

Logical processing of set inclusion relations in meaningful text

RICHARD A. GRIGGS

University of Florida, Gainesville, Florida 32611

The present experiments attempted to resolve some recent conflicting findings in cognitive structure research between processing linear ordering and set inclusion relations described in meaningful paragraphs of text. In a self-paced study-test paradigm, college students studied such paragraphs. Definite processing differences were found for the two set-theoretic relations. Set inclusion test results were found to be due to faulty logical processing and not to memory retrieval problems. Subjects made invalid conversions of universally quantified statements and failed to make valid transitive inferences between such statements. This failure was found to be an increasing function of the distance separating the two terms in the set inclusion. No such problems were observed for linear orderings. The usual distance function was found: Accuracy on test questions about the ordering was an increasing function of the distance between the terms in the question. Results similar to those for linear orderings were observed for set inclusions when subjects were given special instructions about the validity and invalidity of symmetric and transitive inferences.

Recent research in linguistic comprehension has supported the constructivist theory of comprehension which argues that linguistic input is subjected to an abstract, constructive encoding process (e.g., Bransford, Barclay, & Franks, 1972; Cofer, 1973; Paris & Carter, 1973). The constructivist theory proposes that when a person comprehends meaningful verbal material, he educes a cognitive structure for its meaning. Linguistic inputs are viewed as cues to create meanings. Comprehension depends upon the cognitive contributions of the comprehender. He does not simply interpret and store sentences per se, but rather transforms the linguistic input into a cognitive structure which is a joint function of the input information and the comprehender's knowledge of the world. Meaning is constructed. (For more detail, see Bransford & McCarrell, 1974.)

The Bransford and Franks group (e.g., Bransford et al., 1972; Bransford & Franks, 1972; Bransford & Johnson, 1973) have played a major role in reviving constructivist theory and stimulating experimental interest in clarifying and characterizing it. They have devised numerous experiments to demonstrate the validity of this approach, but have failed to achieve a formal characterization of the theory. Their experiments

This paper is based on a PhD dissertation submitted to Indiana University. The author wishes to express his thanks to Frank Restle, the chairman of his dissertation committee, whose continued encouragement and advice made this research and paper possible, to the other members of his committee—Richard Shiffrin, Lloyd Peterson, David Pisoni, and Maynard Thompson—for their helpful suggestions and criticisms, and to Jerry C. Forshee for invaluable programming assistance. Research support was provided by NIMH Grant PHS RO1 MH 16817 to Frank Restle. Requests for reprints should be sent to: Richard A. Griggs, Department of Psychology, University of Florida, Gainesville, Florida 32611.

support an argument for the creation of cognitive structures (i.e., "wholistic semantic ideas," "semantic situations," etc.), but what are these structures? How can they be characterized? As Bransford and Franks (1971) point out, "A very important problem . . . concerns the question of *what is learned* . . . How can one characterize the nature of the semantic ideas that are acquired?" (p. 349). Such a characterization or model would seem necessary if the theory is to advance.

Potts (1972, 1974a, b; Scholz & Potts, 1974) has attempted to model this abstract constructive encoding process and to describe what is stored for at least one type of meaningful verbal material. Potts employed English paragraphs which described linear orderings. The paragraphs contained statements such as "The bear was smarter than the hawk, the hawk was smarter than the wolf, and the wolf was smarter than the deer." The orderings can be represented symbolically as $A > B > C > D$, where A, B, C, and D represent the terms of the ordering and ">" represents the comparative adjective used. Thus, in the above example statement, A, B, C, and D would stand for bear, hawk, wolf, and deer, respectively, and the comparisons would be made on the dimension of smartness. Linear orderings were used because of their transitivity which allows information to be deduced (i.e., $A > C$, $B > D$, $A > D$). The deduced pairs are referred to as remote pairs. The pairs which provide the necessary information about the ordering (i.e., $A > B$, $B > C$, and $C > D$) are called adjacent pairs.

Using a true-false study-test paradigm, subjects were asked to indicate whether a test sentence was true (consistent with the ordering information given in the paragraph) or false (inconsistent with the ordering information). True test sentences were statements of the adjacent and remote pairs (e.g., $A > B$, $A > C$). False

statements were formed by reversing the terms in each of the true statements (e.g., $B > A$, $C > A$). Potts was consistently found that subjects perform better on remote pairs than on adjacent pairs, even if the remote pairs are not presented and have to be deduced. Performance, in this case, may be based on proportion correct or response latency.

Performance, in general, is a function of the spatial distance separating the two terms in the test statement. This is referred to as the step-size effect. As the step size of the statement (the number of adjacent pairs necessary to deduce the statement) increases, reaction time decreases and proportion correct increases. However, the step-size effect is confounded in Pott's research by end-term effects. Performance on test sentences containing an end term (the first or last term in the ordering) is generally better than that on statements not containing an end term. Potts (1974a; Scholz & Potts, 1974), however, has shown that with end-term effects eliminated, performance is a decreasing function of inferential distance.

The opposite result, however, has been observed in research employing a similar set-theoretic relation, set inclusion (Frase, 1969, 1970). In general, Frase's studies demonstrate the superiority of memory for presented adjacent information. Frase employed paragraphs that contained sentences which asserted relations between five logical classes. The four adjacent sentences of his paragraphs can be symbolized As are Bs, Bs are Cs, Cs are Ds, and Ds are Es. Thus, the basic text structure of the paragraph is a series of implication relationships among classes, a set inclusion relationship, and can be symbolized as $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$, where the arrow points to the class which includes the other class. As with Potts' linear orderings, the stimulus material permits deducible information, for example, As are Cs, Bs are Ds.

Frase employed an incidental learning paradigm. Subjects were asked to read paragraphs in order to decide if the conclusion typed at the top of each paragraph was a valid deduction from the text, or, in some experiments, in order to underline the information in the paragraph that was needed to verify the conclusion. This task was followed by a free recall test which was sometimes followed by a recognition test comparable to Potts' true-false test. On free recall tests subjects were asked to write down all the assertions they could recall and any inferences they could generate. On the recognition test, 10 sentences were valid (the four adjacent sentences and the six valid inferences using universal quantifiers, e.g., All As are Bs) and 10 were invalid (the 10 valid sentences with the relation reversed, e.g., All Bs are As).

Frase's recall results are inconsistent with Potts' findings. In Frase (1969, 1970), recall of deducible remote information was definitely inferior to recall of presented adjacent information. The number of assertions that were recalled was a decreasing function of

step size. Frase's recognition results, though still different, are more similar to Potts' findings. Frase (1969, 1970) found no significant differences between recognition of deducible information and recognition of presented information.

These conflicting findings pose a serious problem for cognitive structure theory. Potts' results would appear to have little generality since Frase employed a similar set-theoretic relation, but observed completely different results. Set inclusions and linear orderings share several mathematical properties (e.g., irreflexivity, transitivity) and both can be represented by a simple directed graph. A linear ordering is antisymmetric, while a set inclusion is only asymmetric; but if Frase's statements are interpreted as A is a subset of B, then, according to Restle (1959), a string of such statements induces an ordering (see the case of "nested sets"). Therefore, the difference in cognitive structures in the Frase and Potts experiments is a serious conceptual problem. Granted, the similarity between the relations is not enough to insure similarity of cognitive structures, but the results seem diametrically opposed. Why is the transitivity property handled in such a different way in the set inclusion relation?

EXPERIMENT 1

The differences between Potts' and Frase's results could have been brought about by procedural differences, the actual set-theoretic structures studied, or the difficulty with which the set inclusion and linear ordering structures could be established. People have very powerful "ordering schemata" (DeSoto, 1960). The strong tendency for humans to structure elements by ordering them, the fact that Potts did not insert extraneous material between the presented adjacent pairs, and the fact that subjects studied a paragraph knowing that they were to be tested on the paragraph later make the method and stimulus materials very conducive to the establishment of a linear ordering. Frase's use of an incidental learning paradigm along with the insertion of extraneous material between the presented adjacent pairs in a paragraph appears to make the establishment of a set inclusion structure very difficult.

Experiment 1 attempted to reconcile the conflicting findings of Potts and Frase by investigating both types of set-theoretic material under comparable experimental conditions with procedural and material differences removed. Linear ordering paragraphs and set inclusion paragraphs were equated in the following ways. Paragraphs were of comparable word length, the number of terms in each ordering was equal to the number of sets in each set inclusion, and extraneous material was inserted between the presented adjacent pairs in both types of paragraphs. A self-paced study-test paradigm was employed with a true-false test similar to that used by Potts. Thus, any differences that were found should

have been brought about by the set-theoretic relations and not by procedural differences.

Method

Subjects. The subjects were 124 Indiana University undergraduates who were fulfilling course requirements for experimental participation. Each subject participated in one self-paced session, lasting approximately 20 min. Subjects were randomly assigned to four conditions containing 31 subjects each.

Apparatus. This experiment and those that follow were conducted in the Semantic Memory Laboratory at Indiana University. The experimental room consisted of eight individual booths, each containing a television monitor and response box. Stimuli were presented on a 30-cm Miratel TE video monitor controlled by an ADDS (Applied Digital Data System) MRD-280 video character generator. The displays consisted of black 6-mm-high capital letters on a white background. Responses were obtained using a two-button response box located immediately below each video monitor. Appropriate labels for the response buttons for each question appeared on the television screen directly above the buttons. Stimulus presentation and response collection were controlled by an IBM 1800 process-control computer.

Materials. The stimuli consisted of eight paragraphs.¹ Each of four of the paragraphs described a linear ordering among five items along a unique dimension. Each of the other four paragraphs described a set inclusion relation among five sets of items. Each paragraph presented the intended ordering or set inclusion as four adjacent relations in a chained order (i.e., AB, BC, CD, DE). Such paragraphs permit six remote pairs to be deduced (AC, BD, CE, AD, BE, and AE). A sample linear ordering paragraph is given in Table 1, and a sample set inclusion paragraph in Table 2. Unlike Frase's stimulus material, universal quantifiers are employed in each presented adjacent relation in the set inclusion paragraphs. This was done to prevent any inconsistency between presentation of the set inclusion pairs in the paragraph and in the test statements. In Frase's research, universal quantifiers were not always used to describe the pairs in the paragraphs, but were always used in the test statements.

Each of the five-term orderings and set inclusions used permitted testing 10 distinct relational pairs (the four presented adjacent pairs and the six deducible remote pairs). An equal number of true and false pairs were obtained by presenting the 10 pairs in forward and reverse order, resulting in 20 test statements.

Design. A between-subjects 2 by 2 factorial design was employed. The two factors studied were: (1) type of paragraph structure (linear ordering vs. set inclusion) and (2) memory load (massed presentation vs. single presentation). Four paragraphs were presented for study before testing in the massed presentation conditions, whereas only one paragraph was shown before testing in the single conditions.

Procedure. One to eight subjects participated in each experimental session. All stimulus materials (instructions, paragraphs, arithmetic problems, and questions) were presented on the tele-

Table 1
A Sample Paragraph Describing a Five-Term Linear Ordering Relation Used in Experiment 1

Due to the excessive amount of rainfall that occurred earlier this summer, the agricultural department found that corn was better than tomatoes. The bulk of corn is grown in the Midwestern states. Tomatoes were better than cucumbers. Tomatoes require a very acidic soil for best growth. Cucumbers were better than squash. Cucumbers require much moisture. Squash was better than peas. Squash requires a very long growing climate. Peas tend to grow well under a wide range of conditions. Though the rain was detrimental to some crops, overall it was a very productive year.

Table 2
A Sample Paragraph Describing a Five-Term Set Inclusion Relation Used in Experiment 1

All the Fundalas are outcasts from other tribes in Central Ugala. These people are isolated from the other tribes because it is the custom in this country to get rid of certain types of people. All the outcasts of Central Ugala are hill people. The hills provide a most accommodating place to live. All the hill people of Central Ugala are farmers. The upper highlands provide excellent soil for cultivation. All the farmers of this country are peace-loving which is reflected in their art work. All together, there are about fifteen different tribes in this area.

vision monitor. At the beginning of each session, the experimenter seated the subjects and familiarized them with the experimental setting. Subjects were then given instructions about the experimental task.² Subjects in the two single conditions were told that they would be shown a paragraph that they should read and study carefully for as long as they wished, because later they would be given a true-false test on the information in the paragraph. They were also informed about a short arithmetic task that they would be given after studying the paragraph and before the true-false test and that they must do well on the task or their experimental results could not be used.

The instructions for subjects in the two massed conditions were identical to the instructions for the single conditions except subjects were told that four paragraphs would be presented for study, followed by an arithmetic task and questions on each paragraph according to their order of presentation. After the true-false test, subjects in the single condition were told that another paragraph for study, arithmetic task, and true-false test would follow. This was done until all four paragraphs had been presented and tested.

Subjects were told that a sentence should be considered true if the information contained in it was either presented in the paragraph or logically deducible from the information in the paragraph, and false if the information contained in it was not in the paragraph and could not be logically deduced from the information in the paragraph.

The intervening arithmetic task was inserted between study and test to try to insure that the representation of the set-theoretic information was stored in long-term memory. The task was simply to indicate which of two arithmetic expressions was the larger. One expression appeared on the left side of the television screen and the other appeared on the right. In each case, one of the expressions had to be calculated by adding two two-digit numbers (e.g., 28 + 24), while the other expression was just a two-digit number and required no calculation. The subject was instructed to push one of two buttons labeled "Left" and "Right" to indicate which arithmetic expression was the larger. The results of this task were monitored to insure that subjects were performing the task conscientiously.

Results and Discussion

The linear ordering results were analyzed separately from the set inclusion results. The set inclusions proved to be more difficult than the linear orderings. Overall mean proportion correct for the set inclusions was .630, whereas it was .848 for the linear orderings. An analysis of variance performed on each set of proportion correct data revealed no significant memory load main effect. Thus, the following results are for the massed and single conditions combined. The analyses of variance revealed a significant main effect for the true vs. false factor for the set inclusion data [$F(9,540) = 13.21, p < .001$], but

no significant main effect was found for this factor for the linear ordering data.

Except where specifically noted, statistical analyses for this experiment and those experiments that follow were performed on the mean proportions correct using two-tailed *t* tests. The rest of the analyses will be broken down in terms of "true" and "false" performance because of the significance of the true vs. false factor for set inclusions. Since Potts (1974a) found both a step-size (distance) effect and an end-term effect for linear orderings, both of these types of analyses will be presented.

In general, Potts' step-size results for linear orderings and Frase's step-size results for set inclusions were replicated. Figure 1 presents mean proportions correct for true and false linear ordering test statements as a function of step size. Figure 2 gives these step-size results for true and false set inclusion statements. The step size of a true test statement is the number of adjacent statements necessary to deduce the statement. The step size of a false statement is the same as its true counterpart. Figure 1 indicates an increasing step-size effect for both types of statements, whereas Figure 2 shows this to be the case for false set inclusion statements but clearly not the case for true statements. The step-size function for true set inclusion statements decreases over the first three step sizes and then increases for Step Size 4. Step Size 4, however, consists of test statements containing both end terms of the set inclusion, that is, A and E. Thus, the rise in the decreasing function may be due to an end-term effect. These set inclusion results are similar to those reported in Frase (1969), except that he found subjects performed better on false (invalid) test statements than on true (valid) test statements.

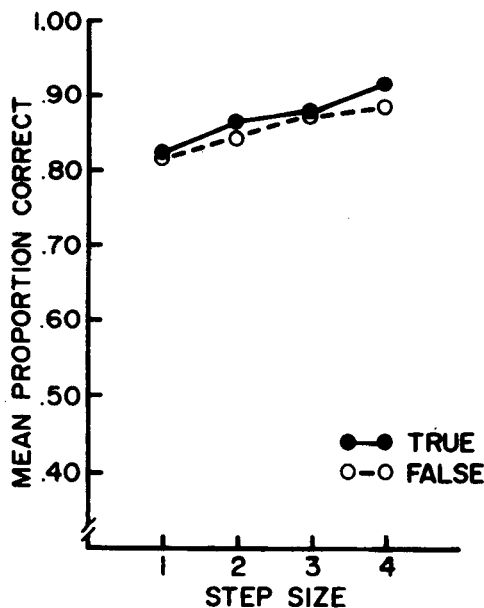


Figure 1. Mean proportion correct as a function of step size for linear orderings in Experiment 1.

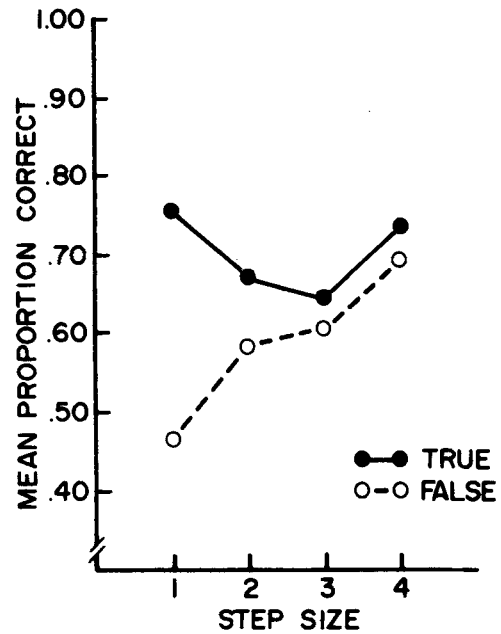


Figure 2. Mean proportion correct as a function of step size for set inclusions in Experiment 1.

Performance on deducible information was determined by computing mean proportion correct for test statements of Step Sizes 2, 3, and 4. For linear orderings, the mean proportion correct on deducible information (true, .878; false, .858) was significantly greater than the mean proportion correct on presented information (true, .821; false, .816) for both true and false test statements [$t(494) = 3.66, p < .001$ and $t(494) = 2.62, p < .001$, respectively]. For set inclusions, this was not the case for both true and false test statements. For false statements, the mean proportion correct on deducible information (.552) was significantly greater than the mean proportion correct on presented information (.465) [$t(494) = 7.46, p < .001$]. The opposite result was found for true set inclusion test statements. For true test statements, the mean proportion correct on presented information (.758) was significantly greater than the mean proportion correct on deducible information (.673) [$t(494) = 5.13, p < .001$]. Thus, Potts' finding that subjects performed better on deducible information for linear orderings was replicated, but definite true-false differences were found for set inclusions: For true statements, subjects performed better on presented information; for false statements, subjects performed better on deducible information.

To check for end-term effects, the mean proportions correct for true and false test statements containing an end term and containing no end terms were computed. For linear orderings, the mean proportion correct for test statements containing an end term (true, .863; false, .847) was greater than the mean proportion correct for test statements containing no end terms (true, .847; false, .828) for both true and false test statements, but

Table 3
Mean Proportions Correct and Standard Errors as a Function of Step Size for Set Inclusion Statements Containing No End Terms

Step Size	True		False	
	M	SE	M	SE
1	.716	.021	.381	.022
2	.556	.032	.488	.032

not significantly greater. These end-term results are not quite in line with those of Scholz and Potts (1974), who found significant end-term effects for linear ordering with proportion correct as the dependent measure. However, Scholz and Potts (1974) studied six-term linear orderings without extraneous material inserted between the presented adjacent pairs. The end-term results of the present experiment were in the correct direction but were not significant.

The set inclusion results were different. Clearly, the set inclusion results in Figure 2 were confounded by end-term effects. The mean proportion correct for test statements containing an end term (true, .726; false, .611) was significantly greater than the mean proportion correct for test statements containing no end terms (true, .663; false, .417) for both true and false statements [$t(494) = 3.45$, $p < .001$ and $t(494) = 9.43$, $p < .001$, respectively]. Table 3 gives the mean proportions correct for true and false set inclusion test statements which do not contain an end term, as a function of step size. The data have been greatly reduced and only two step sizes are left, but the remaining data do indicate a decreasing step-size function for true test statements and an increasing step-size function for false test statements.

EXPERIMENT 2

Since only the adjacent pairs were presented in Experiment 1, it is not clear whether the true-false differences and the low level of performance for set inclusions would appear if all the pairs (both adjacent and remote) of the set inclusion were presented in the study paragraph. Experiment 2 attempted to determine whether these set inclusion results would appear if all the pairs in the set inclusion were presented. Presentation of all the relational pairs in the set inclusion might help subjects to deduce the set inclusion structure from the study paragraph and, hence, possibly remove the true-false differences and raise the level of performance. Presentation of all the pairs should at least raise the level of performance on true remote pairs in the set inclusion, since they are presented and do not have to be deduced.

Method

Subjects. The subjects were 96 Indiana University undergraduates who were fulfilling course requirements for experimental participation. None of the subjects had participated in Experiment 1. Each subject participated in one self-paced session, lasting approximately 15 min. Subjects were randomly assigned to four conditions containing 24 subjects each.

Materials and procedure. The stimuli consisted of four paragraphs. Modified versions of two set inclusion paragraphs used in Experiment 1 were employed. One paragraph discussed native tribes and the other discussed recent space discoveries. The sentence describing the last adjacent pair (DE) in the set inclusion was deleted so that the paragraphs described four-term instead of five-term set inclusions. This was done because of experimental necessity. The insertion of the remote pairs into a paragraph increased the length of the paragraph sufficiently that a comparable paragraph for a five-term set inclusion would not fit on the television screen. Thus, to make experimental conditions comparable, four-term inclusions were employed.

Two types of paragraph structure were studied: (1) adjacent only (the paragraph presented the three adjacent pairs in the set inclusion) and (2) adjacent and remote (the paragraph presented all the adjacent and remote relations in the set inclusion). A chained order of presentation of the pairs was used in the adjacent only paragraphs. The order of pairs in the adjacent and remote paragraphs was AB, BC, AC, CD, AD, BD.

A between-subjects 2 by 2 factorial design was employed. The type of paragraph structure (adjacent only vs. adjacent and remote) varied, as well as the order of the paragraphs. The procedure and instructions were identical to those employed for the single presentation conditions in Experiment 1, except subjects were presented two paragraphs, instead of four, for study and test.

Results and Discussion

Presentation of all the relational pairs in the set inclusion did not help subjects deduce set inclusion structures from the paragraphs. As in Experiment 1, overall performance on set inclusions was poor. The mean proportion correct for the adjacent only conditions was .615 and for the adjacent and remote conditions, .613. An analysis of variance revealed no type of paragraph or order of presentation main effects, but a significant true vs. false main effect [$F(1,92) = 39.46$, $p < .01$] was found. Since there was no significant effect for presentation order, the results for the two adjacent only conditions as well as the results for the two adjacent and remote conditions will be combined in the rest of the analyses.

Table 4 gives mean proportions correct for true and false test statements as a function of step size for adjacent only and adjacent and remote conditions. As in Experiment 1, Table 4 indicates increasing step-size functions for false test statements and a decreasing step-size function for true statements in the adjacent only condition. However, the step-size function for true statements in the adjacent and remote condition is nearly flat (slightly U-shaped). The mean proportion correct for test statements of Step Size 1 (.817) was not significantly greater than the mean proportion correct for test statements of Step Sizes 2 and 3 (.744). This is under-

Table 4
Mean Proportions Correct and Standard Errors as a Function of Step Size for Adjacent Only and Adjacent and Remote Conditions

Step Size	Adjacent Only		Adjacent and Remote					
	True		False		True		False	
	M	SE	M	SE	M	SE	M	SE
1	.795	.028	.389	.026	.812	.025	.358	.029
2	.656	.039	.557	.040	.760	.035	.453	.039
3	.646	.049	.750	.044	.803	.041	.625	.050

standable since all the test statements were presented in the adjacent and remote conditions. Thus, a decreasing function would not be expected in this case. No end-term analyses were performed because the four-term set inclusions employed in this experiment did not provide enough data for such analyses.

No presented/deducible results could be computed for the true test statements for the adjacent and remote conditions since all true statements had been presented in these conditions. For adjacent and remote conditions, the mean proportion correct for deducible information (.510) was significantly greater than the mean proportion correct for presented information (.358) for false statements [$t(190) = 4.56, p < .001$].

For adjacent only conditions, the presented/deducible results were the same as those of Experiment 1 for set inclusions. For true test statements the mean proportion correct for presented information (.795) was significantly greater than the mean proportion correct for deducible information (.653) [$t(190) = 4.17, p < .001$]. For false statements the mean proportion correct for deducible information (.622) was significantly greater than the mean proportion correct for presented information (.389) [$t(190) = 7.14, p < .001$].

The mean proportions correct for remote pairs for the two types of paragraphs were also compared. Subjects studying adjacent and remote paragraphs did significantly better on true remote pairs than subjects who studied adjacent only paragraphs [$t(190) = 3.38, p < .001$], but at the same time they did significantly worse on false remote pairs [$t(190) = 2.98, p < .01$]. Hence, presenting the remote pairs in the set inclusion helped subjects on true remote test statements but made the false remote test statements more difficult.

EXPERIMENT 3

The results of Experiments 1 and 2 indicated that subjects did not appear to be integrating the information in the set inclusion paragraphs into set inclusion structures. It could be that the extraneous material inserted between the presented relational pairs is preventing such integration. This hypothesis was examined in Experiment 3. Subjects were presented study paragraphs which contained only the adjacent pairs in the set inclusion and no extraneous material whatsoever. The absence of the extraneous material was expected to facilitate the integration of the information into a set inclusion structure. If subjects do integrate the information into such a structure, the results should be similar to those for linear orderings. The true-false differences should disappear and the overall level of performance should rise.

Method

Subjects. The subjects were 52 Indiana University undergraduates who were fulfilling course requirements for experimental participation. None had participated in Experiment 1 or 2. Each subject participated in one self-paced session, lasting

approximately 15 min. Subjects were randomly assigned to two conditions containing 26 subjects each.

Materials and procedure. The stimuli consisted of eight paragraphs. The paragraphs were identical to the four set inclusion paragraphs used in Experiment 1, except there was no extraneous material in the paragraphs. Each paragraph consisted of only the four sentences which presented the four adjacent pairs in the set inclusion. Two presentation orders of the adjacent pairs for each paragraph were employed: a chained order (AB, BC, CD, DE) and a mixed order (AB, CD, BC, DE).

Using a between-subjects design, only one factor was studied: presentation order of adjacent pairs. The procedures and instructions were identical to those employed in the single presentation conditions in Experiment 1.

Results and Discussion

The results of this experiment are essentially the same as those of Experiment 1. The overall level of performance was low. The mean proportion correct for the chained order of presentation was .688 and for the mixed order of presentation, .607. An analysis of variance showed a significant main effect for presentation order of adjacent pairs [$F(1,50) = 4.65, p < .05$] and a significant main effect for the true vs. false factor [$F(150) = 72.58, p < .001$]. Overall performance was better for chained order and this superiority held up even when only presented pairs were considered (.851 for chained order compared to .722 for the mixed order).

The step-size effects observed for set inclusions in Experiments 1 and 2 were found again for both types of presentation order. Table 5 presents mean proportions correct for true and false test statements as a function of step size for both conditions. Table 5 indicates an increasing step-size function for false statements and a decreasing step-size function for true statements for the chained-order condition and the mixed-order condition.

As can be seen in Table 5, however, subjects in the mixed-order condition did better on false deducible information than on true deducible information. Other than this, the presented/deducible results for both the chained- and mixed-order conditions were the same as in Experiment 1. For true statements, the mean proportion correct on presented information (chained order, .851; mixed order, .722) was significantly greater than the mean proportion correct on deducible information (chained order, .758; mixed order, .590) for both types of presentation order [$t(206) = 3.75, p < .001$ and $t(206) = 6.94, p < .001$, for chained and mixed orders,

Table 5
Mean Proportions Correct and Standard Errors as a Function of Step Size for Chained-Order and Mixed-Order Conditions

Step Size	Chained Order		Mixed Order		Chained Order		Mixed Order	
	True	False	True	False	True	False	True	False
1	.851	.021	.502	.033	.772	.023	.406	.027
2	.772	.027	.596	.037	.644	.031	.628	.033
3	.745	.034	.639	.040	.543	.037	.635	.037
4	.740	.043	.731	.044	.519	.049	.740	.043

respectively]. For false statements, the mean proportion correct on deducible information (chained order, .633; mixed order, .649) was significantly greater than the mean proportion correct on deducible information (chained order, .502; mixed order, .406) for both types of presentation order [$t(206) = 3.84$, $p < .001$ and $t(206) = 8.49$, $p < .001$, respectively].

An end-term effect was observed for both types of presentation order. For chained order, the mean proportion correct for test statements containing an end term (true, .810; false, .611) was significantly greater than the mean proportion correct for test statements containing no end terms (true, .760; false, .510) for both true and false test statements [$t(206) = 1.98$, $p < .05$ and $t(206) = 2.83$, $p < .01$, respectively]. The results for the mixed-order condition are slightly different. A significant end-term effect was found for false statements but not for true statements. The mean proportion correct for false test statements containing an end term (.598) was significantly greater than the mean proportion correct for false test statements containing no end terms (.446) [$t(206) = 4.95$, $p < .001$]. For both conditions, when end-term effects are removed, the true step-size function is still decreasing, and the false step-size function is still increasing, and the two functions are almost symmetrical, indicating a definite step-size effect for both true and false statements, independent of end-term effects.

GENERAL DISCUSSION FOR EXPERIMENTS 1-3

The first three experiments lead to one conclusion: Subjects are not processing the set inclusion paragraphs in the same manner as the linear ordering paragraphs. Performance on presented information was consistently superior for true statements about the set inclusions, whereas performance on deducible information was superior for false statements. Overall performance on set inclusion material was not very good, with better performance on true statements than on false statements. Performance on false adjacent test statements was especially poor.

One hypothesis to account for these results is that the difficulty with set inclusion material is not in memory but in the subjects' logical processing of the set inclusion information. The statement All As are Bs could refer to either Relation 1 or 2 in Figure 3. It is possible that such statements are being interpreted to mean that the two sets, A and B, are equal (Relation 1). With such an interpretation one can deduce that the converse, All Bs are As, is true. Such illogical conversions would account for the poor performance on false adjacent statements and partially account for the overall poor performance on false statements. They would also account for the drop in performance on false remote test statements in the adjacent and remote conditions in

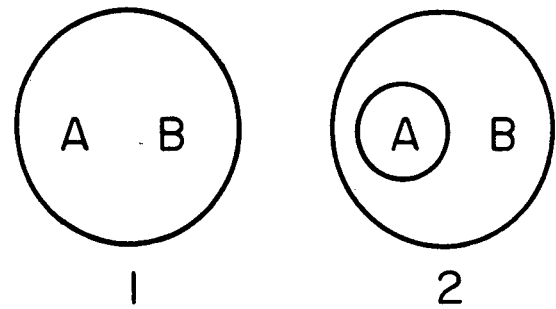


Figure 3. Possible A-B relations for the statement "All As are Bs."

Experiment 2, in which the remote pairs are actually presented in the study paragraph.

The reasoning error of illogical conversion is a common one in syllogistic reasoning tasks (Johnson, 1972; Wason & Johnson-Laird, 1972). This invalid conversion in which symmetry between sets is incorrectly deduced may help explain the poor performance on false adjacent questions but, certainly, will not account for the decreasing step-size function for true test statements and the increasing step-size function for false test statements. To account for these step-size functions, it is also hypothesized that there is a general tendency for subjects to reject test statements as step size increases. This agrees not only with the present results but with those of Frase (1969, 1970), who found that as step size of the test statements increased, subjects tended to reject more test statements as invalid. Hence, we find a decreasing step-size function for true statements and an increasing function for false statements. Therefore, it is hypothesized that subjects are failing to make the valid transitive inferences from the set inclusion information in the paragraph and this failure is an increasing function of step size. That is, the probability that the subject will say a true test statement is false increases as the step size of the test statement increases. Thus, the logical processing hypothesis proposes two faults in subjects' reasoning to account for the set inclusion results: (1) Subjects fail to make valid transitive inferences and this failure is an increasing function of step size, and (2) subjects make invalid symmetric inferences about the set relations.

EXPERIMENT 4

Experiment 4 was designed to investigate the logical processing hypothesis. If the difficulty is in the logical processing of the set inclusion information, and not in memory retrieval processes, then the same results should be observed when retrieval processes are not involved. Experiment 4 examined this possibility by presenting the questions simultaneously with the paragraph. Thus, there was no memory load at all, and subjects merely

had to calculate answers to the questions with the paragraph in front of them.

Method

Subjects. The subjects were 20 Indiana University undergraduates who were fulfilling course requirements for experimental participation. None of the subjects had participated in any of the preceding experiments. Each subject participated in one self-paced session, lasting approximately 15 min.

Materials and procedure. Only one paragraph was studied. The set inclusion paragraph that described native tribes in Experiment 1 was used. The procedure employed was very similar to Frase's incidental learning task. The paragraph was presented along with each question, combining the study and test phases. Thus, memory of the paragraph was not needed to answer the questions. Subjects received instructions which were similar to those used in the single presentation conditions in Experiment 1, except for the following difference: They were told that they would be shown a paragraph several times, each time with a question on the information in the paragraph, and should read as much of the paragraph as necessary in order to answer the question. Each of the 20 test statements was presented with the paragraph in a random order.

Results and Discussion

The results indicate that the major source of difficulty with set inclusions is not memory retrieval but, rather, faulty logical processing. The same pattern of true-false differences observed for set inclusions in the first three experiments was observed in this experiment, in which the subject had the study paragraph in front of him at all times. An analysis of variance revealed a significant main effect for the true vs. false factor [$F(1,18) = 17.17, p < .01$].

Table 6 presents the mean proportions correct for true and false test statements, as a function of step size. The data in Table 6 indicate a decreasing step-size function, with a slight increase for Step Size 4, for true test statements and an increasing step-size function for false test statements. When end-term effects are removed, the same type functions are still found.

The presented/deducible results also agree with those found for set inclusions in the first three experiments. For true test statements the mean proportion correct on presented information (.943) was significantly greater than the mean proportion correct on deducible information (.826) [$t(42) = 3.18, p < .01$]. For false test statements the mean proportion correct on deducible information (.659) was significantly greater than the mean proportion correct on presented information (.511) [$t(42) = 2.52, p < .05$].

Table 6
Mean Proportions Correct and Standard Errors as a Function of Step Size for True and False Test Statements

Step Size	True		False	
	M	SE	M	SE
1	.944	.023	.511	.060
2	.849	.046	.576	.063
3	.796	.054	.728	.091
4	.818	.084	.773	.091

EXPERIMENT 5

Experiment 5 was designed to further check the faulty logical processing hypothesis. If it is assumed that subjects are actually making errors of invalid conversion, then the erroneous reasoning should be prevented if counterexamples to the erroneously deduced symmetry are inserted into the study paragraph. Experiment 5 investigated this possibility by using paragraphs containing not only the adjacent pairs which establish the set inclusion but also counterexamples which destroy symmetry between the adjacent sets. For example, if the statement All As are Bs was in the paragraph, then the counterexample Some Bs are not As was also in the paragraph. Hence, subjects who studied the paragraph should be prevented from inferring All Bs are As. If the logical processing hypothesis is correct, then subjects who study paragraphs containing such counterexamples should perform better on false statements, and especially on adjacent false statements, than subjects who do not receive study paragraphs containing such counterexamples.

Method

Subjects. The subjects were 50 Indiana University undergraduates who were fulfilling course requirements for experimental participation. None of the subjects had participated in any of the preceding experiments. Each subject participated in one self-paced session, lasting approximately 20 min. Subjects were randomly assigned to two conditions containing 25 subjects each.

Materials and procedure. The stimuli consisted of four paragraphs. The two four-term adjacent only paragraphs used in Experiment 2, and a modified version of each, were employed. Counterexamples which destroyed the symmetry between the presented adjacent pairs were inserted into each of the two modified paragraphs.

A between-subjects design was employed. One factor (presence or absence of counterexamples) was studied, making two conditions. In one condition, subjects received the two paragraphs with counterexamples. In the other condition, subjects received the two paragraphs with no counterexamples. The procedure and instructions were identical to those used in the single presentation conditions in Experiment 1.

Results and Discussion

Experiment 5 clearly shows that the insertion of counterexamples that destroy the symmetry between sets improves performance on false statements. An analysis of variance revealed significant main effects for the absence vs. presence of counterexamples factor [$F(1,48) = 31.3, p < .001$] and the true vs. false factor [$F(1,48) = 51.4, p < .001$]. Overall performance was better on paragraphs containing counterexamples. The mean proportion correct for test statements for paragraphs containing counterexamples (.688) was greater than the mean proportion correct for test statements for paragraphs not containing counterexamples (.558).

This superiority was brought about by better performance on false test statements. The mean proportion correct for false statements for paragraphs containing

counterexamples (.777) was significantly greater than the mean proportion correct for false statements for paragraphs with no counterexamples (.410) [$t(98) = 7.93, p < .001$]. The mean proportion correct for adjacent false statements for paragraphs with counterexamples (.727) was over twice as large as the mean proportion correct for adjacent false statements for paragraphs with no counterexamples (.347). There was no significant difference between the mean proportions correct for true statements. Hence, it is clear that the insertion of counterexamples improved performance on false test statements, as predicted by the erroneous logical processing hypothesis, and did not affect performance on true test statements.

The true-false differences in step-size effects were found again for both types of paragraphs. Table 7 presents mean proportions correct, as a function of step size, for true and false test statements for both conditions. Table 7 shows decreasing step-size functions for true test statements and increasing step-size functions for false statements. Table 7 clearly demonstrates the superiority of performance on false test statements for paragraphs containing counterexamples. For all step sizes, mean proportion correct on false test statements was much greater than mean proportion correct on true test statements. No end-term analyses were performed because sufficient data was not available for such analyses, as four-term set inclusions were employed.

The presented/deducible results were also in agreement with the set inclusion results of the other experiments. First, consider paragraphs with no counterexamples. For true test statements, the mean proportion correct on presented information (.833) was significantly greater than the mean proportion correct on deducible information (.580) [$t(98) = 5.74, p < .001$]. For false statements, the mean proportion correct on deducible information (.473) was significantly greater than the mean proportion correct on presented information (.347) [$t(98) = 2.85, p < .01$]. The same results were found for paragraphs with counterexamples. For true statements, the mean proportion correct on presented information (.707) was significantly greater than the mean proportion correct on deducible information (.493) [$t(98) = 5.39, p < .001$]. For false statements, the mean proportion correct on deducible information (.827) was significantly greater than the mean proportion correct on presented information

(.727) [$t(98) = 2.99, p < .01$]. Hence, regardless of whether counterexamples are present, performance is better on presented information when true test statements are considered and better on deducible information when false test statements are considered.

EXPERIMENT 6

In Experiment 5, performance improved on false test statements, but mean proportion correct remained a decreasing function of step size for true statements and an increasing function for false test statements. Thus, invalid conversions partially account for the difficulty subjects have with set inclusions, but not entirely. It appears that subjects do not make the transitive inferences for set inclusions that they make for linear orderings. Therefore, in the preliminary instructions in Experiment 6, subjects were told that such transitive inferences are valid, in the hope that this information would encourage the use of the transitive inferences during the experiment. To prevent invalid symmetric inferences, subjects were also informed that these inferences are invalid. Dickstein (1975) used a similar strategy for syllogistic reasoning. He found that instructions designed to reduce errors of illogical conversion and probabilistic inference in syllogistic reasoning significantly improved performance.

If the primary causes of the set inclusion results in the first five experiments were erroneous symmetric inferences and failure to make transitive inferences, then subjects receiving instructions that symmetric inferences are invalid and transitive inferences are valid should be expected to reduce set inclusion structures from the study paragraphs. The results should then be similar to the linear ordering results: There should be no true-false differences in step-size functions and performance on false test statements should be improved.

Method

Subjects. The subjects were 40 Indiana University undergraduates who were fulfilling course requirements for experimental participation. None of the subjects had participated in any of the preceding experiments. Each subject participated in one self-paced session, lasting approximately 15 min. Subjects were randomly assigned to two conditions containing 20 subjects each.

Material and procedure. The stimuli consisted of the four set inclusion paragraphs used in Experiment 1. A between-subjects design was employed. One factor was studied: type of instructions. Subjects in the control condition received regular instructions, and subjects in the experimental condition received special instructions about valid and invalid deductions.

The procedure was identical to that employed in the single presentation conditions in Experiment 1, with one exception. Subjects in the experimental condition received instructions similar to those in the single presentation conditions in Experiment 1, except the following additional information about the validity and invalidity of deductions was provided:

"The paragraph will describe relationships between sets of items. The following is a sample of the type of material you will be studying—All brown elephants are African. All African elephants have long trunks. All elephants with long trunks have

Table 7
Mean Proportions Correct and Standard Errors as a Function of Step Size for Paragraphs With Counterexamples and Without Counterexamples

Step Size	No Counterexamples		Counterexamples		No Counterexamples		Counterexamples	
	True	False	True	False	True	False	True	False
1	.833	.030	.347	.037	.707	.039	.727	.037
2	.600	.045	.360	.054	.500	.043	.760	.038
3	.540	.071	.700	.065	.480	.071	.960	.028

large tusks. Consider the first sentence. This only means that all brown elephants are African. You cannot logically deduce that all African elephants are brown. The same is true for the other relational statements. Reversible statements are not logically deducible. There are, however, statements which can be logically deduced from this set of relations. They are the following: (1) All brown elephants have long trunks, (2) All African elephants have large tusks, and (3) All brown elephants have large tusks. (1) and (2) are two-step logical deductions. That is, they are the result of combining two of the relational statements. For example, (1) is deducible by combining the two statements 'All brown elephants are African' and 'All African elephants have long trunks.' (3) is a three-step deduction. (1), (2), and (3) are the only statements that can be logically deduced from the information in the sample material."

Results and Discussion

Overall level of performance was raised by the special instructions. An analysis of variance of the proportion correct data revealed a significant main effect for type of instructions [$F(1,38) = 12.34, p < .01$]. Mean proportion correct for the condition using special instructions was .828, while it was only .630 for the condition using regular instructions. The overall mean proportion correct for the special instructions condition is comparable to the mean proportion correct for linear orderings in Experiment 1 (.843).

The special instructions also removed the differences in overall mean proportion correct for true and false test statements. For regular instructions, the mean proportion correct for true statements (.730) was significantly greater than the mean proportion correct for false statements (.530) [$t(158) = 7.30, p < .001$]. For the special instructions, however, the mean proportion correct for true statements (.841) was not significantly different from the mean proportion correct for false statements (.815).

The special instructions removed the true-false differences in step-size effects. Table 8 presents the mean proportions correct, as a function of step size, for true and false statements for both conditions. As shown in Table 8, for regular instructions the usual set inclusion results for step size were found. For true statements mean proportion correct was found to be a decreasing function of step size with a slight increase for Step Size 4. For false statements mean proportion correct was found to be an increasing function of step size. However, under special instructions the step-size results were different: The step-size function for true state-

ments increased; the step-size function for false statements increased and then became almost flat over the last three step sizes. The overall level of performance was high for both functions.

If end terms are removed from the step-size analyses, a definite decreasing step-size function for true statements and a definite increasing step-size function for false statements are still found in the regular instructions condition. Under special instructions, however, unconfounded step-size functions for both true and false statements increased.

For the regular instructions condition, the usual presented/deducible set inclusion results were also obtained. For true statements, the mean proportion correct on presented information (.809) was significantly greater than the mean proportion correct on deducible information (.677) [$t(158) = 4.84, p < .001$]. For false statements, the mean proportion correct on deducible information (.610) was significantly greater than the mean proportion correct on presented information (.409) [$t(158) = 5.98, p < .001$]. Different results were obtained under special instructions: For true statements, the mean proportion correct on deducible information (.865) was significantly greater than the mean proportion correct for presented information (.806) [$t(158) = 2.06, p < .05$]. For false statements, the mean proportion correct on deducible information (.828) was greater, but not significantly greater, than the mean proportion correct on presented information (.791).

The results obtained for the condition employing special instructions are very similar to those for the linear orderings in Experiment 1. It appears that giving subjects special instructions enabled them to educe a simple set inclusion structure from the study paragraph, bringing about results similar to those for linear orderings due to the similarity between the two relations. Without such instructions, subjects do not appear to make the transitive inferences necessary to educe a set inclusion structure. As the number of relational statements that have to be combined to make a transitive inference increases, subjects seem more hesitant to make it. Along with this hesitancy to make transitive inferences, subjects also appear to be making invalid symmetric inferences. These two factors seem to be responsible for the true-false differences and the poor performance observed for set inclusions. When subjects are instructed that transitive inferences are valid and symmetric inferences invalid, the true-false differences disappear and the level of performance is raised to one comparable to that for linear orderings.

SUMMARY

The present experiments attempted to resolve some recent conflicting findings in cognitive structure research. Research by Potts (1972, 1974a) and Frase (1969, 1970) indicated that two very similar set-

Table 8

Mean Proportions Correct and Standard Errors as a Function of Step Size for Regular Instructions and Special Instructions

Step Size	Regular Instructions		Special Instructions		Regular Instructions		Special Instructions	
	True	False	True	False	True	False	True	False
1	.809	.024	.409	.033	.806	.031	.791	.034
2	.708	.031	.571	.039	.846	.028	.833	.036
3	.619	.040	.625	.042	.869	.032	.825	.038
4	.700	.105	.700	.109	.913	.032	.838	.042

theoretic relations described in meaningful paragraphs of text were being processed in very different ways. The present experiments lend support to this finding. Under comparable experimental conditions, set inclusions were processed very differently than linear orderings. Subjects appeared to make invalid conversions of universally quantified set inclusion statements but fail to make valid transitive inferences between such statements. The failure to make transitive inferences was found to be an increasing function of step size. With linear orderings, the usual distance effect was observed: Test accuracy was an increasing function of step size. Results similar to those for linear orderings were found for set inclusions, however, when subjects were instructed about the validity and invalidity of symmetric and transitive inferences before studying a paragraph. Hence, the conflicting findings were attributed to faulty logical processing of set inclusion information.

The invalid conversion error observed for set inclusions agrees with the syllogistic reasoning literature, but there does not appear at present to be a good analytical explanation of the subjects' failure to make the transitive inferences. This failure was observed not only in the present experiments but also in Frase's experiments. A memory for the premises explanation would not seem applicable in the present experiments, since, in Experiment 4, both invalid conversion and the failure to make transitive inferences were observed when the subjects were tested on the set inclusion information with the paragraph directly in front of them. Thus, the same results were found when memory retrieval processes were not involved.

As Erickson (1974) points out, "The average college student doesn't seem to be very 'logical.' That is, there exist a number of very simple situations in which the typical experimental subject doesn't behave the way a logician would say he ought to behave" (p. 305). The experimental conditions investigated in the present experiments appear to bring about such a situation, and further research is needed to account for such behavior.

REFERENCES

- BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, 3, 193-209.
- BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350.
- BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas: A review. *Cognition: International Journal of Cognitive Psychology*, 1972, 1, 211-249.
- BRANSFORD, J. D., & JOHNSON, M. K. Considerations of some

- problems of comprehension. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.
- BRANSFORD, J. D., & MCCARRELL, N. S. A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes*. Washington, D.C: Lawrence Erlbaum, 1974.
- COFER, C. N. Constructive processes in memory. *American Scientist*, 1973, 61, 537-543.
- DESOTO, C. B. Learning a social structure. *Journal of Abnormal and Social Psychology*, 1960, 60, 417-421.
- DICKSTEIN, L. S. Effects of instructions and premise order on errors in syllogistic reasoning. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 1, 376-384.
- ERICKSON, J. R. A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium*. New York: Halsted Press, 1974.
- FRASE, L. T. Structural analysis of the knowledge that results from thinking about text. *Journal of Educational Psychology Monograph Supplement*, 1969, 60(6, Part 2).
- 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 errors in comprehending set inclusion relations in meaningful text (Doctoral dissertation, Indiana University, 1974). *Dissertation Abstracts International*, 1975, 35B, 5155B. (University Microfilms No. 75-8942, 103).
- JOHNSON, D. M. *Systematic introduction to the psychology of thinking*. New York: Harper & Row, 1972.
- PARIS, S. G., & CARTER, A. Y. Semantic and constructive aspects of sentence memory in children. *Developmental Psychology*, 1973, 9, 109-113.
- 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. Storing and retrieving information about ordered relationships. *Journal of Experimental Psychology*, 1974, 103, 431-439. (a)
- POTTS, G. R. Incorporating quantitative information into a linear ordering. *Memory & Cognition*, 1974, 2, 533-538. (b)
- RESTLE, F. A metric and an ordering on sets. *Psychometrika*, 1959, 24, 207-220.
- SCHOLZ, K. W., & POTTS, G. R. Cognitive processing of linear orderings. *Journal of Experimental Psychology*, 1974, 102, 323-326.
- WASON, P. C., & JOHNSON-LAIRD, P. N. *Psychology of reasoning: Structure and content*. Cambridge: Harvard University Press, 1972.

NOTES

1. These paragraphs were provided by George R. Potts, are modifications of paragraphs provided by him, or are modifications of paragraphs used in Frase (1969). Copies of the exact paragraphs used in this experiment and subsequent experiments described in this paper may be obtained upon request from the author.

2. The exact instructions are available in Griggs (1974).

(Received for publication March 29, 1976;
revision received June 16, 1976.)