paragraph contains four sentences of the form, "The As are Bs," for example, "The hill people of Central Ugala are farmers." If these sentences are interpreted as assertions of class inclusion, then the paragraph describes a nested set of five classes (call them A, B, C, D, and E, in the order they are mentioned in the paragraph). Six additional class inclusions can be inferred from the four presented in the paragraph. These inferences have been referred to as "remote" relationships to distinguish them from the "adjacent" relationships presented in the paragraph. The latter terms imply a notion of rank or distance in the set hierarchy. Distance can be measured on an ordinal scale in terms of step size, which we will define as the number of intervening classes required by an inference. Thus, the presented sentences describe adjacent relations and represent Step Size 0; remote relations such as "As are Cs" represent Step Size 1, "As are Ds," Step Size 2, and "As are Es," Step Size 3.

The finding that has provoked the greatest interest with respect to set inclusion paragraphs like those used by Frase (1969) is the unwillingness or even inability of college students to draw inferences from the material along the lines described above. Frase manipulated the instructions given before the paragraphs were presented. He told subjects in some instances to determine whether certain inferences could be drawn from the material (e.g., "Are Bs Es?") and then tested to see whether intermediate inferences (e.g., "Cs are Es") were recalled or recognized. He found that, in general, college students remembered only what had been presented in the paragraphs and the specific inferences they were asked to check.

Subsequently, Griggs (1976) directly tested whether

  0090-502X/79/050390-11$01.35/0

Table 1
Example of an Artificial Set Inclusion Paragraph Used by Frase (1969)

The Fundalas are outcasts from other tribes in Central Ugala. It is the custom in this country to get rid of certain types of people. The outcasts of Central Ugala are all hill people. There are about fifteen different tribes in this area. The hill people of Central Ugala are farmers. The upper highlands provide excellent soil for farming. The farmers of this country are peaceloving people, which is reflected in their art work.

subjects could verify the truth of the inferred relationships. After reading paragraphs similar to those of Frase (1969) (except that the assertions took the form "All As are Bs"), the subjects responded "true" or "false" to a set of test sentences representing all the various combinations of elements and inclusions. Thus, there were true adjacent relations that had appeared in the paragraph and true remote relations that the presented sentences implied, as well as false adjacent and remote relations. Griggs found that the proportion of correct responses was generally greater for true relations. In addition, there was an interaction of truth value and distance (adjacent vs. remote). For true relations, proportion correct was typically greater for adjacent relations than for remote relations. For false relations, the opposite result was found. Similar findings have been reported by Carroll and Kammann (1977) and Potts (1976).

One reason for the interest in this failure of inference is that investigations of similar paragraphs asserting a logically similar set of relational propositions have found that subjects' performance tends to improve with step size. For example, Potts (1972) reported that when paragraphs composed of four assertions of the form "The A is smarter than the B" were tested in the same way, subjects correctly verified the remote comparisons and verification time decreased with step size. Moreover, Griggs (1976) and Potts (1976) have reported experiments that directly compared the two types of relations and found that the proportion of correct verifications for comparisons (linear orderings) was the same for true and false relations and increased with step size. Because both linear orderings and set inclusions are built from transitive relations that are not symmetric, parallel results might be expected. Indeed, some authors have implied that the two relations are cognitively equivalent (cf. Hayes-Roth & Hayes-Roth, 1975). The discrepancy in the results using the two types of materials has stimulated a search for an explanation.

Griggs (1976) first established that the problems subjects have with paragraphs like the one in Table 1 are not due to memory. When subjects had the paragraphs in front of them while answering questions, they still produced the same pattern of errors as when they read the paragraphs and answered from memory (see also Carroll & Kammann, 1977). Griggs reasoned that subjects must have difficulty interpreting the meaning of set inclusion statements. First, he noted that subjects appeared to treat the stated relations as symmetrical. From the sentence, "All As are Bs," they concluded that it was also true that Bs are As, thus assuming identity of the two sets. This interpretation, sometimes referred to as "illicit conversion," is well documented in research on syllogistic reasoning (see, e.g., Ceraso & Provitera, 1971; Chapman & Chapman, 1959; Wason & Johnson-Laird, 1972). Second, Griggs noted subjects' apparent unwillingness to treat set inclusion as a transitive relation. If all A are B and all B are C, then all A must be C. However, Griggs suggested that subjects might not be applying this reasoning in comprehending the set inclusion paragraphs.

In two experiments Griggs (1976) more or less subtly taught subjects the intended interpretation of the set inclusion sentences. In the least subtle manipulation, he explicitly instructed subjects on the validity of transitive inferences and the invalidity of symmetric inferences using a concrete example about brown African elephants with long trunks and large tusks. For subjects who received this special instruction, Griggs found no difference in the proportion of correct responses for true and false relations, with performance improving as step size increased. Potts (1976) has shown that subjects' logical errors can be overcome in a slightly different way. In one experiment Potts utilized a practice paragraph that described a real set inclusion and included the sentence "All collies are dogs." Feedback was given following a block of practice questions. The experimental paragraphs, however, substituted nonsense syllables for the elements of the set (e.g., "All DAX are MEP"). Feedback also followed the blocks of questions. Under these conditions subjects were correct a high proportion of the time and their response times showed a decrease with step size that resembled subjects in the same experiment who were given sentences containing comparisons (linear orderings).

The studies we have reviewed suggest that Frase's (1969) set inclusion paragraphs were an unfortunate choice for the study of the role of inference in text processing. This is not to say that the logic underlying the construction of these materials is faulty. Any set of elements can be assigned to a class on the basis of a common attribute. But the paragraph about the Fundalas (Table 1) can be understood in two ways. Careful analysis reveals a hierarchy of nested sets as intended by Frase, but superficially the passage reads like a list of attributes of the Fundalas rather than an anthropological taxonomy. However, much of the interest in experiments using these paragraphs is prompted by the fact that a set inclusion interpretation of the paragraphs makes this research appear relevant to the semantic memory literature, where the semantics of English imply natural taxonomies—collies, dogs, animals; thrones, chairs, furniture; and so on. Frase's experiments would have something to do with how people use texts to add to their semantic memories if people understood simple declarative sentences to be assertions of set inclusion (cf. Dawes, 1964). But the weight of the evidence suggests they are not very likely to adopt this interpretation in the absence of special instruction.

One explanation of why in the absence of instruction college students misinterpret artificial set inclusion relations in text is that people in general do not understand the logician's interpretation of "All A are B." Support for this conclusion could be drawn from work on how people reason with syllogisms and interpret set inclusion statements in ordinary usage (e.g., Wason & Johnson-Laird, 1972). However, this explanation fails to take account of the fact that instructed subjects can and do process such statements according to canons of logic. It is noteworthy that Griggs (1976) and Potts (1976) used instructions that drew on subjects' prior knowledge. This fact suggests that the proper interpretation of set inclusion statements is available to most college students but is not tapped by the typical experiment on text containing artificial set inclusions. Perhaps a critical aspect of text processing has thus far been overlooked.

When a subject first encounters a text, the information in the individual sentences does not interact with a neutral background of previously stored information. Instead, the subject has at his disposal a large number of schemata in which the new information can be interpreted and stored. We will use the term "schema" here, in preference to other currently popular terms such as "frame" or "script," because of the weight of historical precedent (Bartlett, 1932) and because we feel the term has a greater generality that the concept deserves. As a text is being read, the subject selects one of the available schemata, and thereafter, the information in the text is encoded in a distinctive way reflecting the specific schema selected. The inferences a subject can and does draw from a text depend critically on the choice of schema.

The notion of schema selection suggests that there should be settings in which uninstructed subjects would process texts containing set inclusion relations in a logically correct fashion. Although, strictly speaking, the foregoing situation is not required by a schema selection hypothesis, finding such texts would constitute strong support. A pilot study suggested that appropriate texts could be constructed from statements referring to the geographical or political subdivisions of a fictional world (e.g., "NAV is in REL").[1] A series of such statements describe what may be called topological inclusion, but, in principle, the latter can be reduced to a set-theoretic interpretation. Specifically, the foregoing example means that all the points in NAV are included in the set of points corresponding to REL. A group of 22 subjects given a paper-and-pencil set inclusion task utilizing such materials performed at a high level with no special instructions (mean proportion of questions correctly answered was .937, with a range from .79 to 1.00).

The specific goal of the present study was to explore some of the additional conditions that determine whether or not college students select a set inclusion schema to represent an inclusion hierarchy presented in a paragraph. In Experiment 1, a direct comparison was made between the topological inclusion materials we developed in our pilot studies and the type of materials used in the majority of earlier studies. The procedure required the subjects to respond to the inclusion questions from memory. Experiment 2 looked at the effect of providing additional context information and varying the order in which the relations were presented. Here, the subjects were not asked to work from memory. Experiment 3, using measures of study response time, demonstrated that the processes used to construct a mental representation of an inclusion relation and in retrieving information from the representation are much the same as those in the processing of comparison relations (linear orderings). The final experiment examined performance on a topological inclusion that involved multiple subsets at several levels of an inclusion hierarchy.

EXPERIMENT 1

Method

Subjects were given booklets containing all instructions and materials, allowing them to proceed through the task at their own pace. They were instructed to read and study a short paragraph, then to turn over the page and answer a set of questions concerning the paragraph without referring to the previous page. Each subject read and answered questions concerning two paragraphs in succession. In the paragraphs, all subjects were asked to pretend that a civilization of Plutons had just been discovered on the back side of the planet Pluto. For half of the subjects, the rest of the paragraph explained that the Plutons divided their planet into geographical and political units, and it presented four sentences describing the relationship among five units on one part of the planet in sentences of the form "A is in B." For the remaining subjects, the paragraph explained that the Plutons had developed a complex tribal system and it presented four sentences describing the organization of five tribes in one country in sentences of the form, "All A are B." The terms used as units or tribes were nonsense syllables taken from Noble (1961), ranging in meaningfulness (m') from 3.20 to 3.33. The sentences describing inclusion relations were presented in the order AB, BC, CD, DE, where "AB" stands for either "A is in B" or "All A are B," as appropriate. Frase (1970) has shown that this order is most likely to encourage subjects to make correct inferences in recall. Two different sets of sentences, each describing an inclusion relation, were constructed for each condition. To counterbalance materials half the subjects within each condition received one set of sentences first, and half received the other set first.

After studying each paragraph as long as desired, the subjects turned the page, read 20 statements, and made a check indicating whether they thought each statement was true or false based on the paragraph they had just seen. They were told to consider a statement true if the information in the statement either was presented in the paragraph or could be logically deduced from the information in the paragraph. A statement was to be considered false if the information it expressed was not in the paragraph or could not be logically deduced from the paragraph. The 20 statements included all possible combinations of the five terms in both forward (true) order and reverse (false) order. Thus, half of the statements were worded so that a "true" answer would be correct, and half so that a "false" answer would be correct. Within both the true and false statements, four involved adjacent relations (Step Size 0), three involved relations of Step Size 1, two involved relations of Step Size 2, and one involved a relation of Step Size 3. The order of the statements was random, with all subjects receiving the same order.

The subjects were 52 undergraduates who were required to participate in psychological research as a part of a course in introductory psychology.

Table 2
Mean Proportion of Correct Responses to Test Questions for Subjects in the Two Conditions of Experiment 1

Type of Question	Condition	
	All A are B	A is in B
True Adjacent	.92	.97
True Remote	.66	.79
False Adjacent	.52	.89
False Remote	.65	.93

Results and Discussion

The proportion of statements correctly verified by each group of subjects for the four types of questions is shown in Table 2. "Correct" answers were defined by formal logic, which treats "All A are B" as a transitive, asymmetric relation between A and B. The correct response to forward adjacent and remote sentences was "true," and to reverse adjacent and remote sentences, "false." However, the label "correct" should be seen as no more than a useful way of classifying a response according to a particular criterion. It is not intended as an indication of subjects' inability to understand English or to reason. For example, English usage permits a sentence such as "All deities are gods" to be interpreted as set identity (a symmetric relation). If a subject chooses this interpretation, then reverse adjacent and remote sentences are "true," in spite of the fact that they would be scored "incorrect" here. (See Revlis, 1975, for a discussion of the separate roles of encoding and deductive processes in reasoning with quantifiers.)

An analysis of variance was conducted with proportions of correct responses using the .05 level of significance. (All subsequent statistical analyses used the .05 level.) For both groups combined, proportion correct was greater for true relations than for false [$F(1,25) = 315.66$, $MSe = .523$]. Within the true relations, proportion correct was greater for adjacent relations than for remote relations [$F(1,25) = 16.54$, $MSe = .526$]. For false relations, the opposite effect was found: Proportion correct was greater for remote relations [$F(1,25) = 11.63$, $MSe = .214$]. Although neither of these distance effects interacts with group, performance overall is at a significantly higher level for the A-is-in-B group [$F(1,25) = 4.29$, $MSe = 1.14$]. In addition, the difference between proportion correct on true and false questions is considerably less (and in the opposite direction) for the A-is-in-B group, as shown by a significant interaction of group and truth value [$F(1,25) = 11.06$, $MSe = 1.88$].

A closer look at the data, however, indicates that it is very misleading to theorize about how college students process inclusion relations on the basis of group data. Although Potts (1976) did present an analysis of consistent responses by subjects for each type of question, he did not look at subjects' individual response patterns. An examination of individual subjects' responses in the present study reveals that they produce identifiable patterns of responses, reflecting the use of different processing strategies. In Table 3 the data are presented as a function of experimental group, subjects' strategies, and type of question. A subject was assumed to be using a strategy if the same response, true or false, was given to each type of question 67% of the time. A total of 16 different strategies is possible based on patterns of consistent responses to the four types of questions. However, if it is assumed that every subject who follows

Table 3
Proportion of Correct Responses to the Four Types of Questions for Subjects Assuming Different Strategies in Experiment 1

Subject Strategy	Number of Subjects	Type of Question: True Adjacent	True Remote	False Adjacent	False Remote
		All-A-are-B Condition			
Transitivity and Asymmetry	7	.93	.94	.96	.99
Transitivity and Symmetry	6	1.00	.96	.02	.06
Intransitivity and Asymmetry	4	.88	.10	.91	.94
Intransitivity and Symmetry	2	.88	.04	.12	.96
Transitivity or Symmetry	1	1.00	1.00	.12	.92
Inconsistent	6	.85	.54	.46	.77
		A-is-in-B Condition			
Transitivity and Asymmetry	21	.96	.96	.90	.97
Transitivity and Symmetry	1	1.00	1.00	.00	.00
Intransitivity and Symmetry	3	1.00	.00	1.00	.00
Inconsistent	1	1.00	1.00	.38	1.00

the instructions will respond correctly to true adjacent (presented) relations, the number of possible strategies is reduced to eight. Only five of these strategies appeared among the subjects in the two groups.

A subject who responded correctly to all four types of questions was assumed to be using a transitivity-and-asymmetry strategy. That is, such a subject presumably interpreted the relations described in the text as transitive and asymmetric in nature. A subject who answered every question with a "true" response was assumed to be using a transitivity-and-symmetry strategy. Subjects using this strategy responded as though the relationship among all the elements in the inclusion was one of identity. A subject who made no inferences at all, responding "true" to presented sentences only, was assumed to be using an intransitivity-and-asymmetry strategy. A fourth strategy, intransitivity and symmetry, involved responding "true" to both true and false adjacent relations, and false to any remote relations. A fifth strategy was used by only one subject and differs in a basic way from the other four strategies. This subject consistently drew both transitive and symmetrical inferences from presented relations, but when both logical operations were required together (false remotes), consistently responded "false." This strategy has been labeled "Transitivity or Symmetry" in Table 3, to emphasize the interacting use of the two operations.

A comparison of the number of subjects employing the various strategies in the two groups reveals the source of the data presented in Table 2. Of the 26 subjects in the A-is-in-B group, 21 correctly assumed transitivity and asymmetry, 4 used incorrect strategies, and 1 subject could not be classified. In the all-A-are-B group, only 7 of the 26 subjects used the correct strategy, 13 used an incorrect strategy, and 6 subjects responded inconsistently. The difference between the number of subjects in each condition who used the correct strategy is statistically significant using Fisher's exact test ($p < .01$).

The fact that there were six subjects in the all-A-are-B group who responded inconsistently, and only one such subject in the A-is-in-B group, is a further indication of the confusing nature of text utilizing all-A-are-B relations. However, the failure to discover a consistent strategy for so many subjects may also be due in part to the use of a memory procedure. Inconsistent subjects judged as true only 87% of the true adjacent relations, while subjects with consistent strategies were correct on 94% of the true adjacent questions. A second source of inconsistency could have been that subjects did not adopt a consistent strategy during the first trial. However, using data from only the second trial, there is no change in the distribution of strategy types among subjects in the A-is-in-B group, while only two additional subjects in the all-A-are-B group can be categorized. One of these subjects assumed a transitivity-and-symmetry strategy, and the second, an intransitivity-and-asymmetry strategy on the second trial.

EXPERIMENT 2

The results of Experiment 1 make it evident that the use of topological inclusion and the sentence frame "A is in B" evokes logically correct responding from many more subjects than does the all-A-are-B sentence frame. Such good performance is in marked contrast to earlier research using the Frase-type paragraphs and points out the great influence that the sentence frame can have on the selection of a schema by subjects. However, 19% of the subjects who received set inclusion relations in the A-is-in-B form failed to respond in a logically correct way. The question is whether some additional manipulation might further improve performance.

Both Griggs (1976) and Potts (1976) were successful with instructions that provided a link or parallel to the subjects' knowledge of real classifications. One purpose of the second experiment was to explore the role of these connections to the real world in the absence of training in logic. Subjects were given texts similar to that used in the A-is-in-B condition of Experiment 1. However, half of the subjects were told beforehand that certain elements were specific geographical-political units: cities, counties, states, countries, or continents. It was expected that this manipulation would help to make relationships among the elements clearer and aid the subject in choosing the appropriate schema for understanding the task.

The results of Experiment 1 also did not make it clear whether failures to respond correctly were due to subjects' inability to reason, to failure to select the correct schema, or to memory failure. Thus, in Experiment 2 subjects were given access to all materials throughout the experimental session in an attempt to eliminate the confounding of having the subject respond from memory.

Method

Subjects were given test booklets consisting of a page of instructions, a page containing four sets of four sentences describing the set inclusion relations, and a page of questions with spaces to check off the answers. The instructions were similar to those of the A-is-in-B condition of Experiment 1, except that subjects were told that they could refer to previous pages in the booklet and that they could take notes. Half of the subjects, those in the "categories" condition, were also told that four of the units were continents, four were countries, four were states, four were counties, and four were cities. Lists of the units fitting into each category were given in sentence form. The remaining subjects were simply given a list, in sentence form, of the 20 units. The latter subjects constituted the "no-categories" condition.

For all subjects, the second page of the booklet contained four sets of four sentences, each set describing one inclusion relation, or nest. Every sentence was of the form, "A is in B," where A and B were nonsense syllables from the same source as Experiment 1. Each of the four nests was presented in a different presentation order. Although there are 24 possible presentation orders for sets of four inclusion relations, four representative orders were used. These were AB, BC, CD, DE; DE, CD, BC, AB; BC, DE, CD, AB; and DE, BC, CD, AB, in which each letter pair stands for a sentence presenting an adjacent pair of elements. Varying orders of presentation were used because a pilot study using the same materials and procedures but only the presentation order of Experiment 1 (AB, BC, CD, DE) had shown that real-world referents had no effect. Presentation orders and their effects are further discussed in Experiment 3.

Using Graeco-Latin squares, groups of four subjects received materials counterbalanced for order of presentation, order of the nests on the printed pages, and assignment of nest terms (nonsense syllables) to presentation orders. Three different Graeco-Latin squares were used to generate set inclusion paragraphs for 12 subjects. The same paragraphs were given to one subject in each of the two instruction conditions (categories or no categories). Thus a total of 24 subjects were used. The subjects were drawn from the same population as in Experiment 1.

A total of 40 questions, half true and half false, were presented. Of the true questions, 7 tested adjacent relations and 13 tested remote relations of various step sizes. Of the false questions, four tested adjacents and four tested remotes. Six additional false questions dealt with terms taken from two distinct nests, but the sequential order between the two terms was maintained. For example, the question might ask if a city term from one nest was "in" a state term from a second nest. This type of question will be labeled "between ordered." The remaining six false questions involved terms from two different nests presented in a nonsequential order. For example, the question might ask if a continent term from one nest was "in" a country term from a second nest. This type of question will be labeled "between reversed." The between-nest questions were intended to be indicators of whether or not subjects clearly understood that four separate nests had been described. The entire set of questions was selected from the set of all possible questions in such a way that terms from each nest were tested an equal number of times and were equally mentioned in both true and false questions to the greatest extent possible.

Results and Discussion

The results for the second experiment are presented in Table 4, which is organized by group, type of question, and strategy employed by the subject. As in Experiment 1, subjects within each condition were classified as to strategy used in answering questions based upon consistent responses. In comparison with Experiment 1, in which subjects used five different strategies and in which approximately 15% of the subjects did not respond consistently, subjects in the present experiment used only three different strategies, and none responded inconsistently. The increased consistency can be in part attributed to the fact that subjects did not have to respond from memory.

The effect of the categories/no-categories manipulation was subtle. Using Fisher's exact test, the number of subjects in the categories condition who used the correct strategy (10) is only marginally different from the number in the no-categories condition (6) ($p = .129$). However, none of the subjects in the categories condition assumed intransitivity or responded to the true remote assertions inconsistently, whereas one-third of the subjects in the no-categories condition assumed intransitivity. This difference is significant by Fisher's exact test ($p = .047$). Apparently, without a clear indication of the nature of the relationships being presented, the use of a variety of presentation orders causes a substantial proportion of subjects to assume intransitivity.

The relatively high proportion of subjects (two in each condition) who treated "A is in B" as symmetrical was something of a surprise. The "illicit conversion" of set inclusion statements of the form "All A are B" to assertions of identity (thereby implying that all B are A) is both intuitively plausible and well documented in previous research. The phenomenon appears to have something to do with the way in which quantifiers like "all" are used in ordinary discourse, as opposed to the technical discourse of logic, and the way in which "are" is interpreted (Revlis, 1975). However, to treat "A is in B" as an assertion that A is the same as B seems unprecedented by intuition or ordinary language. A rare example of such usage is found in the statement, "The city of Washington is in the District of Columbia." Some insight into the responses of the subjects who assumed symmetry was provided by a subject who stated that she thought the units were related in unique ways because they were on Pluto and not on Earth. It is possible that the particular cover story used in conjunction with the nonsequential presentation orders led four of the subjects to interpret the assertions of inclusion as symmetrical.

Different orders of presentation did not produce any statistically reliable differences in overall performance, nor were there any significant interactions. However, presentation order was manipulated within subjects, and its effect on performance could be observed only if most subjects differed in their error rates to various orders. The foregoing analysis suggests that differences

Table 4
Proportion of Correct Responses to the Six Types of Questions for Subjects Assuming Different Strategies in Experiment 2

		Type of Question					
		True		False		Between	
Subject Strategy	Number of Subjects	Adjacent	Remote	Adjacent	Remote	Ordered	Reverse
		Categories Condition					
Transitivity and Asymmetry	10	.95	.94	.91	.91	.95	.97
Transitivity and Symmetry	2	.93	.92	.12	.12	.83	.92
		No-Categories Condition					
Transitivity and Asymmetry	6	.97	.91	.98	.90	.95	.98
Transitivity and Symmetry	2	.93	.96	.06	.06	.83	.92
Intransitivity and Asymmetry	4	.89	.02	.94	1.00	1.00	1.00

in error rate are associated with individual subjects, not with conditions within each passage of text. The effect of presentation order is clarified somewhat by the results of the pilot study, in which the presentation order AB, BC, CD, DE was used throughout. Only 14% of these subjects failed to interpret inclusions correctly, none assumed symmetry, and the presence of geographical categories had no effect. It therefore appears that variations in presentation order can influence the selection of an interpretation (schema), but not in a direct way. Perhaps varied presentation orders increase subjects' general confusion about the meaning of inclusion. Experiment 3 was designed to investigate presentation order more systematically.

EXPERIMENT 3

The rationale of the third experiment was as follows: If certain presentation orders make it difficult to construct a hierarchy of inclusion relations, subjects otherwise predisposed to interpret "A is in B" correctly (in logical terms) will tend to seek other less appropriate schemata that make the construction task easier. The difficulty of construction was hypothesized to reflect the same processes people use in constructing linear orderings under comparable circumstances. To support this line of reasoning, subjects who had already demonstrated that they correctly interpreted statements of topological inclusion constructed inclusion hierarchies on the basis of varying orders of sentence presentation.

A series of recent papers has outlined an analysis of the effects of different presentation orders on the construction of linear orderings. Foos, Smith, Sabol, and Mynatt (1976) presented an explanation of the differences in error rates for different presentation orders in terms of five basic types of construction processes (see also Smith & Foos, 1975). Their approach was then extended by Mynatt and Smith (1977) to account for variations in study times when the individual comparisons were presented one at a time under the control of the subject (see also Smith & Mynatt, 1977). Smith has proposed a model of these constructive processes in a section of the paper by Potts, Banks, Kosslyn, Moyer, Riley, and Smith (1978). The model distinguishes five types of situations that arise when individual comparisons are presented one at a time, and it shows how these situations vary in difficulty as a function of the length of the search for matching elements in the comparisons, the number of matching elements involved, and the amount of mental rearrangement of previously ordered information required to achieve construction. This model was used in analyzing the data on construction.

Because subjects in the third experiment had been selected on the basis of their correct interpretations of inclusion relations, a further test concerned with distance effects was also conducted with the same subjects. Potts (1976) found that in some conditions subjects who used a transitivity strategy in processing inclusion relations of the form "All A are B" produced the distance effect typically found with linear arrays. That is, subjects responded more quickly to questions concerning remote relations than to adjacent relations. The presence of a distance effect indicates that the questions are being answered on the basis of a holistic or organized representation of the information, and not by combining the presented information in the course of answering the questions (Bower, 1971; Potts et al., 1978). If a distance effect can be shown when the relation "is in" is used, it would demonstrate the similarity of the memory structure and retrieval strategies associated with linear orders, all-A-are-B set inclusion relations, and the A-is-in-B inclusion relations.

Method

Seven undergraduates who had previously participated in a somewhat similar unreported experiment were recruited as subjects. The seven subjects were solicited from among those who assumed transitivity in answering a set of questions, and who missed fewer than 5% of the questions. Subjects were run in individual sessions and were paid $3 for participation. The experiment consisted of two phases: a test of hierarchy construction and a verification test.

The construction task was almost identical to the experiments of Mynatt and Smith (1977), except that the instructions presented the geographical cover story of Experiments 1 and 2 of the present paper, the pool of elements consisted of eight CVC nonsense syllables used in the previous experiments, and sentences had the form, "A is in B," where A and B were nonsense syllables. Subjects were told that the syllables were names of geographical or political units, but they were not told which syllables were names of cities, states, and so on.

The sentences appeared one at a time on a Digivue plasma display screen under the control of a NOVA 1220 computer. The subject responded by pressing keys on a typewriter console interfaced with the computer, which recorded study times in milliseconds, advanced the display to successive sentences, and, following the last sentence in a series, signaled the subject to write down the correct hierarchy on an answer sheet containing five vertically arranged lines. They were told to order the units either from city to continent or from continent to city, as long as they were consistent throughout the session.

Each subject completed six blocks of four trials. The first of these blocks was treated as practice, and data from this block were not used in any analysis. Within each block, the four presentation orders shown in Table 5 occurred once. These are the same orders used in Experiment 2. The presentation orders were ordered randomly within blocks, and on each trial, five of the eight nonsense syllables were selected randomly for use on that trial.

The second phase of the experiment was a verification test and immediately followed completion of the 24 construction trials. Subjects were first given a card containing printed sentences describing a single inclusion hierarchy similar to those they had just been constructing. They were asked to study the sentences until they felt they knew them well enough to answer questions about the relations between the five geographical names. Subjects then answered five blocks of 20 yes-no questions concerning all possible true and false adjacent and remote relations in the hierarchy. All questions were of the form "Is A in B?" and were presented on the Digivue screen. The subject responded by pressing one of two labeled keys, and reaction time from onset of the question was recorded in milliseconds. The word "Wrong" appeared on the screen whenever an incorrect answer was given; it remained there along

Table 5
Mean Number of Errors on the Presentation Orders Used in Experiment 3

Presentation Order	Example**	Mean Errors	Constructive Process Involved After Sentence*	
			2	3
		Match Orders		
Forward	AB, BC, CD, DE	.00	M1	M1
Reverse	DE, CD, BC, AB	.03	M2	M2
		Nonmatch Orders		
Confirmation	BC, DE, CD, AB	.03	N	D1
Disconfirmation	DE, BC, CD, AB	.17	N	D2

**Constructive process based on the analysis of Foos, Smith, Sabol, and Mynatt (1976).*
***The hierarchy is A B C D E, and "AB" means "A is in B."*

with the question for 2 sec. Speed and accuracy were equally stressed in the instructions.

Results and Discussion

The proportion of sets incorrectly recalled per subject was .057. The mean number of errors for each of the four orders is shown in Table 5. (A total of five was possible for each subject.) Table 5 also presents an analysis of the constructive processes ascribed by the theory to the subject following the second and third sentences of each trial. (The first sentence was not considered because no construction, i.e., combining relational information, can occur until the second sentence has been presented. The fourth sentence was also not considered. Because this was the last sentence in the sequence, study times might reflect modified construction processes.) A detailed account of the five processes has been given in the papers cited above. However, a brief summary follows.

The five processes fall into three classes, depending upon the number of elements in the relationship just presented that match elements previously stored in working memory. Thus, one, two, or no elements can match. The five processes are presented schematically in Table 6. Processes M1 and M2 occur when exactly one element matches. Specifically, when AB is stored and BC is presented, the match occurs on B and the hierarchy ABC is constructed (process (M1). When BC is stored[2] and AB is presented, the match occurs on B as in M1, but according to the theory, BC and AB must first be reordered before construction can be completed. Hence, process M2 takes longer and is more prone to error. When there is no matching element, the theory assumes that time is consumed in an exhaustive search for a match (process N, for "nonmatch"). Finally, following a nonmatch situation, both elements of the new relationship can match elements of an incomplete ordering in memory. Foos et al. (1976) distinguished two double-match processes, D1 and D2, paralleling processes M1 and M2.

Table 6
Mean Study Time (in Seconds) as a Function of Constructive Process in Experiment 3

Process	Sentence 2	Sentence 3
M1: AB + BC → ABC	5.11	4.42
M2: BC + AB → ABC	5.29	5.90
N: AB + CD → AB – CD	6.44	
D1: (AB – CD) + BC → ABCD		6.51
D2: (CD – AB) + BC → ABCD		13.58

Note–In each case, the constructed string is either ABC or ABCD in the example. Information preceding the plus sign is stored in memory; information following is to be added. The dash represents a marker element.

The mean study times for the second and third sentences are combined in Table 6 according to the constructive process involved. (Details of this type of analysis can be found in Mynatt and Smith, 1977.) A set of planned comparisons was used to analyze these data, along with the study times for the first sentence in each trial. The mean study time for the latter (4.26 sec) was significantly shorter than the mean for the second and third sentences combined (6.71 sec) [$F(1,6) = 9.95$, $MSe = 58.71$]. This was expected because no construction can occur on the first sentence. Additional comparisons showed that sentences involving the single-match processes (M1 and M2) required significantly less time than sentences involving nonmatch processes (N, D1, and D2) [$F(1,6) = 9.63$, $MSe = 67.99$]. The difference between processes M1 and M2 was not significant, but it shows M1 to be faster, as expected (means = 4.77 and 5.60, respectively). Process D1 took significantly less time than process D2, as was also expected [$F(1,6) = 6.05$, $MSe = 145.04$]. Thus, the data from the construction phase of Experiment 4 fit quite well into the pattern predicted by the theory.

It appears that when subjects correctly understand the inclusion relationship, they construct an inclusion hierarchy in much the same way as they construct linear orderings. The fact that process D2 took longer than process D1 is also evidence that subjects were performing the construction in a rehearsal system, rather than manipulating images. Mynatt and Smith (1977) found that when materials encouraged the formation of images, study times for process D1 were greater than for process D2. The outcome of the present experiment favors a more symbolic type of representation.

In the second phase of the experiment, the mean proportion of yes-no questions answered incorrectly was .03. Response times for incorrect answers were replaced with the mean of response times for correct answers of the same type within each subject. The mean response times as a function of truth value and ordinal distance (i.e., step size) are given in Table 7. As can be seen, there was a monotonic decrease in time as ordinal distance increased. An analysis of variance of these times displayed a significant linear trend in times for ordinal distance [$F(1,6) = 9.48$, $MSe = 4.19$]. The presence of a clear distance effect indicates that subjects were responding on the basis of a holistic representation of the inclusion relations rather than memory for the individual sentences. Truth value had no reliable effect [$F(1,6) = 2.37$]; however, there was a significant decrease in response times across blocks of trials [$F(40,240) = 3.66$, $MSe = .968$].

EXPERIMENT 4

One difference between set inclusion relations encountered in everyday use and artificial set relations is the number of subsets in each set. Most people are well aware that collies are not the only kind of dog, and that dogs are not the only type of mammal. Yet in previous studies, artificial inclusion hierarchies have been constructed from sets containing only one subset. For example, the passage of Table 1 describes only the Fundala tribe, and only farming as an occupation. No other tribe is mentioned, nor are the hunters, traders, or warriors of Central Ugala discussed. The subject in the second experiment who concluded that geography was somehow different on Pluto may be representative of many subjects in these experiments. Such subjects may treat the first one or two inclusion statements appropriately. However, after several sentences, it becomes obvious that the paragraph is not going to present a taxonomy or other structure for which a set inclusion schema seems appropriate. Subjects who then alter their interpretations of the sentences and adopt some other schema would, of course, produce logical errors.

Although the foregoing hypothesis explains why previously studied set inclusion materials lead to errors,

Table 7
Mean Response Time (in Seconds) to Adjacent and Remote Test Questions in Experiment 3

Type of Relation	Step Size: Adjacent 0	Remote 1	Remote 2	Remote 3
True	2.36	2.23	2.05	1.65
False	2.68	2.51	1.98	1.88

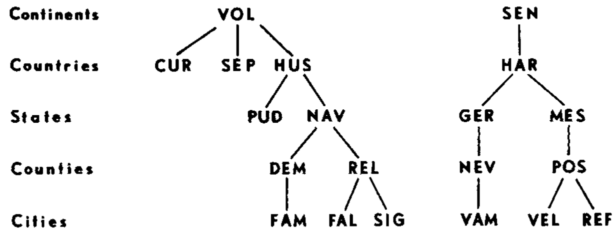

Figure 1. Hierarchical tree diagrams representing the inclusion relations of Experiment 4.

no clear prediction follows with regard to performance on more appropriate material. The problem is that systems of inclusion relations in which some sets contain two or more nonoverlapping subsets may be difficult to understand and remember for other reasons. To make these issues more concrete, consider the inclusion hierarchies represented as tree structures in Figure 1. In Experiment 4, subjects worked with paragraphs that described the geography reflected by the two hierarchies. Each node of the graphs is occupied by a nonsense-syllable name, with the elevation of the node representing its specific geographical level. These were the levels assigned to names in the categories condition of the second experiment. Thus, FAL and SIG are cities in REL county. REL and DEM counties are part of the state of NAV, and so on. An example of a true remote test item, or inference, would be that SIG is in NAV (Step Size 1). Inclusion hierarchies such as the two in Figure 1 will be referred to as "complex" to distinguish them from the simple nested series of elements that have been investigated previously.

The first property of complex hierarchies to be considered is that none of the possible orders of presentation is obvious and straightforward, as is the case with simple hierarchies. At some point in the presentation series, it is necessary to backtrack or reverse the direction of movement through the graph. Little is presently known about the effects of different orders of presentation on constructing complex networks of relations, and what has been reported (see Smith & Mynatt, Note 1) used comparisons of the form "A is greater than B." The latter results may not apply to inclusion relations. Moreover, DeSoto (1961) was one of the first to argue that with comparisons, hierarchies such as Figure 1 are intrinsically more difficult to understand and remember than simple linear orderings. Subjects in his experiments typically reduced partially ordered information to complete (linear) orderings.

There appear to be several conflicting predictions concerning performance on complex set inclusion hierarchies. Considerations of presentation order clearly suggest that they will be more difficult, but only by extrapolation from work on simple systems of comparisons (linear orderings). The overall complexity of the structure has more ambiguous implications. Generalizing from work with comparisons (partial orderings) suggests they may be more difficult, but it is just this aspect of complex structures that makes the appropriate schema selection more likely and therefore increases the chances that inclusion relations will be properly interpreted. It appears that complex structures hold enough interest to warrant an exploratory study, if for no other reason than to compare performance on them with previous findings for simple structures. In this spirit, a comparison of categories with no-categories conditions was also included in Experiment 4.

Method

The procedure and instructions were identical to those of Experiment 2. The elements used (nonsense syllables) and the form of the sentences were also identical to the earlier study. However, the sentences described the two inclusion hierarchies shown in Figure 1. Each hierarchy was described in a separate paragraph. The sentences within each paragraph were ordered so that the structure was described beginning with the bottom-most elements or areas and working upward (i.e., from cities to continent) to the greatest extent possible.

The subjects were tested on 36 questions. Although the use of complex hierarchies creates the possibility of indeterminate relations when the continent-to-city framework is not provided, all of the questions used were determinate. Of the 18 true questions, 6 concerned presented, adjacent relations, and 12 concerned remote relations. Of the 18 false questions, 2 concerned adjacent and 4 concerned remote relations from within a nest. Six false questions were of the between-ordered type, and six were of the between-reverse type.

A total of 32 subjects were used, half in the categories condition, and half in the no-categories condition.

Results and Discussion

The proportion of correct responses is presented in Table 8 as a function of type of question, condition, and strategy used by the subject in answering the questions. More than twice as many subjects in the no-categories condition assumed an incorrect strategy or were inconsistent as in the categories condition, a significant difference using Fisher's exact test ($p = .033$). Thus subjects given the categories as a framework were more likely to use the correct strategy in answering the questions than were subjects not given such a framework.

A comparison of outcomes between the present experiment and Experiment 2 suggests a remarkable parallel. The proportion of subjects who responded correctly in the categories condition was .81 and .83 in Experiments 2 and 4, respectively. The remaining subjects in this condition of both experiments treated the relationship as intransitive. The proportions in the no-categories condition were .50 and .56, with 33% and 25% subjects resorting to an intransitive interpretation, respectively. While the similarity is certainly not conclusive, there was also no clear evidence that complex hierarchies were much more difficult than simple ones. What is apparent is that a great deal more research is needed to determine whether the materials and conditions for the complex hierarchies are a special case or whether the present findings are truly representative.

CONCLUSIONS

A notable aspect of the experiments presented here is the individual differences observed among the subjects. In most cases subjects' responses to the tasks were consistent and readily classifiable according to strategy. We believe that looking at the data in this way gives a much more realistic and insightful view of how people process text containing set inclusions than does any analysis of group data. What becomes apparent is that the variables studied affect the distribution of subjects across a small number of different, but highly consistent, patterns of responding. We have interpreted these shifting distributions in various terms: schemata, strategies, interpretations of the relational term, and so on. However, the basic idea is simply that text processing does not occur in a cognitive vacuum. Subjects try to fit material in artifical texts into already established systems of representation. Instructions, sentence frames, additional ties to categories of familiar concepts, and presentation order all influence the representational machinery that is chosen, and this in turn influences the pattern of responding that is observed. Until the selection process is better understood, it will be easy to misinterpret the results of studies using artificial texts.

Table 8
Proportion of Correct Responses to the Six Types of Questions for Subjects Assuming Different Strategies in Experiment 4

		Type of Question					
		True		False		Between	
Subject Strategy	Number of Subjects	Adjacent	Remote	Adjacent	Remote	Ordered	Reverse
				Categories Condition			
Transitivity and Asymmetry	13	.99	.92	.96	.92	.97	.97
Intransitivity and Asymmetry	3	.87	.00	1.00	1.00	1.00	1.00
				No-Categories Condition			
Transitivity and Asymmetry	9	1.00	.95	1.00	.97	.98	.98
Intransitivity and Asymmetry	4	.96	.02	1.00	.93	1.00	.92
Inconsistent	3	1.00	.47	.83	.75	1.00	1.00

As a small contribution to this field, the present experiments have demonstrated the importance of at least three variables. (1) Very minor differences in the phrasing of a relational statement and the background information can have substantial effects on the representational schema selected. (2) When artificial material is tied to familiar concepts in a particular way (e.g., in the categories conditions), the chances are increased that the subject will adopt the schema from which the concepts are drawn. (3) With somewhat less confidence, we can say that complexity does not necessarily increase the difficulty of understanding artificial texts. In fact, where the schema is matched to the complexity of the information, complex sets of relations may be no more difficult than simple ones.

Finally, Experiment 3 provides an important link between two bodies of experimental results on understanding and remembering transitive, asymmetric relations. Among subjects who demonstrated an understanding of the logic of inclusion relations, the processes used in constructing a unified mental representation were similar to those used with linear orderings. In addition, the presence of distance effects in the retrieval of inclusion information indicates that artificial linear orderings and set inclusions may have similar underlying representations in memory.

REFERENCE NOTE

1. Smith, K. H., & Mynatt, B. T. *Effects of presentation order on construction of complete and partial orders.* Paper presented at the meeting of the Psychonomic Society, Denver, November 1975.

REFERENCES

Bartlett, F. C. *Remembering: A study in experimental psychology.* Cambridge, England: Cambridge University Press, 1932.

Bower, G. H. Adaptation-level coding of stimuli and serial position effects. In M. H. Appley (Ed.), *Adaptation-level theory.* New York: Academic Press, 1971.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, **3**, 193-209.

Carroll, M., & Kammann, R. The dependency of schema formation on type of verbal material: Linear orderings and set inclusions. *Memory & Cognition*, 1977, **5**, 73-78.

Ceraso, J., & Provitera, A. Sources of error in syllogistic reasoning. *Cognitive Psychology*, 1971, **2**, 400-410.

Chapman, L. J., & Chapman, J. P. Atmosphere effect reexamined. *Journal of Experimental Psychology*, 1959, **58**, 200-226.

Collins, A. M., & Quillian, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Dawes, R. M. Cognitive distortion. *Psychological Reports*, 1964, **14**, 443-459.

Dawes, R. M. Memory and distortion of meaningful written material. *British Journal of Psychology*, 1966, **57**, 77-86.

DeSoto, C. B. The predilection for single orderings. *Journal of Abnormal and Social Psychology*, 1961, **62**, 16-23.

Foos, P. W., Smith, K. H., Sabol, M. A., & Mynatt, B. T. Constructive processes in simple linear order problems. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 759-766.

Frase, L. T. Paragraph organization of written materials: The influence of conceptual clustering upon the level and organization of recall. *Journal of Educational Psychology*, 1969, **60**, 394-401.

Frase, L. T. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. *American Educational Research Journal*, 1970, **7**, 307-319.

Griggs, R. A. Logical processing of set inclusion relations in meaningful text. *Memory & Cognition*, 1976, **4**, 730-740.

Harris, J., & Monaco, G. E. Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General*, 1978, **107**, 1-22.

Hayes-Roth, B., & Hayes-Roth, F. Plasticity in memorial networks. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 506-522.

Mynatt, B. T., & Smith, K. H. Constructive processes in linear order problems revealed by sentence study times. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 357-374.

Noble, C. E. Measurements of association values, rated associations, and scaled meaningfulness for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, **8**, 487-521.

Potts, G. R. Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 727-740.

Potts, G. R. Bringing order to cognitive structures. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, N.J: Erlbaum, 1975.

Potts, G. R. Artificial logical relations and their relevance to semantic memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 746-758.

Potts, G. R., Banks, W. P., Kosslyn, S. M., Moyer, R. S., Riley, C. A., & Smith, K. H. Encoding and retrieving ordered relationships. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J: Erlbaum, 1978.

Revlis, R. Syllogistic reasoning: Logical decisions from a complex data base. In R. J. Falmagne (Ed.), *Reasoning: Representation and process.* Hillsdale, N.J: Erlbaum, 1975.

Smith, K. H., & Foos, P. W. Effect of presentation order on the construction of linear orders. *Memory & Cognition*, 1975, **3**, 615-618.

Smith, K. H., & Mynatt, B. T. On the time required to construct a simple linear order. *Bulletin of the Psychonomic Society*, 1977, **9**, 435-438.

Wason, P. C., & Johnson-Laird, P. N. *Psychology of reasoning: Structure and content.* Cambridge, Mass: Harvard University Press, 1972.

NOTES

1. We are aware of two studies that have used somewhat similar wording. Karl Scholz (cited in Potts, 1975) presented geographical information in a similar form but embedded in complex paragraphs. Revlis (1975) reported an informal study with sentences of the form, "All A are included in B," as an alternative phrasing of the set inclusion relation. Neither of these papers provides sufficient information to determine whether the use of "in" improves performance on inclusion relations.

2. The stored information can consist of a string of any length. For example, BCDE can be in memory, and process M2 would be required to add AB. Foos et al. (1976) even found that process M2 was comparable when adding elements to orderings containing a marker due to a nonmatch (e.g., in the third sentence of the order, AB. DE. CD.).

(Received for publication January 26, 1979;
revision accepted May 30, 1979.)