

The dependency of schema formation on type of verbal material: Linear orderings and set inclusions

MARIE CARROLL and RICHARD KAMMANN
University of Otago, Dunedin, New Zealand

Two paragraph structures, linear orderings (Potts, 1972) and set inclusions (Frase, 1969), were equated on relevant independent variables to examine recognition memory for relationships between pairs of terms. Experiment 1 replicated Potts' findings: Proportion correct was greater on remote than on adjacent linear pairs. For set pairs, however, there was an interaction between remoteness and truth, giving better performance on remote false and adjacent true pairs. It was suggested that the memory representation (schema) for linear structure was predicted by Potts' end-term anchoring and rating-scale strategies, while set structure performance reflected logical errors in processing. Experiment 2, an "open book" test, confirmed this view: In the nonmemory task, there was no evidence of schema formation for linear structure, whereas set performance still merely reflected logical errors in processing.

There is widespread support for the idea that linguistic information is stored in memory in a form resulting from a semantic integration of the ideas contained in a number of sentences. Potts (1972) attempted to clarify the nature of the memory representation of linguistic information structured as a four-term ordering of the form $A > B > C > D$, where the letters stand for the first to fourth terms in the ordering. The series $A > B > C > D$ can be broken into six pairs, three of which (the *adjacent* pairs $A > B$, $B > C$, $C > D$) are essential to the establishment of the ordering, and three of which (the *remote* pairs $A > C$, $A > D$, $B > D$) are redundant in the sense that they can be deduced from some subset of the necessary pairs. Potts found that subjects' true-false recognition test performance, measured both as proportion correct and response latency, was better on the remote pairs than on the adjacent pairs, even when the subjects had originally encountered only the adjacent pairs. He concluded that subjects could not have remembered the paragraphs by storing only the individual sentences; instead, they imposed on the whole paragraph a structure bearing a different form from the original. The proposed structure was suggested by the latency profile in which recognition response times were generally found to be a decreasing function of the degree of remoteness of the pair. To account for this, Potts proposed that the subjects first transformed the structure given in the paragraph into an imaginary spatial continuum (or rating scale) to represent the relations $A > B > C > D$. The occurrence of such spatial representation has been reported by other researchers (cf. Huttenlocher, 1968). Next, to account for the superior recognition of the most remote pairs, Potts reasonably deduced that in the serial ordering the end terms A and D have better representation or accessibility in memory (end-term anchoring effect) than do the middle terms,

B and C; thus, pairs involving the end terms would be more accurately and quickly recognized as true or false.

However, in a related series of studies, Frase (1969, 1970) found a reversed effect, that is, the adjacent pairs presented in the paragraph were superior to the remote pairs for both recall and recognition. Whereas Potts' paragraphs stated linear orderings (e.g., As are smarter than Bs, or $A > B$), Frase presented set inclusion information about the terms in this paragraph (e.g., all As are Bs, all Bs are Cs, etc.). If Frase's subjects followed the processing model proposed by Potts, the paragraph would have been encoded into a spatial representation for A B C D E (Frase used five terms requiring four adjacent pairs), and thus would have produced better recognition for the remote pairs and for those involving the end terms A and E. But Frase generally found that the more remote the pair, the less readily it was recalled and recognized.

As there were a number of other differences between the two types of experiments, Griggs (Note 1) attempted to reconcile the conflicting findings in a series of studies which removed some of the procedural and material differences between the two. Briefly, he found that subjects did not process set inclusions as they did linear orderings, since the truth or falsity of a pair interacted with the remoteness of the terms for set inclusions and with the remoteness effect obtained for false, but not for true, sentences. Griggs concluded that when subjects are confronted with set relations, such as "All As are Bs," they erroneously deduce the symmetrical relationship, "All Bs are As." That such errors are probable is supported by Ceraso and Provitera (1971). Griggs also showed that when the subjects were given special instructions to prevent the reversal error for set inclusion statements, the expected remoteness effect reappeared, as Potts' model would predict.

The present experiments have a purpose similar to

those of Griggs, although they were conducted without knowledge of Griggs' work and, therefore, differ in several useful respects. First, the two types of paragraph were constructed identically in every respect except the type of structural relation involved. Further, latency, as well as proportion correct, was analyzed as a performance measure. Since the memory representation for linear orderings was assumed to be a reconstruction of the original terms into a serial array, the question is whether, under precisely comparable conditions, the set inclusions are processed in the same fashion, and, if not, how they are processed. A rating-scale explanation predicts superior performance on the remote pairs for set inclusions, as for linear orderings, and an end-term anchoring explanation predicts that set inclusion pairs containing end terms should have the shortest latencies, again as for linear pairs.

EXPERIMENT 1

The experimental methods and materials of Frase (1969, 1970) and Potts (1972) were made as comparable as possible, in order to investigate the different memory effects of their experiments. The task (from Potts, 1972) required the subject to respond "true" or "false" to sentences, following reading of a paragraph structured (from Frase, 1970), as a five-term linear ordering or a five-term set inclusion, using identical terms in both conditions.

Method

Subjects. Fifteen subjects were randomly assigned to each of the four conditions. The 60 first-year university students (average age 18 years) were tested individually.

Materials. A paragraph, adapted from Frase (1970), was presented either as a five-term linear ordering or a five-term set inclusion. In addition, the order of presentation of the sentences within each type of paragraph was varied: In the ordered condition, the pairs occurred in a sequence which reflected their natural order (AB, BC, CD, DE), while the pairs in the disordered condition were presented in a sentence order which did not involve a simple sequence of the pairs to be linked (BC, DE, CD, AB). Both Potts and Frase manipulated the presentation order variable, but only Frase found an interaction of order with type of information remembered. There were, then, four paragraphs, each constructed as either a set inclusion or a linear ordering, in both a disordered and an ordered sequence. An example of one of the linear-ordered paragraphs follows:

New astronomical discoveries have been made about the temperatures of life-supporting planets, so-called "Blue" planets, atmospheric planets, planets within 15 light years' distance from Earth, and planets in Galaxy IV. Among these, atmospheric planets are found to be warmer than life-supporting planets. However, life-supporting planets are warmer than "Blue" planets. And the "Blue" planets are warmer than planets in Galaxy IV. Finally, the Galaxy IV planets are warmer than planets within 15 light years' distance.

The paragraphs were equated on several important variables, aside from content, which had differed in the Frase and Potts materials. These differences are: (1) Frase used "filler" sentences between the crucial sentences; Potts presented only the crucial sentences. The present paragraphs excluded filler sentences so that subjects could more readily perceive the structure. (2) Potts presented either the three adjacent pairs or those three plus the remote pairs, while Frase's material used only adjacent pairs.

Only adjacent pairs were used here, since, under these conditions, superiority of remote over adjacent constitutes even stronger evidence for a transformation in memory. (3) Frase used five-term and Potts four-term paragraphs. Five terms were used in the present paragraphs, so that there might be one remote pair which did not contain an end term.

From each paragraph, 10 true (4 adjacent and 6 remote) and 10 false sentences were derived. The latter contained the terms of the true sentences in a reversed order. A three-term linear ordering and a four-term set inclusion practice paragraph (again adapted from Frase, 1969) was constructed, differing in content from the test paragraph but appropriate in structure to the particular condition to which the subject had been assigned. Data from pilot studies indicated that subjects would probably have more difficulty with the set inclusion relations and, in particular, would be prone to treat them as symmetrical and therefore reversible. This observation suggested that the structural difference between the two sets of experiments on linear orderings and set inclusions, rather than some other (procedural) difference, produced the conflicting findings. However, this hypothesized structural difference could be based upon a weak effect which could readily be removed by instruction, or a strong effect which would be difficult to overcome. In order to eliminate the weak form of the hypothesis, the subjects were given a special instruction on the set inclusion paragraphs which was not included for the linear orderings: "Note that when you are answering the questions, it is not valid to make backward inferences. For example, if all As are Bs, it does not necessarily follow that all Bs are As."

Furthermore, each set-training paragraph had one more sentence, and thus one more term, and twice as many test sentences as the linear. Twelve sentences were used for the set training, six true and six false; six sentences, three true and three false, were used for the linear training. After practice training to a criterion of two or fewer errors, subjects proceeded to the test paragraph. All paragraphs were typed on individual sheets of paper, and all sentences on individual 15 x 10 mm cards.

Apparatus. Subjects viewed the test sentences through a Cambridge Mark 3 (Unit Type 3) tachistoscope, with a connected response-time clock. The timer started at the onset of a stimulus card (also signaled by a buzzer) and stopped when the subject pressed one of the two response keys, labeled "true" or "false."

Procedure. Each subject was informed that he was participating in an experiment on the comprehension of, and memory for, written material. He read the practice paragraph and viewed the practice sentences individually. After each sentence, the keypress response was made and feedback was given. Following attainment of the criterion, for which no subject required more than two trials, the test paragraph was presented. The subject then responded to the 20 questions, presented in random order, with no feedback. After the cards were reshuffled, a second trial of the study-test sequence was given. Reading time for both practice and test paragraphs was unrestricted, and the subject was allowed to take notes while reading (this was optional), but the notes were not available to him during testing.

The subject was instructed to respond "true" to a sentence if it seemed consistent with the information in the paragraph and to respond "false" if it seemed inconsistent. Speed and accuracy of response were stressed. The interval between the response to one sentence and the onset of the next was approximately 4 sec, and an experimental session lasted from 30-40 min.

Results

The mean reading time for set passages (157 sec) did not differ significantly from the time (135 sec) for linear passages [$t(58) = .55$].

Proportion correct. The mean proportions correct for linear orderings were significantly greater than those for set inclusions [Trial 1: .87 and .74, $F(1,28) = 7.99$,

$p < .01$, $MSe = .11$; Trial 2: .93 and .83, $F(1,28) = 7.14$, $p < .02$, $MSe = .08$]. These proportions are great enough to allow a meaningful analysis of latencies for correct responses. Preliminary analyses of variance indicated that order had no effect for either linear or set; therefore, it was dropped from subsequent analyses.

Adjacent-remote differences. Table 1 presents the mean proportions correct for the four adjacent pairs (henceforth called Step Size 1) along with the six remote pairs (called Step Sizes 2, 3, and 4, depending on whether they require one, two, or three mediating terms, respectively). In all cases, proportions correct are greater on remote pairs than on adjacent pairs for the linear condition, although this difference is significant for false pairs only [$t(29) = 3.50$, $p < .01$, $SE = .03$]. However, for the set condition, remote performance is superior to adjacent performance only on false pairs [$t(29) = 7.43$, $p < .01$, $SE = .03$]. For true pairs, the opposite effect obtains, with adjacent performance better than remote [$t(29) = 3.39$, $p < .01$, $SE = .02$].

Step-size effects. Potts' findings show an increasing probability of correct responses with increasing structural distance; Frase's results demonstrate the opposite effect. Figure 1 shows the step-size effects for mean proportions correct for set inclusions and linear orderings. There is a problem in performing the step-size analysis, since data obtained from subjects at Step Size 4 (which represents only one sentence) can take only one of two values (0 or 1.00), while the proportions obtained from the other step sizes can take more than two values. To overcome this problem, data from pairs of subjects chosen at random were combined, giving 15 "super-subjects" in both conditions, and thus reducing the corresponding degrees of freedom from 58 to 28 in the analysis.

Inspection of Figure 1 shows that performance improves with increasing structural distance between the pairs for true and false linear sentences. In contrast, there is a decreasing step-size function for true set pairs and an increasing function for false set pairs. This interaction occurs on both trials [Trial 1: $F(3,84) = 11.43$, $p < .01$, $MSe = .02$; Trial 2: $F(3,84) = 5.87$, $p < .01$, $MSe = .01$] and supports Frase (1969, Figure 7), who

Table 1
Mean Proportions Correct on Adjacent and Remote Pairs (Experiment 1)

	Linear		Set	
	True	False	True	False
Trial 1				
Adjacent	.87	.77	.87	.55
Remote	.88	.90	.72	.80
Trial 2				
Adjacent	.91	.89	.90	.62
Remote	.95	.94	.87	.86

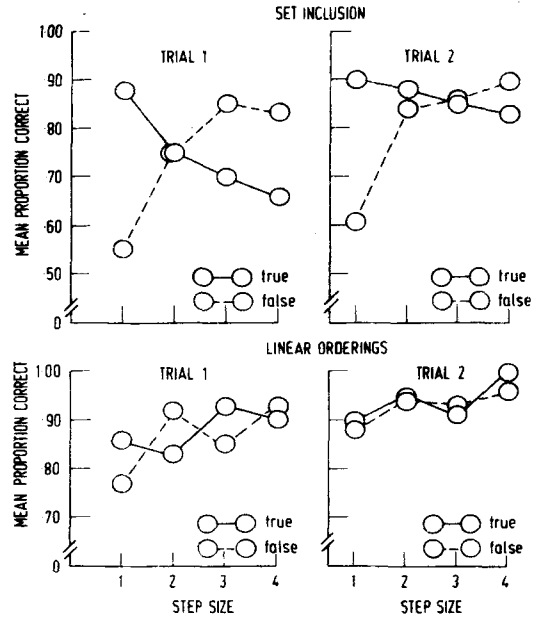


Figure 1. Mean proportions correct for true and false linear and set pairs as a function of step size.

Table 2
Mean Proportions Correct and Mean Latencies for Pairs with No End Terms (Experiment 1)

Proportion Correct	Linear		Set	
	True	False	True	False
Step 1	.87	.83	.86	.66
Step 2	.90	.91	.80	.88
Latency (Seconds)				
Step 1	2.99	2.95	2.93	3.20
Step 2	2.82	2.81	3.09	2.94

found that false sentences were rejected more frequently with increasing structural distance between pairs and that true sentences were accepted less frequently with increasing structural distance. The interaction also replicates Griggs' (Note 1) findings, and seems best interpreted as an overall increasing tendency to respond "false" to all items, as remoteness of the pairs increases.

Latency profiles. As noted in the introduction, Potts (1972) predicts that both end-term effects and structural distance operate jointly to determine performance on individual pairs. However, remoteness of the pair and end-term effects are confounded, and it is necessary to examine the step-size function for remote pairs which do not contain end terms, that is, Step Size 1 with pairs BC and CD, and Step Size 2 with pair BD. Table 2 shows the proportions correct and reaction times for correct answers only (averaged over order and trials). For true and false linear pairs, mean proportions correct are greater, and mean latencies are shorter on remote than on adjacent pairs. This is also the case for false set pairs.

The reverse effect, longer latencies and less accuracy on remote than on adjacent pairs, still obtains for true set pairs. Thus, the functions still closely parallel those obtained when all points are plotted.

On the other hand, remoteness by itself is not sufficient to predict latencies for individual sentences. Here, more precise information is afforded by examining *all* pairs. Tables 3 and 4 show the mean latencies for correct answers to individual linear and set sentences, respectively. The data are compared with those of Potts (1974, Figure 3), though comparison is possible only to a limited extent, since Potts used a six-term ordering, thus deriving 30 pairs. Nevertheless, when the response times for only those same 20 pairs in the present experiment are compared with Potts' (omitting pairs with an F term), the correspondence between the two profiles is remarkably good for true linear sentences [$r(8) = .78, p < .01$], but not for false linear sentences [$r(8) = .19, p > .05$]. According to the Potts model, false sentences beginning with E should be uniformly short, since, if the subject has successfully stored the information that E is the last term, he may immediately respond "false" to any pair beginning with E. However, as Table 3 shows, the false pairs beginning with the D term have the shortest latencies. Potts (1974, Experiment 1) suggests, as a possible explanation for this anomaly in his own data, that almost all subjects store A as the first term, but only some code E as the last term.

The set profile is quite different. Set true pairs have a significant negative correlation with Potts' true pairs [$r(8) = -.69, p < .05$]; the set false pairs are not significantly correlated [$r(8) = .36, p > .05$]. Thus, set latencies are predicted by neither a rating-scale nor an end-term anchoring explanation, since the profile bears no similarity to that for linear orderings found by Potts and in the present study. Indeed, in contrast to the linear pairs, the set pairs beginning with A have longer latencies than other true pairs, and latency does not vary inversely with remoteness for either true or false pairs.

Thus, with the materials equated in all respects except that of structure, set inclusions show a different

pattern of recognition memory than that of linear orderings.

Discussion

Experiment 1 confirmed the findings of Potts for linear orderings: Performance on remote pairs was superior to performance on adjacent pairs. This effect was not obtained for set inclusions, even when both kinds of paragraph structure were equated on important experimental variables. If the memory representation (henceforth called the schema) for set inclusions paralleled that for linear orderings—so that subjects transformed the original input into an ordered array of five terms and retrieved from memory not the original, but the transformed, information—then one would expect that true set sentences would be affected in the same way as false set sentences. Instead, only false set sentences showed the predicted improved performance with increasing structural distance. Further, the latency profiles indicate dissimilarity between the schemata formed from set and linear information.

An alternative explanation for the set inclusion data is that subjects did not form a schema for the terms in the paragraph at all. Instead, the data might reflect logical errors in the initial processing of the information. Despite having had practice with examples of permissible valid inferences, set subjects may nevertheless have misunderstood the subordinate-superordinate relationship of terms in a true set pair, and thus its asymmetry. The poor performance on false adjacent pairs and good performance on true adjacent pairs (Figure 1) indicates that subjects were relying mainly on their memory for the pairs of terms which had actually appeared together in the paragraph, and tended to respond "true" to all of these, assuming they were reversible. That is, contiguity at input appears to have been the principal determinant of "true" responding in the set condition. As pairs of terms became less like the original presentation condition (i.e., more remote), subjects became more likely to respond "false," giving an increasingly better hit rate for false sentences and a greater number of false alarms for true sentences with increasing step size. The next experiment examines the alternative explanation that subjects erroneously treated the set relation as reversible.

EXPERIMENT 2

To further test the idea that set inclusion results in Experiment 1 reflected logical errors in processing, and not a true memory effect, it is instructive to look at the processing of such information when no memory task is required. If set subjects were responding largely on the basis of their memory for actually presented information, their performance on remote and adjacent pairs should be the same in both a memory and a non-memory task. Presumably there should be the same

Table 3
Mean Latencies (Seconds) for Individual Pairs of Terms in the Set and Linear Conditions (Experiment 1)

Type of Pair	Set		Linear	
	True	False	True	False
AB	3.11	2.99	2.47	2.99
BC	3.12	3.25	3.13	3.14
CD	2.74	3.15	2.85	2.75
DE	2.93	3.78	2.94	3.03
AC	3.07	3.11	2.69	2.84
BD	3.09	2.94	2.82	2.81
CE	2.92	3.76	2.90	2.93
AD	3.16	3.29	2.38	2.69
BE	3.07	3.48	2.90	2.83
AE	3.33	3.36	2.78	3.01

tendency to reject all inferences, giving poorer performance on true remote pairs relative to true adjacent and better performance on false remote pairs relative to false adjacent. On the other hand, the evidence supports the hypothesis that a schema for linear information was formed in Experiment 1. Therefore, the prediction for a nonmemory task is that there will be better performance on both true and false adjacent pairs than on remote pairs because no schema need be formed; subjects may respond mainly on the basis of what they actually read without having to use a memory representation.

Experiment 2, then, is an "open book" comprehension task, where the two types of paragraph are actually present for inspection when the subject is answering questions about them. Though the task is similar to Griggs' Experiment 4, his study differs in two respects: (1) He presented no training instructions for set materials; (2) he used only one paragraph structure, the set inclusion, and omitted the important linear comparison.

The present study affords a direct comparison of the two paragraph structures by presenting both set and linear information.

Method

Subjects. Fifty-two female high school students (average age 15 years) were tested as a group. Thirteen subjects were assigned randomly to each of the four conditions.

Materials and procedure. Practice and test materials were the same as those used in Experiment 1, however, the practice and test sentences appeared beneath the corresponding paragraphs. The same instructions, to respond consistently with the information in the paragraph, were given, except that subjects were also told to refer to the paragraph when answering questions. Time to complete the task was unrestricted.

Results

Proportion correct. Table 4 shows the mean proportion correct (the only performance measure) in each group. The proportion correct for linear (.89) was significantly greater than that for set (.55) [$F(1,48) = 81.48, p < .01, MSe = .03$]. Nevertheless, both proportions are very low, considering the nature of the task, and may partly reflect the age difference between the subjects in the two experiments and the less satisfactory conditions

Table 4
Mean Proportions Correct (Experiment 2)

Test Pairs	Presentation Order		M
	Ordered	Disordered	
	Set		
True	.69	.64	.66
False	.35	.56	.45
M	.52	.60	
	Linear		
True	.95	.88	.91
False	.94	.81	.87
M	.95	.84	

Table 5
Mean Proportions Correct for Adjacent and Remote Pairs (Experiment 2)

	Set		Linear	
	True	False	True	False
Adjacent	.79	.40	1.00	.95
Remote	.57	.49	.85	.81

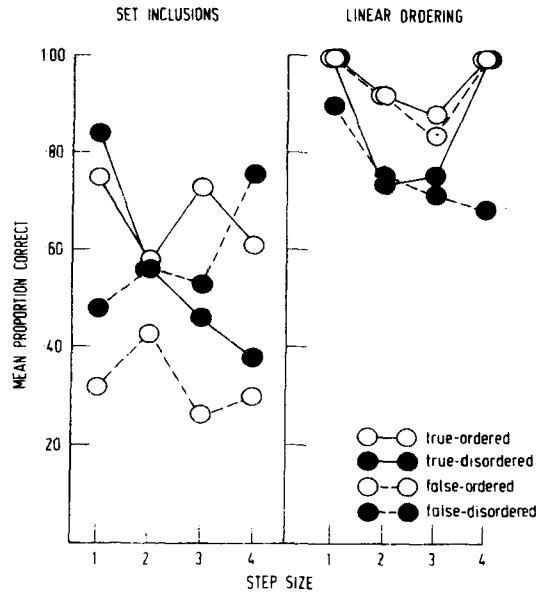


Figure 2. Linear and set mean proportions correct for ordered and disordered true and false pairs as a function of step size.

of being tested in a group. In contrast to Experiment 1, order of presentation interacted significantly with paragraph type [$F(1,48) = 6.08, p < .05, MSe = .03$]. Performance was better in the linear ordered and in the set disordered conditions. This last finding is anomalous under any theory, but may reflect a greater realization that the disordered set paragraph required more careful and exact problem solving.

Adjacent-remote difference. Table 5 shows that the adjacent vs. remote differences for linear pairs are in the opposite direction. Performance is significantly better on adjacent than on remote pairs [true: $t(24) = 3.77, p < .01, SE = .04$; false: $t(24) = 3.00, p < .01, SE = .04$]. In the set condition, however, the interaction of truth value with the type of pair found in Experiment 1 still obtains: Performance is better on remote than on adjacent false pairs [$t(25) = 1.78, p < .05, SE = .05$] and better on adjacent than on remote true pairs [$t(25) = 5.14, p < .01, SE = .04$].

Step-size effects. Figure 2 shows the step-size functions separated into the two order conditions. In the linear condition, the upturn of three of the four otherwise decreasing functions at Step Size 4 may be due simply to an immediate visual "anchoring" effect of the pair containing the first and last terms.

For set pairs, only the disordered condition shows the interaction between truth value and step size, with decreasing accuracy for true and increasing accuracy for false pairs as remoteness increases. The set findings are consistent with the hypothesis of logical errors: As the task becomes more difficult (disordered presentation), subjects increasingly reject inferential material, which accounts for the crossover of the two functions.

GENERAL DISCUSSION

The open-book test provides further evidence that set information is not retrieved from memory as a linear schema. In contrast to the linear case, set performance in a nonmemory task was no different from that in a memory task when the adjacent-remote difference was considered. The conclusions support those of Griggs (Note 1): When answering questions in either a memory or nonmemory task, set subjects (1) appear to assume incorrectly that a symmetrical relationship exists between adjacent pairs, and (2) fail to make valid transitive deductions. What does appear is a tendency to reject pairs as false as they become more remote, giving an increasing accuracy function for false pairs and a decreasing accuracy function for true pairs over step size. The latency profile gives further support to the hypothesis that set material is not retrieved from memory in the same way as is a linear ordering.

Griggs' (Experiment 6) special instructions included specific examples of correct processing: only when these were included were the set step-size functions similar to those for linear orderings. Merely instructing subjects about valid inferences and even allowing practice in dealing with set inclusions (as in the present studies) seems to have been ineffective in insuring task comprehension. The question arises of why people are so poor at drawing logical inferences from set information. Perhaps the difficulty reflects a possible ambiguity in the presumed asymmetry of set relations as used and interpreted in daily language. Most often, extralinguistic knowledge can be used to determine the asymmetry of subject and object in such a sentence as "All men are mammals." Such knowledge can also determine the symmetry of terms in sentences based on identity or definitional relations, such as "Ice is frozen water." No

such extralinguistic knowledge is available, however, when a person reads that "Googledydunks are ragleys" or "Atmospheric planets are life-supporting."

The present experiments may be taken to mean that the schema-formation process occurs with some types of verbal material but not with others. An alternative interpretation is that the encoding process tends to occur pervasively, but with some materials (e.g., set inclusions) the reader does not construe the sentences or form the schema expected on the basis of the logician's or linguist's analysis of the implied relations.

While there is a natural and sanctioned tendency for experimenters to choose stimulus materials which are most likely to reveal the operation of a hypothesized mental process, the present experiments suggest that such processes might best be considered stimulus bound until their generality across stimulus materials is confirmed. This comment might be considered applicable to laboratory experiments in general, and not solely to tasks of memory processes and abstraction.

REFERENCE NOTE

1. Griggs, R. A. *Logical errors in comprehending set inclusion relations in meaningful text* (Report No. 74-7). Bloomington, Indiana: Indiana University, Indiana Mathematical Psychology Program.

REFERENCES

- CERASO, J., & PROVITERA, A. Source of errors in syllogistic reasoning. *Cognitive Psychology*, 1971, 2, 400-410.
- FRASE, L. T. A structural analysis of the knowledge that results from thinking about text. *Journal of Educational Psychology*, 1969, 60 (Monograph Supplement 6).
- 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.
- HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, 75, 550-560.
- 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.

(Received for publication June 30, 1975;
revision received June 21, 1976.)