Journal of Experimental Psychology
1969, Vol. 82, No. 2, 205–215

INFLUENCE OF LANGUAGE ON SOLVING THREE-TERM SERIES PROBLEMS[1]

HERBERT H. CLARK[2]

Carnegie-Mellon University

Three linguistic processes have been previously proposed to account for the influence of language on deductive reasoning. To test for these processes, 100 *S*s were each asked to solve 64 different forms of the three-term series problem (e.g., "If John isn't as bad as Bill, and Pete isn't as good as Bill, then who is worst?"). The *S*s were allowed only 10 sec. per problem. The kind and relative number of errors *S*s made support the existence of an implicit problem solving strategy containing the three proposed linguistic processes. The strategy consists of (*a*) interpreting the propositions and question of a problem in terms of their deep structure relations, (*b*) retaining the information in this form, and (*c*) later searching memory for information that is congruent in deep structure with the information asked for in the question. For some problems, this strategy leads to longer solution times, hence more errors. For many problems, it also leads to preliminary solutions that show up as specific errors.

People are often led to reason fallaciously merely by the way reasoning problems are worded. Wording is particularly important in the so-called three-term series problems. To cite an example, the problem, "If John is better than Pete, and John is worse than Bill, then who is best?," encourages far more errors in problem solvers than the problem, "If John is better than Pete, and Bill is better than John, then who is best?," even though both problems superficially present the same information (Burt, 1919; DeSoto, London, & Handel, 1965; Hunter, 1957). The immediate aim of the present study was to account for the errors people make in solving such problems. But the wider purpose was to explain the memory processes of storing and retrieving abstract information retained from sentences. In a recently proposed theory for these memory processes (Clark, 1969), it was argued that the way in which a person stores and retrieves the information in a reasoning problem determines how quickly he solves the problem.

The theory proposed by Clark (1969) consists of three principles, briefly summarized here. The first is the principle of the primacy of functional relations. It asserts that the functional relations underlying a sentence, like the logical subject, verb, and object, are more available after comprehension than other less fundamental kinds of information. This principle, derived from linguistic theory (e.g., Chomsky, 1965), was first proposed in slightly different form by Miller (1962) and was supported subsequently by Mehler (1963), McMahon (1963), Gough (1965, 1966), Savin and Perchonock (1965), Slobin (1966), and Clark and Begun (1968).

Comparative sentences, like other constructions, have underlying functional relations. As several linguists (Chomsky, 1965; Doherty & Schwartz, 1967; Huddleston, 1967; Lees, 1961; Smith, 1961) have shown, the sentences "John is worse than Pete" and "Pete isn't as bad as John" both consist linguistically of two "base strings," *John is bad* and *Pete is bad,* that are conjoined in a comparative or equative construction. The first principle, then, asserts that a person "knows" more readily that John and

[1] The work reported here was supported in part by United States Public Health Service Research Grant MH-07722 from the National Institute of Mental Health. The author wishes to thank Stuart K. Card and Eve V. Clark for their suggestions on the manuscript and Judy Grumet for her assistance and advice in running the experiment.

[2] Requests for reprints should be sent to Herbert H. Clark, who is now at Stanford University, Department of Psychology, Stanford, California 94305.

Pete are bad (the functional relations expressed in the base strings) than that John is more extreme in badness than Pete.

The second principle, that of lexical marking, specifies the semantic complexity of words like *good* and *bad*. Several generations of linguists (e.g., Bierwisch, 1967; Greenberg, 1966; Lyons, 1963, 1968; Sapir, 1944) have observed that word pairs like good and bad are not symmetrical. Good, the so-called unmarked member of the pair, can be neutralized in some contexts, as in "How good was the movie?," whereas bad, the marked term, cannot. Good, but not bad, can also be neutralized in comparatives: "John is better than Pete" can mean that John and Pete are only being compared evaluatively, although "Pete is worse than John" presumes Pete and John to be bad. The principle of lexical marking is that the neutral senses of unmarked adjectives, like good, are coded in memory in a simpler form than the senses of marked adjectives, like bad. The main consequence is that storage and retrieval should be quicker for comparatives containing unmarked adjectives.

Lastly, the principle of congruence specifies how a search through memory will be made. Information can only be retrieved when it is congruent with the information being sought. Furthermore, this search demands congruence, not of superficial information like words or phrases, but of underlying functional relations. For example, the proposition "John isn't as bad as Pete" implies that John is bad and Pete is bad; the question "Who is best?" requires an X fitting the description *X is good.* Here there is a lack of congruence of underlying functional relations in proposition and question. The question must be implicitly reformulated to read "Who is least bad?" It is only at this point that congruent information can be found and the answer retrieved. Reformulation of a question, of course, should take more time.

A particular strategy for solving problems is implied in the three principles previously described. It goes as follows: The problem solver, in understanding a problem, retains information in such a manner that the underlying base strings are most accessible. (Comprehension time per se is presumably equal for all the comparative sentences used in the present study, and equal for all the negative equative sentences, since the same linguistic constructions are involved in all these comparatives.) It takes him longer in this process to store and retrieve semantically more complex, marked adjectives (like bad) than it does unmarked adjectives (like good). The question directs him then to search for an answer. First, he searches for stored information that is congruent with the question. If he succeeds, he can begin to look at the second-order comparative information and attempt to form an answer; if he fails to find con-

TABLE 1

Types of Determinate Three-Term Series Problems

Problem type	Form of problem	Analysis	Problem type	Form of problem	Analysis
I	A better than B B better than C	A is good B is good C is good	I′	A not as bad as B B not as bad as C	A is bad B is bad C is bad
II	C worse than B B worse than A	A is bad B is bad C is bad	II′	C not as good as B B not as good as A	A is good B is good C is good
III	A better than B C worse than B	A is good B is good, bad C is bad	III′	A not as bad as B C not as good as B	A is bad B is bad, good C is good
IV	B worse than A B better than C	A is bad B is bad, good C is good	IV′	B not as good as A B not as bad as C	A is good B is good, bad C is bad

gruent information and a solution, he reformulates the question and continues his search. It is not meant by the term "strategy" that the solver has consciously chosen this means of solving the problem. More likely, the strategy is a normal part of processing language and is used without awareness.

The presence of this strategy can be tested by the many possible forms of the three-term series problem, each of which consists of two propositions followed by a question. Traditionally, this allows 16 possible problems. The four pairs of propositions, numbered I–IV in Table 1, can occur in either order, and the question can be either "Who is best?" or "Who is worst?" These traditional problems, however, can be supplemented by replacing *is better than* by *isn't as bad as* and *is worse than* by *isn't as good as.* The resulting 16 problems are shown on the right of Table 1 as Problem Types I′–IV′. The lettering convention used in Table 1 (and Table 2) is that A is always best, C worst, and B in the middle.

The 32 three-term series problems of Table 1 are related to each other in important ways. First, the problems can be matched for the order in which the three terms appear in the surface structure of the sentences (Chomsky, 1965; Postal, 1964). Problem Type I, e.g., is superficially similar to I′: both types arrange the terms A, B, and C in the same order, and both contain relational terms meaning "strictly greater in goodness than." Types II and II′, III and III′, and IV and IV′ are similar in precisely the same way. But the problems can also be matched for deep structure, roughly the underlying functional relations expressed in the propositions. Problem Type I is built on the functional relations *A is good* and *B is good*, as shown in the Analysis column. In this respect, I is like II′, which is built on the same base strings. Furthermore, Types II and I′, III and IV′, and IV and III′ are similar in deep structure. The most important consequence of the proposed strategy is that deep structure takes precedence over order in surface structure in three-term series problems. This point is implied in all the predictions that follow.

Reasoning by the proposed strategy, a problem solver allowed only 10 sec. per problem should make more errors on some problems than on others. Consider first the problems with two *good* or two *bad* propositions: I, II, I′, and II′. Since the storage and retrieval of information from propositions with *good* should be quicker than from ones with *bad,* the problem solver should make fewer errors overall on I and II′, both problems containing *good,* than on II and I′, both problems containing *bad.* The search stage of the strategy also has its effects on these problems. There is congruence when the two propositions in Problems I and II′ are followed by "Who is best?" and when those in II and I′ are followed by "Who is worst?" The problem solver should make few errors on these combinations, but many errors on the remaining incongruent combinations.

The strategy has important consequences for the problems with heterogeneous propositions (III, IV, III′, IV′) as well. Consider Problem Type III. With the question "Who is best?," the problem solver will search for congruent information and come up immediately with the correct answer A; with "Who is worst?," he will likewise find congruent information in the correct answer C. But consider Problem Type IV. With the question "Who is best?," the search for a congruent solution will produce B (of the second proposition), which, on testing, is clearly wrong. Only after "Who is best?" is implicitly reformulated as "Who is least bad?" is the correct solution (A) directly forthcoming. The same difficulty will also occur with the question "Who is worst?" Type III problems, therefore, should encourage few errors since the strategy leads directly to the correct answer, but Type IV problems should encourage many errors. Type IV′ problems are similar to Type III problems in deep structure (note their identity in the Analysis column), and Type III′ problems are similar to Type IV problems. So, for the same reasons, there should be fewer errors on IV′ than III′.

One kind of problem that has not been studied previously is the so-called indeterminate problem. It is called indeterminate because it contains an incomplete rank order of the three people being compared, as in "John is better than Pete, and Dick is better than Pete." Note that when the question for this pair of propositions is "Who is best?," the correct answer is "can't tell," although when it is "Who is worst?," the correct answer is "Pete." Altogether, there are 16 possible indeterminate problems containing *better* or *worse* or both and 16 others containing *isn't as bad as* or *isn't as good as* or both.

The reasoning strategy proposed here has particularly interesting effects on indeterminate problems. In the problem, "If John is better than Pete, and Pete is worse than Dick, then who is best?," a solver will sometimes make an error, choosing "John" or "Dick" rather than the solution "can't tell." But which term will he choose? The strategy describes the solver as searching first for information congruent with the question. His efforts to satisfy the conditions of the question will lead him to answer "John," but not "Dick": "John" satisfies the congruence conditions, whereas "Dick" does not. It is only in indeterminate problems like this one that the strategy produces errors in such a direct way. Here a congruent and an incongruent solution, both incorrect, can be presented in the propositions in a symmetrical form.

Method

Three problems were constructed for each of the 64 possible problem types—the 32 determinate and 32 indeterminate types. In addition, 16 practice problems, 8 determinate and 8 indeterminate, were constructed to have one comparative and one negative equative proposition each. Common four-letter English men's names were used as terms in the problems, with the restriction that no two names could occur together in more than 1 problem. Each problem was printed on an IBM card in the following form:

If Pete isn't as good as Jack,
And Jack isn't as bad as Dick,
Then who is best?
Dick Pete Jack Can't tell

Each *S* solved the 16 practice problems and then 64 other problems, 1 from each of the 64 problem types. Both the practice and experimental problems were in a random and different order for each *S*. There were three groups of *S*s, so that the order of the three names following the question could be counterbalanced both across problems within *S*s and across groups of *S*s within problems; "can't tell" always occurred last among the possible solutions.

Each *S* was given a deck of problems face down. On a signal every 10 sec., he turned a new problem face up, read the problem to himself, circled the answer he thought was correct (or, if he had no answer at the signal, circled nothing), and went on to the next problem. There was a short break after every sixteenth problem. The *S*s, 100 introductory psychology students fulfilling a course requirement, had explained to them beforehand what constituted a "can't tell" answer. They were run in groups of 5–20.

Results

The solutions to both determinate and indeterminate problems contained errors supporting the existence of the reasoning strategy proposed in the present paper. The two kinds of problems are discussed separately. In all that follows, the percentage of errors are compared statistically with Seeger and Gabrielsson's (1968) extension of Cochran's *Q* test; this extension allows comparisons involving repeated observations on the same *S*.

Determinate problems.—The error data in Table 2 first show that *S*s retained the underlying deep structure relations from the propositions of each problem. Among the homogeneous problems, Type I problems produced fewer errors than Type II problems; this closely replicates the results of DeSoto et al. (1965). Now if the order of the terms in surface structure had been responsible for this difference, then Type I′ problems should have produced fewer errors than Type II′ problems, but if deep structure had been responsible, II′ should have produced fewer errors than I′. Clearly, deep structure takes precedence over surface structure, for I and II′ produced fewer errors, respectively, than II and I′, χ^2 (1) = 24.37, $p < .001$. The same result holds for Heterogeneous Problems III, IV, III′, and IV′. Type III problems were easier than Type IV problems, again replicating DeSoto et al.) in parallel with this in deep but not surface structure, Type IV′ problems were

TABLE 2

PERCENTAGE OF ERRORS IN SOLVING DETERMINATE THREE-TERM SERIES PROBLEMS

Problem type	Form of problem	Form of question		Average	Overall average
		Best?	Worst?		
I	(*a*) A better than B; B better than C	15	29	22	17
	(*b*) B better than C; A better than B	7	18	12	
II	(*a*) C worse than B; B worse than A	43	48	46	33
	(*b*) B worse than A; C worse than B	21	21	21	
III	(*a*) A better than B; C worse than B	8	10	9	10
	(*b*) C worse than B; A better than B	10	12	11	
IV	(*a*) B worse than A; B better than C	30	25	28	32
	(*b*) B better than C; B worse than A	32	43	38	
I′	(*a*) A not as bad as B; B not as bad as C	28	21	24	31
	(*b*) B not as bad as C; A not as bad as B	39	36	38	
II′	(*a*) C not as good as B; B not as good as A	14	32	23	26
	(*b*) B not as good as A; C not as good as B	26	31	28	
III′	(*a*) A not as bad as B; C not as good as B	34	45	40	40
	(*b*) C not as good as B; A not as bad as B	34	45	40	
IV′	(*a*) B not as good as A; B not as bad as C	26	30	28	28
	(*b*) B not as bad as C; B not as good as A	23	35	29	

easier than Type III′ problems, χ^2 (1) = 57.30, $p < .001$.

Some of the same results also confirm that unmarked adjectives (like good) are stored and retrieved more quickly than marked adjectives (like bad). Note that the retention of deep structure relations would have been confirmed even if Problem Types I and II′ had been solved with more, rather than fewer, errors than Types II and I′, respectively; the important thing for the confirmation was that I and II′ had the same deep structure relations, and these relations were different from those of II and I′. But since I and II′, containing *good* in their deep structure, *were* solved with fewer errors than II and I′, containing *bad* in their deep structure, this also constitutes evidence for lexical marking; *good* was apparently more quickly comprehended and retrieved than *bad.*

And finally, the data in Table 2 support the proposed search stage of the reasoning strategy. Problem Types III and IV′ are internally congruent: the question always requests information congruent with that specified in the propositions. By the same criterion, Problem Types IV and III′ are internally incongruent. The congruent problems III and IV′ produced significantly fewer errors than the incongruent problems IV and III′. The principle of congruence also finds confirmation in the homogeneous problems. The question "Who is best?" is congruent with the propositions in I and II′, but incongruent with the propositions in II and I′. "Who is best?," rather than "Who is worst?," was answered with fewer errors in I and II′, but with more errors in II and I′, overall χ^2 (1) = 9.04, $p < .01$. In detail, there was one exception to the last result, with "Who is best?" eliciting slightly fewer errors than "Who is worst?" on Problem Type II, a homogeneous *bad* problem. But this is only an apparent exception since "Who is best?," which contains the unmarked adjective *good,* produced fewer incorrect solutions, overall, than "Who is worst?" (24% to 30%). In line with predictions, the difference between *best* and *worst* questions was not as large in Type II problems (2.5%) as it was overall (6%).

It should be noted here that in a previous study (Clark, 1969), *S*s were timed from the beginning of reading to their spoken solutions for each of the 32 determinate problems. The significant differences in errors of the present study are paralleled by significant differences in solution times in the previous study. The problems that produced more errors here had longer solution times previously.

The specific errors *S*s made on these problems further support the existence of the reasoning strategy. The errors can be of three kinds: (*a*) failures to respond, (*b*) "can't tell" answers, or (*c*) incorrect men's names. If *S*s reason as proposed, they should take longer to solve some problems than others. This in turn means that in 10 sec., fewer *S*s will come to any answer whatsoever on the longer problems. And *S*s with no answers will most likely make *a* errors, but also *b* or *c* errors. Indeed, the percentages of *a* errors increased on longer problems. There were 6% such errors on I, compared with 10% on longer II; 10% on II′, compared with 12% on longer I′; 8% on the homogeneous problems (I, II, I′, II′) with congruent questions, compared with 10% on those with incongruent questions; 4% on internally congruent III, compared with 12% on incongruent IV; and 11% on internally congruent IV′, compared with 18% on incongruent III′. The addition of *b* responses merely increases each of these differences, respectively, to 11% to 18%, 16% to 20%, 14% to 18%, 6% to 22%, and 22% to 26%. The further addition of the *c* errors, of course, has been discussed previously.

The *c* errors, incorrect men's names, can indicate the preliminary solutions *S*s tend to choose on some problems or can indicate mere guessing. In determinate problems, the proposed strategy allows only very weak predictions as to what the preliminary solutions will be. In the homogeneous problems (I, II, I′, II′), both of the incorrect men's names should occur, perhaps with the middle B term predominating, because it answers the question when the wrong proposition is considered alone. The results showed here that B occurred more often than the other incorrect term in 12 of the 16 problems, 6% vs. 5% of the time. On the heterogeneous problems (III, IV, III′, IV′), however, B should occur slightly more often than the other term because in III and IV′ B is the only other congruent term and in IV and III′ it is the only congruent term satisfying the question when the wrong proposition is considered alone. The results showed that B was given more often than the other incorrect term on 12 of the 16 problems, 6% vs. 3% of the time.

Indeterminate problems.—Errors in indeterminate three-term series problems also give evidence that *S*s reason by the proposed strategy. Table 3 lists the percentages of errors for each type of problem making use of a new lettering convention in which J is the repeated term and G and H are arbitrarily assigned to the other two terms.

The data here again show that *S*s retain underlying functional relations from the problems they have read: deep structure, rather than order of terms in the surface structure, predicts the differences in these problems. Problem Types V and VI′, which have the same deep structure, were solved with fewer errors than Types VI and V′, respectively, χ^2 $(1) = 4.85$, $p < .05$. The similarity of surface structure between V and V′ and between VI and VI′ had no perceptible effect on their solutions. The heterogeneous problems VII, VIII, VII′, and VIII′ cannot be compared in the same way since whether the question was answerable or not is confounded with surface structure; answerability will be shown subsequently to be very important in the production of errors.

The effect of lexical marking—making *good* more accessible than *bad*—was again found when Problem Types V and VI′, with underlying *good,* were solved significantly more easily than Problem Types VI and V′, with underlying *bad.*

The strategy of searching for congruent information derives its best and most direct support from the problems with heterogeneous propositions (VII, VIII, VII′, VIII′ in Table 3). Consider the problem, "If D is better than E, and E is worse than F, then who is best?," whose correct solution is "can't tell." (In a temporary lettering convention for this kind of problem, the D term is congruent with the question, the F

TABLE 3

PERCENTAGE OF ERRORS IN SOLVING INDETERMINATE THREE-TERM SERIES PROBLEMS

Problem type	Form of problem	Form of question		Average	Overall average
		Best?	Worst?		
V	(*a*) G better than J; H better than J	4[a]	15	10	9
	(*b*) J better than G; J better than H	13	4[a]	8	
VI	(*a*) G worse than J; H worse than J	20	14[a]	17	14
	(*b*) J worse than G; J worse than H	7[a]	14	10	
VII	(*a*) G better than J; J worse than H	49[a]	36	42	42
	(*b*) J worse than H; G better than J	45[a]	37	41	
VIII	(*a*) J better than G; H worse than J	22	44[a]	33	35
	(*b*) H worse than J; J better than G	23	52[a]	38	
V′	(*a*) G not as bad as J; H not as bad as J	15[a]	28	22	24
	(*b*) J not as bad as G; J not as bad as H	28	25[a]	26	
VI′	(*a*) G not as good as J; H not as good as J	18	14[a]	16	19
	(*b*) J not as good as G; J not as good as H	14[a]	29	22	
VII′	(*a*) G not as bad as J; J not as good as H	66[a]	42	54	56
	(*b*) J not as good as H; G not as bad as J	66[a]	49	58	
VIII′	(*a*) J not as bad as G; H not as good as J	35	61[a]	48	45
	(*b*) H not as good as J; J not as bad as G	28	56[a]	42	

[a] *The correct solution is "can't tell" to these problems and "J" to the rest.*

term is incongruent, and the E term is the repeated term.) When *S*s make mistakes on this problem, they should answer the "half correct" D and F more often than E. Moreover, the proposed reasoning strategy predicts that in searching for information congruent with the question, *S*s will choose the D term as their preliminary solution. In agreement with this prediction, the results showed that all eight heterogeneous problems with "can't tell" answers produced considerably more congruent D errors than incongruent F errors ($p < .01$, by a sign test). The D errors occurred, on the average, 26% of the time, E errors 3%, and F errors 9%. The ratio of D to F was almost 3:1; on the four appropriate practice problems, the ratio was 4:1.

For each pair of propositions in Table 3, one question is answerable with a name and the other is unanswerable, i.e., the solution is "can't tell." The difficulties of answerable and unanswerable problems seemed somewhat paradoxical. When the propositions were homogeneous, each of the eight answerable problems produced more errors than its correlative unanswerable problem ($p < .01$, by a sign test). On the other hand, when the propositions were heterogeneous, each of the eight answerable problems produced *fewer* errors than its correlative unanswerable problem ($p < .01$, by a sign test). Both effects, it is argued, constitute evidence for the search for congruence as a stage in the reasoning strategy.

First, in the homogeneous problems (V, VI, V′, VI′), it is easy, in comprehending the two propositions one after the other, to compare the information in them, for the second proposition is exactly congruent with the first except in the first or last term (cf. V or VI in Table 3). By comparing these two almost identical statements, then *S* immediately knows that the three-term series is indeterminate and that the solution might well be "can't tell." If time is running short, therefore, he will give his preliminary solution of "can't tell"; otherwise he will work towards the correct answer. This tendency to answer "can't tell," however, leads to correct solutions on unanswerable problems, whose correct solution *is* "can't tell," but incorrect solutions on answerable problems. The results confirm this prediction. The *S*s made errors on 19% of the answerable problems but on only 12% of the unanswerable ones. Yet they failed to respond equally often (5% of the time) on both. Most of the remaining errors on the answerable problems were the predicted "can't tell" answers, which were produced on 13% of the problems, thus accounting for the greater number of errors in answerable problems than in unanswerable problems.

In the heterogeneous problems (VII, VIII, VII′, VIII′), on the other hand, it is very difficult to compare the information in the two propositions, for the second proposition is completely incongruent with the first. Consider the pair of propositions, "Jack isn't as bad as Pete, and Pete isn't as good as Dick." When the question is "Who is worst?," *S* will search and find a preliminary congruent solution in the "Pete" of the first proposition. That solution also happens to be correct. When the question is "Who is best?," he will search and find a preliminary congruent solution in "Dick." But this solution happens to be incorrect. The search for congruent information, then, leads to correct preliminary solutions on answerable problems but to incorrect ones on unanswerable problems. This prediction from the proposed reasoning strategy is confirmed by the errors. The *S*s made errors on 34% of the answerable problems but on 55% of the unanswerable ones. Nevertheless, they failed to respond equally often (17% of the time) on both. The remaining 17% of the errors for the answerable problems were more or less evenly distributed among the other possible errors. But the remaining 38% of the errors for the unanswerable problems were not: *S*s produced the congruent, but incorrect, answer on fully 26% of the unanswerable problems, as noted previously, thus accounting for the greater number of errors in the unanswerable problems than in the answerable ones.

Order of two propositions.—The present *S*s also appeared to make use of a memory-saving strategy in which the information of each proposition was compressed into a less cumbersome form. Instead of storing all the information of a first proposition, "John is better than Bill," they stored only an abbreviated version which might be glossed as "John is the better one." If this *John* then appeared in the second proposition, e.g., "Pete is better than John," alignment of the three terms was relatively easy, for Pete is better than John who was already the better one. But if this *John* did not appear in the second proposition, then alignment was not immediate. The *S*s had to backtrack for the full version of the first proposition or revert to some other time-consuming strategy.

The strategy of compressing information, in general, makes one ordering of two propositions easier than the other (see Clark, 1969). In Problem Types I and II, the *b* order should be easier than the *a* order since the *b* order allows the compressed version of the first proposition to be used in aligning the terms of the second. This prediction was borne out in the data, χ^2 (1) = 31.02, $p < .001$. But the compressed version of the proposition, "Bill isn't as good as John," is also implicitly "John is the better one." In this case, it is the second term, and not the first, that serves as the term in the compressed version. So in Problem Types I′ and II′, it is the *a* order that should be the easier order. This prediction, too, was borne out in the data, χ^2 (1) = 7.58, $p < .01$. The two orderings of the propositions of III, IV, III′, and IV′, on the other hand, do not differ with the application of the compressed strategy, and, except for IV, the data show only slight differences in errors between the two orders. Note here that the compressed version of a proposition is not directly related to its surface structure: it is neither the subject nor predicate term which serves as the focus of the compressed information, but the term more extreme on the underlying goodness or badness scale.

The compression strategy also affected the error rates in the indeterminate problems. In each of the eight Roman-numeraled pairs of problems listed in Table 3, one problem was favorable to the application of the compression strategy and the other was not. In all eight pairs, the favorable problem produced fewer errors than the unfavorable problem, ($p < .01$, by a sign test), although the differences averaged only 5%.

Discussion

There is good evidence in the present results for all three stages of the proposed strategy for deductive reasoning. First, in comprehending the propositions of a problem, *S*s retained underlying functional relations like *John is bad*. The evidence for this was that the relative number of errors for two reasoning problems was predictable in each case from the deep structure, not the superficial characteristics, of the problem's propositions. Furthermore, deep structure congruence, rather than incongruence, between a question and the information it asked for resulted in fewer errors in a fixed interval of time. Both kinds of results are evidence for the reality of underlying functional relations in reasoning.

Once a person has stored the abstract elements underlying a sentence, he can later retrieve that information. A priori, such a search must have principles to guide it, for it must decide when it has found the object it is seeking. One of the few reasonable assumptions is that the search terminates when it finds information congruent to the information sought. This congruence must be at an abstract level. In three-term series problems, the underlying functional relations of the question have to be congruent with those of the propositions. In this respect, "Who is best?" requires information congruent with the information in "Jack isn't as good as Pete." Several strong tests of this principle in the present experiment were all confirmatory.

One final aspect of the reasoning strategy is that some kinds of information are more easily stored and retrieved than others. The abstract semantic features of unmarked adjectives, like good, are less complex than those of marked adjectives, like bad, and so it was expected that retrieval should reflect this complexity. Four tests of this principle for good and bad all showed that retrieval time did reflect semantic complexity. This evidence alone would not be enough to support the principle of lexical marking, for good–bad is only one such unmarked–marked pair. But previous experiments (e.g., Clark, 1969; DeSoto et al., 1965; Handel, DeSoto, & London, 1968; Hunter, 1957; Huttenlocher, 1968) have required *S*s to solve problems with homogeneous propositions (problems of Types I and II) containing better–worse, more–less, faster–slower, farther–nearer, happier–sadder, warmer–colder, taller–shorter, and deeper–shallower. The problems containing an unmarked comparative, the first member of each pair, were solved more quickly or with fewer errors in a fixed interval of time than the problems containing its marked counterpart. Other evidence for lexical marking is discussed in Clark (1969).

The strategy proposed here, however, predicts more than just the *number* of errors *S*s should make on three-term series problems: in many instances it also predicts *which* errors they should make. Since *S* searches his memory for underlying base strings congruent with those of the question, he should sometimes come to preliminary solutions that fulfill the congruence conditions, but which are nevertheless incorrect. Specific errors confirming these predictions were found, especially in the indeterminate problems. These errors demonstrate that in some problems *S*s try at least one incorrect solution before they come to the correct one and that the ones they try are determined by their search strategy.

The present results are in disagreement with several previous theories of the solution of three-term series problems (DeSoto et al., 1965; Hunter, 1957; Huttenlocher, 1968). According to previous results (Clark, 1969), each of these theories incorrectly predicts the relative solution times of certain three-term series problems, especially those with "isn't as good as" and "isn't as bad as" propositions. The present error results, because they fully parallel these solution times, constitute further disconfirmation of these past theories (see Clark, 1969, for details).

Another older theory (Chapman & Chapman, 1959; Sells, 1936; Woodworth & Sells, 1935) asserts that there is an "atmosphere effect" that accounts for certain difficulties in reasoning. This effect occurs, e.g., in traditional syllogisms that contain two negative premises (or propositions). The negatives are said to induce a "negative atmosphere" in this problem, causing the solver to draw a negative, hence often invalid, conclusion. The present three-term series problems have difficulties that also look like the result of a subtle atmosphere effect. But this effect can only *describe* what happens; by itself, it explains nothing. The present proposal instead attempts to account for these difficulties by general linguistic processes. Neither the "atmosphere effect" nor a similar descriptive statement can be said to explain the difficulties in that sense. On the contrary, the atmosphere effect itself may eventually be explained by linguistic processes similar to the ones proposed here to account for solving three-term series problems.

Finally, the present study appears to put important limitations on general theories of linguistic memory. The emphasis here has been on the fact that what a person retains from a sentence is highly abstract, even though he uses the information immediately. Memory for linguistic material over longer periods of time is also highly abstract in form, as has been shown by Brown and McNeill (1966), Fillenbaum (1966, 1968), Sachs (1967), Clark and Clark (1968), Henley, Noyes, and Deese (1968), Clark and Stafford (1969), and Clark and Card (in press), as well as by those mentioned previously, as confirming the importance of deep structure. It is difficult to see how long-term memory could be otherwise, given the present results. If what a person retains for immediate use is so abstract, he must certainly retain the same kind of information over longer periods of time. The more concrete attributes of a sentence, its words, phrases, or surface structure, appear to be done away with soon after its having been heard, unless there is a special reason to keep them. What remains is an abstract interpretation for use either immediately or later.

REFERENCES

BIERWISCH, M. Some semantic universals of German adjectivals. *Foundations of Language,* 1967, **3**, 1–36.

BROWN, R., & MCNEILL, D. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 325–337.

BURT, C. The development of reasoning in school children. *Journal of Experimental Pedogogy,* 1919, **5**, 68–77 and 121–127.

CHAPMAN, L. J., & CHAPMAN, J. P. Atmosphere effect re-examined. *Journal of Experimental Psychology,* 1959, **58**, 220–226.

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge: M.I.T. Press, 1965.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, **76**, 387–404.

CLARK, H. H., & BEGUN, J. S. The use of syntax in understanding sentences. *British Journal of Psychology,* 1968, **59**, 219–229.

CLARK, H. H., & CARD, S. K. The role of semantics in remembering comparative sentences. *Journal of Experimental Psychology,* 1969, in press.

CLARK, H. H., & CLARK, E. V. Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology,* 1968, **20**, 129–138.

CLARK, H. H., & STAFFORD, R. A. Memory for semantic features in the verb. *Journal of Experimental Psychology,* 1969, **80**, 326–334.

DESOTO, C., LONDON, M., & HANDEL, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology,* 1965, **2**, 513–521.

DOHERTY, P., & SCHWARTZ, A. The syntax of the compared adjective in English. *Language,* 1967, **43**, 903–936.

FILLENBAUM, S. Memory for gist: Some relevant variables. *Language and Speech,* 1966, **9**, 217–227.

FILLENBAUM, S. Sentence similarity determined by a semantic relation: The learning of "converses." *Proceedings of the 76th Annual Convention of the American Psychological Association,* 1968, **3**, 9–10.

GOUGH, P. B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 107–111.

GOUGH, P. B. The verification of sentences: The effects of delay on evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 492–496.

GREENBERG, J. H. *Language universals.* The Hague: Mouton, 1966.

HANDEL, S., DESOTO, C., & LONDON, M. Reasoning and spatial representation. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 351–357.

HENLEY, N., NOYES, H., & DEESE, J. Semantic structure in short-term memory. *Journal of Experimental Psychology,* 1968, **77**, 587–592.

HUDDLESTON, R. D. More on the English comparative. *Journal of Linguistics,* 1967, **3**, 91–102.

HUNTER, I. M. L. The solving of three-term series problems. *British Journal of Psychology,* 1957, **48**, 286–298.

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review,* 1968, **75**, 550–560.

LEES, R. B. Grammatical analysis of the English comparative construction. *Word,* 1961, **17**, 171–185.

LYONS, J. *Structural semantics.* Oxford: Blackwell, 1963.

LYONS, J. *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press, 1968.

MCMAHON, L. E. Grammatical analysis as part of understanding a sentence. Unpublished doctoral dissertation, Harvard University, 1963.

MEHLER, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 346–351.

MILLER, G. A. Some psychological studies of grammar. *American Psychologist,* 1962, **17**, 748–762.

POSTAL, P. Underlying and superficial linguistic structure. *Harvard Educational Review,* 1964, **34**, 246–266.

SACHS, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics,* 1967, **2**, 437–442.

SAPIR, E. Grading: A study in semantics. *Philosophy of Science,* 1944, **11**, 93–116.

SAVIN, H. B., & PERCHONOCK, E. Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 348–353.

SEEGER, P., & GABRIELSSON, A. Applicability of the Cochran *Q* test and *F* test for statistical analysis of dichotomous data for dependent samples. *Psychological Bulletin,* 1968, **69**, 269–277.

SELLS, S. B. The atmosphere effect: An experimental study of reasoning. *Archives of Psychology,* 1936, **16**, 1–83.

SLOBIN, D. I. Grammatical transformations in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 219–227.

SMITH, C. S. A class of complex modifiers. *Language,* 1961, **37**, 342–365.

WOODWORTH, R. S., & SELLS, S. B. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology,* 1935, **18**, 451–460.

(Received January 10, 1969)