

LINGUISTIC PROCESSES IN DEDUCTIVE REASONING¹

HERBERT H. CLARK²

Carnegie-Mellon University

The present paper develops a theory to specify in part how a person stores and searches through information retained from sentences. The theory states that (a) functional relations, like the abstract subject-predicate relation which underlies sentences, are more available from memory than other, less basic kinds of information; (b) certain "positive" adjectives, like *long*, are stored in memory in a less complex and more accessible form than their opposites, like *short*; and (c) listeners can only retrieve, from memory, information which is congruent at a deep level to the information they are searching for. The present theory, unlike previous ones, correctly predicts the principal differences in the solution times of 8 types of two-term series problems and 32 types of three-term series problems (e.g., *If John isn't as bad as Pete, and Dick isn't as good as Pete, then who is worst?*). It also accounts for previous observations on children solving these problems and explains other phenomena in deductive reasoning.

Deductive reasoning has often been studied in particular types of reasoning problems. The strategies suggested for their solution have therefore often been of limited generality: they apply in one kind of problem and that kind alone. The present paper proposes, instead, that reasoning is accomplished mainly through certain very general linguistic processes, the same mental operations that are used regularly in understanding language. Furthermore, the present paper demonstrates in several experiments that these processes, rather than the strategies proposed in the past, correctly account for the difficulties in a variety of reasoning problems.

When a person has comprehended a sentence, he is said to "know what it means." It is this knowledge that is at the heart of the theory developed here. The theory specifies in part both the form this knowledge takes in memory and the process by which

it is later retrieved for other purposes. Knowledge of this kind is presumed to be quite abstract. Thus, to answer a question about the content of a sentence, one must know more than the phonological shape of the sentence: one must have come to an interpretation of it. The distinction here is the same as that in the linguistic concepts of "surface" and "deep" structure (Chomsky, 1965; Postal, 1964). The surface structure of a sentence is the structure which allows it to take on phonological shape; but it is the more abstract deep structure which is necessary for its interpretation. The abstract entities one is presumed to know after interpreting a sentence, then, are closely related to certain linguistic facts about deep structure and the lexicon. And it is at this abstract level that a search for previous information is carried out.

For their study of reasoning, many investigators (Burt, 1919; DeSoto, London, & Handel, 1965; Donaldson, 1963; Handel, DeSoto, & London, 1968; Hunter, 1957; Huttenlocher, 1968) have chosen the so-called three-term series problem, which consists of two propositions and a question, e.g., *If John is better than Dick, and Pete is worse than Dick, then who is best?* The wording of these problems is critical. In all past studies, for example, the above problem has been easier than the following one: *If Dick is worse*

¹The research 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 many colleagues, especially Daryl J. Bem, Stuart K. Card, Eve V. Clark, and William C. Watt, for their constructive comments on the paper, and Joel Gordon and Stephen H. Ellis for their assistance in running the experiments.

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

than John, and Dick is better than Pete, then who is best? The difference occurs even though both problems present exactly the same information, at least superficially. Despite the importance of wording in these problems, however, past accounts of reasoning have neglected to deal directly with the logically prior process of how the language of the problems is itself understood. In one way or another, the past accounts all have the subjects (*Ss*) solving the problems with something less than an abstract interpretation of the propositions. The experiments to be reported here, then, besides lending support to the present theory, also appear to disconfirm the earlier explanations. The contradictory evidence comes mainly from a previously untouched set of three-term series problems in which the customary propositions, like *John is better than Pete*, are replaced by new ones, like *John isn't as bad as Pete*. Although these two propositions have a superficially similar appearance and seem almost synonymous, they have radically different abstract interpretations. Because of this property, they allow strong tests of the previous theories as well as of the present one.

The present theory will be formulated as three principles: two specify what it is that the listener knows of a sentence he has heard and a third specifies how he searches his memory for the wanted knowledge. These three principles will then be used as a basis for predicting the relative times it takes *Ss* to solve two-term series problems (e.g., *If John is better than Pete, then who is worse?*) and three-term series problems. Finally, the theory will be applied to previous data on three-term series problems as well as to other, less directly related phenomena in deductive reasoning.

THE THREE PRINCIPLES

Principle of the Primacy of Functional Relations

Functional relations are the primitive conceptual relations out of which sentences are constructed. Chomsky (1965) lists four such relations which he claims are universal: Subject-of, Predicate-of, Direct-object-of,

and Main-verb-of. For example, in both *John watched the monkey* and *The monkey was watched by John*, a listener knows that *John*, *watch*, and *monkey* are in the relation subject, verb, and direct object: it was John who watched, what John did was watch, and it was the monkey which was watched. But the listener also knows that the theme of the first sentence—what the sentence is about (Halliday, 1967)—is John, whereas the theme of the second is the monkey; this information, of a quite different sort, is not to be found in the functional relations that underlie a sentence. The principle of the primacy of functional relations asserts simply that functional relations, like those of subject, verb, and direct object, are stored, immediately after comprehension, in a more readily available form than other kinds of information, like that of theme.

This principle, first proposed by Miller (1962) in the language of an earlier linguistic theory, is formally related to two different kinds of information present in the deep structure of a sentence. In one generative grammar of English (similar to Chomsky's, 1965), deep structure, a product generated by the rules of the "base component" of the grammar, consists of (*a*) so-called base strings, like *John Past watch the monkey*, which fully specify functional relations, and (*b*) directions for the eventual transformation of these strings into a surface structure, like the passive *The monkey be + Past watch + en by John*. The principle proposed here specifies that the information of *a* is more available than the information of *b*. Significantly, there is only a small number of functional relations and possible base strings—linguists differ as to the actual number—but there is a great number of transformations available to shape them into numerous different surface forms.

Miller's (1962) proposal was that a sentence is understood and stored in memory as two kinds of independent information—base strings and transformations. In subsequent research, Miller and others (Gough, 1965, 1966; McMahan, 1963; Mehler, 1963; Miller & McKean, 1964; Savin & Percho-nock, 1965) have concentrated on demonstrating that a sentence requiring more trans-

formations in mapping deep onto surface structure is more complex psychologically, taking more time to understand and more space in immediate memory. The generality of this relationship has recently been questioned by Fodor and Garrett (1966) (cf., however, Watt, in press). In contrast, the emphasis in the present paper is on the importance of the base strings themselves: they constitute the essential part of the interpretation of a sentence and should therefore play an important part whenever the interpretation is needed at a later time. Evidence for their importance is found especially in the experiments of McMahan (1963), Clifton, Kurcz, and Jenkins (1965), Clifton and Odom (1966), and Clark and Begun (1968).

Principle of Lexical Marking

According to the principle of lexical marking, the senses of certain "positive" adjectives, like *good* and *long*, are stored in memory in a less complex form than the senses of their opposites. This principle is derived from certain linguistic facts relevant to the lexical component of English, that part of the grammar which defines the senses of words; these words, when inserted in the base component, give phonological shape to the abstract characterizations of the base. It is the lexicon that specifies that *bird* is superordinate to *oriole*, that *man* is animate, that *good* and *bad* are antonymous, and so on.

Antonymous adjectives, like *good* and *bad*, and *long* and *short*, are often found, on close scrutiny, to be asymmetric (Bierwisch, 1967; Greenberg, 1966; Lyons, 1963, 1968; Sapir, 1944; Vendler, 1968). The first piece of evidence for this is that the "positive" member of many such pairs can be neutralized in certain contexts. A speaker asking "How good is the food?" can merely be asking for an evaluation of the food. He will be satisfied whether he is told the food is good or bad. But the speaker asking "How bad is the food?" is implying something more: rightly or wrongly, he is pronouncing the food to be bad and is asking about the extent of its badness. Since *good* can be

neutralized and *bad* cannot, *good* is said to be "unmarked" and *bad* "marked." Other unmarked-marked pairs by the same criterion are *long-short*, *wide-narrow*, and *interesting-uninteresting*; in the last pair, the marking is made explicit morphologically.

The second piece of evidence, obviously related to the first, is that the unmarked member of each pair also serves as the name of the full scale. The names of the *good-bad* and *long-short* scales are *goodness* and *length*; *badness* and *shortness* name only half of their respective scales. Also, in sentences like *The board is six feet long*, *long* names the dimension to be measured and nothing more; *six feet long* is exactly paraphrased by *six feet in length*. The *long* in *six feet long* obviously is in the same class with other dimensional names, like *wide*, *deep*, *thick*, and *high*—other unmarked adjectives—and not with its opposite *short*: the sentence **The board is six feet short*³ is unacceptable to English speakers.

In sum, the unmarked adjective has two senses, but the marked only one. The sense of *good* in the noncommittal *how good?* and that of *long* in the noncommittal *how long?* or in *six feet long* will be called their "nominal" senses, for in these instances only the scale name is indicated. The other senses of *good*, *bad*, *long*, and *short*, as in the simple *This board is long*, will be called "contrastive." Contrastive *long* and *short*, for example, specify length in relation to some implicit standard and contrast with each other. One property of nominal and contrastive senses is that they can occur together, as in *The short board is six feet long*, although two different contrastive senses (without qualification) cannot, as in the unacceptable **The short board is long*. Another property is that the nominal sense is semantically presupposed in the contrastive senses. Note that contrastive *long* and *short* mean "much in length" and "little in length"; both definitions presume that the measured dimension is length—both presume nominal *long*. Yet the nominal sense of *long* carries no presupposition about comparative length.

³ An asterisk indicates that the following sentence is generally unacceptable in English.

This linguistic evidence suggests that nominal and contrastive senses of adjectives ought to have different codings in memory (cf. also Clark & Stafford, 1969). The coding difference is formalized in the principle of lexical marking: (a) a nominal sense is stored as one less entity than its corresponding contrastive senses, hence (b) the nominal sense is stored and retrieved from memory more quickly than the contrastive senses. To indicate this in a notation (see Bierwisch, 1967), nominal *good* would be coded as [+*Evaluative*[*Polar*]], and contrastive *good* as [+*Evaluative*[+*Polar*]]. The semantic feature [+*Evaluative*] means that an evaluation is being made in both cases. The lack of a sign before the feature *Polar* in the nominal case means that the end of the scale is not specified; the plus sign in the contrastive case means that the positive pole of the scale is specified. The only sense of *bad*, similarly, is coded as [+*Evaluative*[-*Polar*]]. Stated in terms of this notation, the principle states that [+*Evaluative*[*Polar*]] alone is easier to store and retrieve than the more complex [+*Evaluative*[-*Polar*]].

The principle of lexical marking has indirect support from several previous studies. Greenberg (1966) and Marshall (1968) proposed that the extra specification of the marked adjective should be more easily dropped than added in free association; the data they examine support this proposal. Clark and Card (1969) have shown a similar loss of the additional feature in the memory for comparative sentences containing antonymous adjectives. Both kinds of results are consistent with the greater simplicity of the memory coding for unmarked adjectives in their nominal sense.

Principle of Congruence

Answering a question requires more of a listener than a mere understanding of the question itself. He must "search" his memory for the wanted information and formulate that information in an answer. It is proposed here that his search is guided by the principle of congruence. What he seeks from his previous knowledge is information congruent, at the level of functional rela-

tions, with the information asked for in the question. He cannot answer the question until he finds congruent information, or until he reformulates the question so that he is able to do so.

Application of the Principles to Comparative Sentences

The role these three principles play in two- and three-term series problems depends first on the role they play in the sentences that make up these problems. The sentences of interest are comparative constructions, such as *John is better than Pete*, negative equative constructions, such as *John isn't as bad as Pete*, and questions such as *who is best?* The three principles will be examined in turn for their application to these types of sentences.

Principle of the primacy of functional relations. By this principle, the information most readily available from an interpretation of a sentence is its underlying functional relations. But what are the functional relations in comparative and negative equative constructions? Lees (1961), Smith (1961), Huddleston (1967), and Doherty and Schwartz (1967) argue that both types of constructions are generated linguistically from two primitive base strings. Underlying *John is better than Dick* are the base strings, *John is good* and *Dick is good*. It is here in the base strings that the subject-predicate relations of *John* and *good* and of *Dick* and *good* are found. The two strings are also designated in the base component as parts of a comparative construction. After the first transformations, there is a result that reads, roughly, *John is more good than Dick is good*. By a further transformation the last redundant *good* is dropped; and finally, *more good* is changed to *better* to form *John is better than Dick is* or, simpler still, *John is better than Dick*. Had the two base strings been *the door is wide* and *the desk is long*, the resulting sentence would have been *The door is wider than the desk is long*; in this case, the second adjective could not be deleted, since it was not identical to the first. The positive equative construction *John is as good as Dick*, as well as the negative *John isn't as good as Dick*, are derived

analogously, with *as-as* in place of *more-than*.

Comparative sentences then contain, roughly, two kinds of information: (a) the functional relations, as in *John is bad*, and (b) the comparison *more than*. Applied to the comparative, the principle of the primacy of functional relations asserts that *a* is more available than *b*. In *John is worse than Pete*, the listener realizes that John and Pete are bad more readily than that John is more extreme than Pete in badness. To emphasize the functional relations in a convenient notation, the present paper will represent *John is worse than Pete* as (*John is bad+*; *Pete is bad*), *John isn't as bad as Pete* as (*John is bad*; *Pete is bad+*), and *who is worst?* as (*X is bad++*). One + is one degree more extreme than none, and two +'s indicate the most extreme.

An important, but surprising, consequence of this linguistic analysis is that *John is better than Pete* and *Pete is worse than John* do not impart the same information, as a logician might suggest. The comprehension of each sentence brings with it different kinds of immediate knowledge. This difference will be shown to affect how well someone can use the knowledge in deductive reasoning.

Principle of lexical marking. This principle asserts that the nominal sense of *good*—the sense found in noncommittal *how good?* questions—is stored in a less complex and more available form than the contrastive senses of *good* and *bad*. But in which sense are we to interpret the *goods* underlying *John is better than Pete* and the *bads* underlying *Pete is worse than John*? It will be argued that the *goods* can be either nominal or contrastive, but the *bads* must be contrastive. Evidence for this is found in three kinds of examples which show a parallel between comparative sentences and *how good?* type questions, whose properties have already been demonstrated.

First, both *How good is John?* and *John is better than Pete* are ambiguous. If we know nothing about John or Pete, both sentences would normally be interpreted as implying nothing about John or Pete. But when prefaced by *I know that John and*

Pete are good but, they are no longer non-committal and therefore contain the contrastive *good*. Note that *How good is John?*, with a stress on *how*, is understood in its contrastive sense. *How bad is John?* and *Pete is worse than John*, however, unambiguously imply negative evaluations of Pete and John. Second-order comparative questions show the same neutralization phenomenon. *How much better is John than Pete?* can be noncommittal or positive, but *How much worse is John than Pete?* is always negative. Pertinent to this point is DeSoto et al.'s (1965) anecdote about the disgruntled baseball fan who was watching two baseball greats playing a bad game. "I came to see which of you two guys was better," he yelled at one of them. "Instead, I'm seeing which is worse."

Second, unmarked adjectives can be used in ways that marked adjectives cannot. Consider the following examples, remembering that *intelligent* is unmarked and *stupid* is marked:

1. This genius is more intelligent than that genius.
2. This idiot is more intelligent than that idiot.
3. *This genius is stupider than that genius.
4. This idiot is stupider than that idiot.

If it is assumed that *stupid* and *intelligent* are being used in only their contrastive senses, then Sentences 2 and 3 should both be unacceptable sentences. The reason is that their respective base strings, **idiots are intelligent* and **geniuses are stupid*, are unacceptable. However, Sentence 2 is clearly acceptable, although Sentence 3 is not (except in an ironic sense). The *intelligent* underlying Sentence 2 must therefore be nominal, so that the base string *this idiot is intelligent* really only means "this idiot has measurable intelligence." Sentence 3 remains unacceptable because the marked *stupid* cannot assume a nominal sense. There are four parallel examples of the *how intelligent?* type questions, and of these, the only unacceptable one—**How stupid is the genius?*—is the counterpart of Sentence 3.

Third, the parallel between *how good?* type questions and comparative sentences is more than superficial, for they both have much the same structure. *How good is John?* can be written as *John is (how) good*, and *John is better than Pete* as *John is (more in extent than Pete is good) good* (cf. Chomsky, 1965; Doherty & Schwartz, 1967). The comparative is a type of quantification on the unadorned adjective, just as *six feet* is on simple *long*. The parallel in the semantic properties of the two is therefore not surprising. In summary, unmarked adjectives in comparatives, just as in *how good?* type questions, can be interpreted either nominally or contrastively, whereas their marked counterparts can only be interpreted contrastively.

The principle of lexical marking, then, has direct application to sentences that contain *better*, *worse*, *isn't as good as*, or *isn't as bad as*. From the above, we know that the *good* underlying *John is better than Pete* can be interpreted either nominally or contrastively. From the principle of lexical marking, however, it follows that since the contrastive sense takes longer to store and retrieve, this *good* will usually be interpreted in its simpler nominal sense. The *bad* underlying *Dick is worse than Jack*, of course, can only be interpreted contrastively. This agrees with intuition, leaving *John is better than Pete* normally noncommittal in tone, but *Dick is worse than Jack* clearly negative in tone. The principle predicts that *better* and *isn't as good as* propositions will be more quickly registered and retrieved than *worse* and *isn't as bad as* propositions.

Principle of congruence. By this principle, information cannot be retrieved from a sentence unless it is congruent in its functional relations with the information that is being sought. This can be illustrated in the following two-term series problem: *If John is better than Pete, then who is best?* (Although this question is "bad English" by grammar school standards, it was used in the experiments that follow because *good*, *better*, and *best*, and *bad*, *worse*, and *worst*, all have different phonological forms and because the three-term series problem uses the same form of questions; as expected, no

S objected to it.) The proposition provides the information, (*John is good+*; *Pete is good*). The question requests an *X* so that (*X is good++*); that is, it requires a search for the term with the most-plussed *good*. The underlying form of the question is congruent with that of the proposition, so the solution is immediately forthcoming—"John is best" or just "John." But when *who is best?* is replaced with *who is worst?*, a question requesting information not congruent with the proposition's information, then the problem solver will search for the most-plussed bad term and, finding none, implicitly reformulate the question to read *who is least good?* So (*X is bad++*) becomes (*X is good--*), in which the minuses direct the search for the term with the least-plussed *good*. In this search, the *S* will find congruent information and will formulate the solution "Pete is worst" or "Pete." The principle of congruence implies, then, that retrieving an answer should take less time when propositions and questions are congruent in their base strings than when they are incongruent.

TWO- AND THREE-TERM SERIES PROBLEMS

Implicit in the previous three principles is a process by which people solve problems in deductive reasoning. Its identifiable stages are (a) comprehension of the propositions; (b) comprehension of the question; (c) search for information asked for in the question; and (d) construction of an answer. The three principles affect the outcome of this process at one or more of its stages. It is convenient, then, to examine the application of the principles to the process as it is supposed to occur in two-term series problems. These problems consist of eight types, the ones formed when one of four simple propositions (*A is better than B*, *B is worse than A*, *A isn't as bad as B*, and *B isn't as good as A*) is each followed by one of two questions (*who is best?* and *who is worst?*).

The first two stages—comprehending the proposition and question—entail setting up a representation like (*A is good+*; *B is good*) and (*X is good++*), by the principle of the primacy of functional relations.

At these stages and later on, the principle of lexical marking predicts that the base-string pair (*A is good+*; *B is good*) should be more quickly registered and retrieved than (*B is bad+*; *A is bad*), since the memory coding for *bad* is more complex than that for *good*. At the third stage, that of searching for information asked for in the question, *S* carries out the instructions implicit in the question. It is at this stage that the principle of congruence comes into play. Whenever the question is congruent with the proposition, *S* should take little time; if he needs to reinterpret the question to make it congruent, he will take more time.

Experiment I: The Two-Term Series Problem

The four propositions of the two-term series problems, shown in Table 1, can be matched on superficial or deep structure. Proposition I, *A is better than B*, has the same order of terms in surface structure as I', *A isn't as bad as B*. In both propositions, *A* is the subject, *B* is the term in the predicate, and the relation between the terms means "strictly greater in goodness than." But Proposition I does not have the same deep-structure as I'. Proposition I is generated from base strings containing *good*, as indicated in the "Analysis" column, whereas I' is generated from base strings containing *bad*. In deep structure, Proposition I is like II', *B isn't as good as A*. The four propositions, then, allow an orthogonal comparison of order in surface structure and of deep structure: pairs I and I', and II and II', have the same order in surface structure; pairs I and II', and II and I', are similar in deep structure. By the three principles, it is claimed that the solution times of the eight problem types should be affected mainly by the proposition's—and the question's—deep structure.

Method. Four examples of each of the eight problem types were constructed using as terms common English four-letter men's names, no pair of which occurred together in more than one problem. Each of the 32 problems was typed in one continuous line on the middle of a blank IBM card in the following form: If Pete isn't as bad as John, then who is best? The problems were arranged in

TABLE 1
MEAN TIME TO SOLVE TWO-TERM SERIES PROBLEMS

Form of problem	Analysis	Form of question		<i>M</i>
		Best?	Worst?	
I <i>A better than B</i>	<i>A is good</i>	.61	.68	.64
	<i>B is good</i>			
II <i>B worse than A</i>	<i>A is bad</i>	1.00	.62	.81
	<i>B is bad</i>			
I' <i>A not as bad as B</i>	<i>A is bad</i>	1.73	1.58	1.66
	<i>B is bad</i>			
II' <i>B not as good as A</i>	<i>A is good</i>	1.17	1.47	1.32
	<i>B is good</i>			

Note.—Mean time is in seconds.

four blocks of eight, each block containing one problem of each type. The order within each block was random and different for each *S*. The first block was considered practice and was later discarded.

The *S*, at a signal, turned over a card, read the problem aloud, and gave an answer as quickly as he could consistent with high accuracy. He was timed from the first signal to his answer in hundredths of a second. After attempting all 32 problems in this manner, he repeated the procedure, omitting the answer. The time duration on the first go-round, minus the time duration on the second, was taken as the solution time for each of the 32 problems; this procedure was meant to correct for possible differences in the reading times of the problems.

For each *S*, the problem of each type with the longest solution time was taken out and the analysis was done on the two remaining ones; this procedure eliminates the effect of the occasional very long solution time resulting from wandering attention or other extraneous factors. Problems with incorrect answers (6% of the nonpractice problems) were also discarded before analysis. The 20 *S*s were students from an introductory psychology course fulfilling a course requirement.

Results. The mean solution times for the two-term series problems are shown in Table 1; arithmetic means were used in place of the more generally used geometric means, because it was possible here to have null or negative solution times. The predictions

made by the foregoing theory are clearly confirmed by these solution times.

First, the principle of the primacy of functional relations predicts that solution times will correlate with underlying, rather than superficial, structure. It was found that Problem Type I took less time than II. If this difference had been the result of the superficial order and meaning of the terms, then I' should take less time than II'; but if it had been the result of their different underlying base strings, then II' should take less time than I'. The data clearly support the second interpretation: I and II' had significantly shorter solution times than II and I', $F = 8.79$, $df = 1/19$, $p < .01$, and there was no significant interaction.

The principle of lexical marking predicts that problems with underlying *good* will take less time than those with underlying *bad*. This is confirmed by the same evidence as above. First, note that the principle of primacy of functional relations would also have been supported if I and II' had had *longer* solution times than II and I', respectively; support for this principle requires only that the two problems with similar deep structure be consistently different in the same direction from the other two problems. But the results are quite specific: overall, the *good* problems, I and II', took significantly less time than the *bad* ones, II and I'. This supports lexical marking.

Finally, the principle of congruence predicts that questions congruent with a proposition in their underlying base strings will be answered more quickly than incongruent questions. This was supported, $F = 11.32$, $df = 1/19$, $p < .005$, with no other significant interactions. The question *who is best?*, rather than *who is worst?*, had the shorter solution times for Problem Types I and II', built on an underlying *good*, but the longer solution times for Problem Types II and I', built on an underlying *bad*.

Deep structure was clearly dominant over the order of terms in surface structure. In Types I and II, the subject term of the proposition was more quickly retrieved than the term in the predicate. But this was not true for Types I' and II'. For them, the

terms in the predicate were more quickly retrieved. Order in surface structure, therefore, is of no detectable importance in these problems.

A final result is that the problems with comparative propositions were more quickly solved than those with negative equative propositions, $F = 18.65$, $df = 1/19$, $p < .001$, with no other significant interactions. There is little doubt that the negative equative is syntactically more complex than the comparative; in current versions of transformational theory, there is at least one more transformation—the negative—needed in generating the negative equative construction. Conceptually, *B isn't as good as A* is the denial of *B is as good as A*, a construction of the same level of complexity as *B is better than A*. As one more piece of semantic information, denial itself takes time to process (Gough, 1965, 1966; Wason, 1961).

Experiment II: The Three-Term Series Problem

Two-term series problems are, for the most part, trivial to solve. But, as it will be seen, the principles which explain the difficulties of these problems also explain most difficulties of the three-term problems. There is one additional assumption needed to account for the further difficulties of storing the information in three-term series problems.

Three-term series problems consist of two propositions and a question, as in *If John is better than Pete, and John is worse than Dick, then who is worst?* The three terms (*John*, *Pete*, and *Dick*) can be placed in the same evaluative order in 16 different problem types which use *better*, *worse*, *best*, and *worst*. As shown in Table 2, there are four basic pairs of propositions, labeled I through IV. Completing the 16 types, the two propositions of each pair can occur in either order, and the question can be either *who is best?* or *who is worst?* In addition, 16 more problem types are possible when *isn't as bad as* is substituted for *is better than*, and *isn't as good as* for *is worse than*. These are the four pairs of propositions labeled I' through IV' listed on the right

TABLE 2
TYPES OF THREE-TERM SERIES PROBLEMS

Form of problem		Analysis	Form of problem		Analysis
I	<i>A</i> better than <i>B</i>	<i>A</i> is good <i>B</i> is good <i>C</i> is good	I'	<i>A</i> not as bad as <i>B</i>	<i>A</i> is bad <i>B</i> is bad <i>C</i> is bad
	<i>B</i> better than <i>C</i>			<i>B</i> not as bad as <i>C</i>	
II	<i>C</i> worse than <i>B</i>	<i>A</i> is bad <i>B</i> is bad <i>C</i> is bad	II'	<i>C</i> not as good as <i>B</i>	<i>A</i> is good <i>B</i> is good <i>C</i> is good
	<i>B</i> worse than <i>A</i>			<i>B</i> not as good as <i>A</i>	
III	<i>A</i> better than <i>B</i>	<i>A</i> is good <i>B</i> is good, bad <i>C</i> is bad	III'	<i>A</i> not as bad as <i>B</i>	<i>A</i> is bad <i>B</i> is bad, good <i>C</i> is good
	<i>C</i> worse than <i>B</i>			<i>C</i> not as good as <i>B</i>	
IV	<i>B</i> worse than <i>A</i>	<i>A</i> is bad <i>B</i> is bad, good <i>C</i> is good	IV'	<i>B</i> not as good as <i>A</i>	<i>A</i> is good <i>B</i> is good, bad <i>C</i> is bad
	<i>B</i> better than <i>C</i>			<i>B</i> not as bad as <i>C</i>	

of Table 2. Although Problem Types I through IV have been studied before, I' through IV' have not. For ease of comprehension the convention is used that the *A* term is best, the *C* term worst, and the *B* term in the middle.

Like the problems in Table 1, those in Table 2 can be paired for the similarity of either superficial order or deep structure. The Roman numerals match the problem types for superficial similarity. Types I and I', for example, are identical except for the relational terms, and in both cases the relation term (*is better than* and *isn't as bad as*, respectively) means "strictly greater than." The deep structure of each pair, however, is different. Underlying Problem Type I is the adjective *good*, and underlying I' is *bad*. In their underlying base strings (shown in the "Analysis" column), I and II' are alike, as are II and I', III and IV', and IV and III'.

The three principles afford a number of predictions about the problems with homogeneous propositions (Types I, II, I', and II'). The problems with *good* relational terms (I and II') should, as in the two-term series problems, be easier than those with *bad* (II and I'). Also, a question which is congruent with the information in the propositions should be easier than an incongruent question: *who is best?* should

be the easier question for Types I and II', and *who is worst?*, for Types II and I'.

The predictions for Problem Types III, IV, III', and IV'—those with heterogeneous propositions—are slightly more involved. Consider Type III problems, which contain the underlying base strings *A is good*, *B is good*, *B is bad*, and *C is bad*. The answer to *who is best?* is *A*, a term which belongs to a base string congruent with that of the question (*X is good++*). The answer to *who is worst?* is *C*, which also fulfills the congruence conditions. Type IV problems, on the other hand, show complete incongruence of the propositions and questions. *A*, the answer to *who is best?*, is part of the base string *A is bad*, and *C*, the answer to *who is worst?*, is part of *C is good*. Because of this internal disagreement, Type IV problems should be harder than Type III problems. In their deep structure, III' is like IV, and IV' is like III, so, for the same reasons, the internally incongruent Type III' problems should be harder than the internally congruent Type IV' problems.

Method. Three problems were constructed for each of the 32 problem types indicated in Table 3. Common four-letter English men's names were used, such that no pair of them would occur together in more than one problem. Each problem

was typed on a blank IBM card in the following form:

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

In addition, there were eight practice problems each containing one comparative and one negative equative proposition. The problems were arranged in four blocks: the practice problems and then three blocks of 32, each latter block consisting of one of each problem type. Within blocks, the problems were random and different for each *S*. The order of the names following the question was counter-balanced across the last three blocks and across problem types.

On a signal, *S* turned a problem card face up, read the problem silently to himself, and produced an answer as quickly as he could without sacrificing accuracy. He was timed from the initial signal to the answer in hundredths of a second. He solved the 104 problems with short breaks between each. Unlike Experiment I, the solution time was taken as the time duration from the signal to the answer; *S*s found that reading aloud was very disruptive on this complex a problem. In the following results, then, reading time is confounded with solution time. If anything, this would militate against the predicted results in one case, for *better* is one syllable longer than *worse*. Again, the longest solution time was discarded for each problem type for each *S*, and so were the errors (7% of the answers). The *S*s were 13 students fulfilling a course requirement for introductory psychology.

Results. The geometric mean solution times in Table 3 confirm each of the predictions of the present theory. The results that follow, however, were further substantiated in a subsequent variation on this experiment (Clark, 1969b), in which 100 *S*s each attempted to solve the 32 problems given here, along with 32 indeterminate problems, in 10 seconds each. The number of *S*s failing to solve each problem within 10 seconds closely parallels the solution times in Table 3: more *S*s made errors on those problems which in the present experiment took more time to solve. The significance level in Clark (1969b) will therefore be given in brackets for each difference that follows.

The principles of the primacy of functional relations and of lexical marking predict that Problem Types I and II' will be solved more quickly, overall, than II and I', respectively. The solution times in Table 3 confirm this

prediction, $F = 5.38$, $df = 1/12$, $p < .05$, [$p < .001$], with no significant interaction. (The analysis of variance was carried out on the logarithms of the solution times.) The principles of the primacy of functional relations and of congruence predict that III and IV' will be solved more quickly, overall, than IV and III', respectively. The prediction was confirmed here too, $F = 4.92$, $df = 1/12$, $p < .05$, [$p < .001$], with no significant interaction. These two principles also predict that solutions will be faster for Problem Types I and II' when the question is *who is best?* and for Problem Types II and I' when the question is *who is worst?* The results support this prediction, $F = 9.73$, $df = 1/52$, $p < .005$, [$p < .01$], with no other significant interactions. Comparative problems (I through IV) had shorter solution times than negative equative problems (I' through IV'), $F = 25.06$, $df = 1/12$, $p < .001$, [$p < .001$]. Again, all results show the relative importance of deep structure over the order of terms in surface structure.

Order of the Two Propositions

An additional phenomenon to be explained is the fact that two homogeneous propositions in one order are easier than the same two propositions in the opposite order. Consider Problem Types Ia and Ib in Table 3. In Ia the terms are arranged in a linear order. Reading from left to right, one finds *A, B, B, C*. In Ib the arrangement is nonlinear—*B, C, A, B*. In Types I and II, problems with comparative propositions, the nonlinear order was easier, $F = 17.56$, $df = 1/52$, $p < .001$, [$p < .001$], but in the negative equative Problems I' and II', it was the linear order that was easier, $F = 4.04$, $df = 1/52$, $p < .05$, [$p < .01$]. Although this phenomenon remains unexplained by the three proposed principles, it can be accounted for by a generalization of some observations Donaldson (1963) made of children solving three-term series problems containing comparative constructions. Her suggestion for certain difficulties that the children had has merely been generalized here, at the level of deep structure, to negative equative propositions as well.

TABLE 3
GEOMETRIC MEAN TIMES IN SOLVING THREE-TERM SERIES PROBLEMS

Form of problem		Form of question		M	Overall M
		Best?	Worst?		
I	(a) A better than B; B better than C	5.42	6.10	5.75	5.49
	(b) B better than C; A better than B	4.98	5.52	5.25	
II	(a) C worse than B; B worse than A	6.27	6.53	6.40	5.91
	(b) B worse than A; C worse than B	5.93	5.04	5.47	
III	(a) A better than B; C worse than B	5.35	5.34	5.34	5.33
	(b) C worse than B; A better than B	4.84	5.84	5.32	
IV	(a) B worse than A; B better than C	5.00	6.02	5.49	5.63
	(b) B better than C; B worse than A	6.12	5.45	5.77	
I'	(a) A not as bad as B; B not as bad as C	6.77	5.95	6.34	6.59
	(b) B not as bad as C; A not as bad as B	7.16	6.56	6.85	
II'	(a) C not as good as B; B not as good as A	5.58	6.63	6.08	6.22
	(b) B not as good as A; C not as good as B	6.11	6.60	6.35	
III'	(a) A not as bad as B; C not as good as B	6.34	6.66	6.50	6.52
	(b) C not as good as B; A not as bad as B	6.73	6.34	6.53	
IV'	(a) B not as good as A; B not as bad as C	6.10	6.18	6.14	6.19
	(b) B not as bad as C; B not as good as A	5.48	7.12	6.25	

Note.—Mean times are in seconds.

It is assumed that Ss try to compress the information of the propositions so that the information is easier to handle in memory. The proposition *John is better than Bill* would be stored, not as the full (*John is good+*; *Bill is good*), but as the compressed (*John is good+*), meaning "John is the better one." So when the second proposition is *Dick is better than John*, the three-term series is easily constructed: since Dick is better than John, who is already the better one, then Dick is best, John next, and the other one least good. But when the second proposition is *Bill is better than Walt*, there is no obvious series, for the term carried

along—"John is the better one"—does not even appear in the second proposition. In this case, S must try to recover the whole first proposition or back-track with the information that "Bill is the better one" to apply it to the first proposition or revert to some other time-consuming strategy. By this analysis, the nonlinear *Ib* and *I Ib* should be easier, respectively, than the linear *Ia* and *IIa*. The data bear out this expectation. In exactly the same way, *Bill isn't as good as John* should be compressed as (*John is good+*), which has the meaning "John is the better one." In such sentences, it is the second term, not the first, which is at the

heart of the compressed version. So, in contrast with the comparative problems, the linear I'a and II'a should be easier, respectively, than the nonlinear I'b and II'b. The data bear out the expectation here also.

Comparatives Other than Better and Worse

Good and *bad* are only two of the many possible adjectives available for three-term series problems. Previous experiments on such problems have included *better-worse*, *lighter-darker* (DeSoto et al., 1965), *more-less*, *faster-slower*, *farther-nearer*, *earlier-later* (Handel et al., 1968), *happier-sadder*, *warmer-colder* (Hunter, 1957), *taller-shorter* (Hunter; Huttenlocher, 1968), and *deeper-shallower* (see below). Almost all pairs showed asymmetries in difficulty when used in homogeneous problems of Types I and II; difficulty for Hunter and Huttenlocher was measured by solution time, and for DeSoto et al. and Handel et al., by proportion of errors made in a fixed interval of time. Also, in an unpublished pilot study,⁴ seven Ss made 128 judgments each of the truth or falsity of statements like *18 isn't as young an age as 23*. These judgments were significantly faster for the relations *older than* and *isn't as old as* than for their opposites ($p < .025$), and for *higher than* and *isn't as high as* than for their opposites ($p < .025$).

The asymmetries in these pairs of adjectives are accounted for by the principle of lexical marking. Problems with the comparatives of *good*, *much*, *fast*, *far*, *tall*, *happy*, *warm*, *deep*, *old*, and *high* were solved more easily than those with their opposites. Each of these adjectives is semantically unmarked, so their ease in comprehension is explained by their simpler semantic code in memory. However, problems with *earlier* were easier than those with *later*, a fact which remains unexplained by this principle, since both *early* and *late* are marked by the criteria of neutralization and nominalization used in the present theory. In this particular case, there appears to be another type of markedness operating, but one which will be pur-

sued no further here. On the other hand, there was one pair—*lighter-darker*—for which no asymmetry could be detected. But since neither *light* nor *dark* (when used for the color of hair) neutralizes or doubles as the scale name, they are both lexically marked. The principle of lexical marking is correct, therefore, in refusing to predict an asymmetry here.

By the analysis proposed previously, the marking of adjectives is irrelevant for the differences between Problem Types III and IV (Table 2 or 3). The other two principles, however, predict III to be easier than IV, regardless of the antonyms used. The data for all 10 previously studied pairs of comparatives conform to this prediction.

Solution of Three-Term Series Problems by Children

Previous investigators interested in the development of reasoning have studied the solution of three-term series problems by children. Their careful observations, though usually more informal in nature, lend considerable support to the present theory.

Burt (1919) originally, and later Piaget (1921, 1928), Hunter (1957), Donaldson (1963), and Luria (1966) have all noticed what Piaget (1921) called "judgment of membership" in the child's interpretation of comparative sentences. Piaget (1921), for example, reports that 9- and 10-year-olds were unsuccessful in solving the following Type IV problem (from Burt, 1919): "Edith is fairer than Suzanne; Edith is darker than Lili. Which of the three has the darkest hair?" To quote Piaget (1928),

It is as though [the child] reasoned as follows: Edith is fairer than Suzanne so they are both fair, and Edith is darker than Lili so they are both dark. Therefore Lili is dark, Suzanne is fair, and Edith is between the two. In other words, owing to the interplay of the relations included in the test, the child, by substituting the judgment of membership (Edith and Suzanne are "fair," etc.) for the judgment of relation (Edith is "fairer than" Suzanne), comes to a conclusion which is exactly opposite of ours [p. 87].

This is to say, the children have understood the functional relations—the base strings—underlying the propositions, but have not

⁴H. H. Clark and S. H. Ellis. Unpublished study, Carnegie-Mellon University, 1968.

grasped the comparative information. Developmentally, the ability to judge membership in comparative statements arrives earlier than the ability to judge relations, and this fact is closely akin to the principle of the primacy of functional relations.

In solving problems out loud, many children verbalize the underlying base strings of comparative statements directly. For example, Donaldson (1963) quotes one child as saying, "It says that Dick is shorter than Tom, so Dick is short and Tom is short too." But in the next breath the child said, "And Dick is taller than John so Dick is tall and John is short [p. 131]." The child appears to be vacillating between an interpretation of the base strings alone and an interpretation of the comparative information. The child, in the second instance, is stating the comparative relation in the only terms she knows how—as the positive adjectives *tall* and *short*.

The children in Donaldson's (1963) studies often made other errors as a result of their comprehension of propositions as base strings. For example, children were given the following Type IV problem: "Dick is shorter than Tom. Dick is taller than John. Which of these three boys is tallest?" Even though the problem explicitly states that there are three boys, many children assumed there were four. They said there were two Dicks—a tall one and a short one—following the analysis of the base strings. One girl's solution to the above problem was, "This Dick [second premise] is tallest, John is next tallest, Tom is third and then it's Dick [first premise] [p. 131]." Although Donaldson's Ss made the two-Dick error on Problem Types I and II, they did so more often on Types III and IV, problems which, because they describe Dick as both tall and short, encourage this kind of error. In all, fully 70% of the errors Donaldson observed in children on Problem Types I through IV can be traced to the children's selective interpretation of the base strings alone.

Although the principle of congruence is implicit in some of the above examples of children's reasoning, children often made it explicit. Given the problem, "Tom is taller than Dick, Dick is taller than John. Which

of these three boys is shortest?" one boy explained, "This means Dick is shorter than Tom, John is shorter than Dick. So that gives the answer—it's John [Donaldson, 1963, p. 121]." When asked why he changed the lines around, he said, "I thought it would help." He, as well as other children Donaldson reports, apparently changed the lines around to make them congruent with the question. In the present study, it has been assumed instead that it is the question that is reformulated and that it is done so implicitly. This assumption was made because the present Ss seemed to process the question after the propositions had been comprehended and stored. The assumption could be reversed, but evidence internal to Experiment II seemed to favor it as it stands.

Finally, there is evidence for the principle of lexical marking in Donaldson's and in Duthie's (1963) protocols. Both experimenters found that children sometimes misinterpreted a sentence like "Betty is older than May" to mean "May is older than Betty." For adults this is a contradiction, but for these children, it is not. The first sentence meant only that Betty is different in age from May and, as a symmetrical relation, it was synonymous with the second sentence. But "Betty is older than May" was also taken to mean "Betty is younger than May." Apparently, both *young* and *old* are interpreted in a nominal sense, so that both sentences can mean, "Betty is different in age from May." Duthie (1963) found children who made this quite explicit in quantified comparative sentences. When one child was asked in the middle of a problem how he knew that Tom was four, he replied, "Because it says that Tom is four years younger than Dick [p. 237]." Children also made this mistake in "five years older." Both Donaldson and Duthie argue convincingly that these errors result, not from misreadings, but from misinterpretations of the sentences in question. This evidence is related to some further observations on comparatives by Donaldson and Wales (in press). They found that young Scottish children could use the comparatives of unmarked adjectives, like *more*, *bigger*, *longer*, *thicker*, *higher*, and *taller*, earlier and more correctly

than their marked counterparts, *less*, *wee-er*, *shorter*, *thinner*, *lower*, and *shorter*. On this and other evidence, it has been argued (Clark, 1969a) that children develop the semantically prior nominal sense of adjectives in comparisons before they do the contrastive senses. Thus, children appear to acquire the more primitive underlying entities of a comparative before they do the more complex ones: just as they understand functional relations before they do comparisons, they understand nominal senses before they do contrastive senses.

ALTERNATIVE THEORIES FOR THREE-TERM SERIES PROBLEMS

Theory of Spatial Paralogic

One theory developed to explain the difficulties of three-term series problems was proposed by DeSoto et al. (1965) as the theory of "spatial paralogic." This theory states, simply, that in solving problems of this sort a person builds up in his mind a spatial representation of the terms involved. By a principle of directional preference, mental representations are easier to build from top down than from the bottom up. And by a principle of end anchoring, they are easier to build up from the extremes inward than from the middle outward. These two principles together are meant to predict the relative difficulty of the various three-term series problems. The present data on negative equative problems, however, appear to disconfirm both principles.

In problems containing *good* and *bad*, Ss usually visualize the three compared objects with the best on top and the worst on the bottom (DeSoto et al., 1965). According to directional preference, then, problems like *A is better than B, and B is better than C* (Type I problems) should be easier than ones like *C is worse than B, and B is worse than A* (Type II problems), because the former are top-down problems and the latter are bottom-up. This is exactly what was found by DeSoto et al. and was further confirmed in the present Experiment II and in Clark (1969b). The crucial comparison of spatial paralogic and the present theory, however, comes in top-down problems like *A*

isn't as bad as B, and B isn't as bad as C (Type I' problems) compared to bottom-up ones like *C isn't as good as B, and B isn't as good as A* (Type II' problems). By DeSoto et al.'s analysis, the former problems should be easier than the latter; the present theory predicts exactly the reverse. The data of Experiment II and of Clark (1969b) clearly support the present theory over spatial paralogic: overall, Type II' problems took significantly less time to solve than Type I' problems. In more detail, each Type II' problem can be compared to its respective Type I' problem in which each adjective is replaced by its opposite; all four comparisons support the present theory and disconfirm the principle of directional preference.

According to the principle of end anchoring, extremes-inward problems should be easier than middle-outward problems. In *A is better than B, and C is worse than B*, a Type III problem, the extremes are mentioned first in both propositions, so it is an extremes-inward problem; similarly, *B is worse than A, and B is better than C*, a Type IV problem, is a middle-outward problem. The data of DeSoto et al., as well as those in Experiment II and in Clark (1969b), bear out the prediction that III is easier than IV. But again, there is a crucial comparison of spatial paralogic and the present theory, and it comes in Problem Types III' and IV'. End anchoring predicts that Type III' problems, which mention the extremes first, should be easier than Type IV' problems, which mention the middle term first; the present theory predicts exactly the opposite. Indeed, Type IV' problems were found to be significantly easier, overall, than Type III' problems both in Experiment II and in Clark (1969b); this confirms the present theory and disconfirms end anchoring. The failure of spatial paralogic comes, apparently, from its assumption that Ss work directly from the linear encoding of terms in surface structure, disregarding deep structure per se completely.

A phenomenon that DeSoto et al. were apparently the first to notice was that Ss are able to assign nonspatial adjectives to an up-down orientation in a consistent way. Their Ss, for example, almost always placed

the "better" of two objects on top and the other on bottom; more generally from the present point of view, they consistently assigned the unmarked adjective to the upper position.

The lexical marking of spatial adjectives suggests an explanation for these consistent judgments. It is perhaps not coincidental that almost all adjectives which refer to an upward direction are unmarked—*high, tall, big, much, large*, and so on. When asked to assign *good* and *bad* to a spatial representation, Ss could easily consult their knowledge of lexical marking to make a consistent assignment with the unmarked *good* on top. On the other hand, there is no linguistic basis for a consistent left-right ordering of these adjectives; in agreement with this, the left-right assignments in DeSoto et al.'s data were rarely consistent across Ss. DeSoto et al. also found one pair that was not consistently assigned for the vertical axis—*lighter-darker*. This is only to be expected, since *light* and *dark* are both marked and afford no linguistic basis for any reliable placement.

One might argue, on the above evidence, that the principle of lexical marking should be replaced by a "principle of spatial assignments" rather than the other way around: one adjective is easier than its opposite, not because it is unmarked, but because it implies the upper position in a spatial representation. The critical test of this weaker proposition of spatial paralogic—one, however, that DeSoto et al. did not make—comes in an examination of *deep* and *shallow*. Although *deep* is unmarked, it is assigned to the *lower* position in spatial representations. This means that *deeper* problems (of Type I) should be easier than *shallower* problems by the principle of lexical marking, but harder by the principle of spatial assignment. This was tested on 21 more Ss by repeating Experiment II with two changes: (a) only Problem Types I, II, III, and IV were used; and (b) the propositions were changed to ones like *Jack has a deeper well than Dick*, and the questions to ones like *Whose well is shallowest?* The results clearly support the principle of lexical marking over its rival

principle of spatial assignment: overall, Type I problems with *deeper* were solved .44 seconds faster, $F = 11.08$, $df = 1/20$, $p < .005$, than Type II problems with *shallower*. Of the 10 Ss who claimed they used spatial imagery, eight used a vertical representation and seven of these eight visualized the deepest object on the bottom.

Theory of Constructing Spatial Images

A close relative of the theory of spatial paralogic is the theory of constructing spatial images (Huttenlocher, 1968). Like the theory of spatial paralogic, it posits that Ss arrange mental objects in imaginary spatial arrays to enable them to solve three-term series problems. The main proposal is an alternative explanation for end-anchoring. It rests on the presupposition that arranging things mentally should show the same difficulties as arranging things physically.

In physical situations, Ss have difficulties under certain instructions in placing a movable object, like a block, in relation to a fixed one (Huttenlocher, Eisenberg, & Strauss, 1968; Huttenlocher & Strauss, 1968). Given the instruction, "Make it so that the red block is under the blue block," children find it easy if the red block is in hand and the blue block is fixed, but difficult if the blue block is in hand and the red block is fixed. To summarize their results, arranging objects from an instruction is easy only when the movable object is the logical subject of a transitive verb or the grammatical subject of a "relational" sentence.

The imagerial counterparts of these manipulations, the theory states, should show the same difficulties. Consider a Type Ib problem, *B is better than C, and A is better than B*. The first proposition fixes the terms *B* and *C* in mind, *B* above *C*. The third term of this array, *A*, is now the "movable" term to be placed in relation to *B* and *C*. Since *A* is the subject of the second proposition, a "relational" sentence, the task is easy and *Ib* is quickly solved. It is not so easy to solve a Type Ia problem, *A is better than B, and B is better than C*. Here *A* and *B* are first fixed in mind with *A* above *B*, then the third term, *C*, is placed in relation to *A* and *B*. In

this case, the "movable" term *C* is not the subject of the second proposition, so it is hard to place *C* in order to solve the problem. Just as *Ib* should be easier than *Ia*, *IIb* should be easier than *IIa*, *IIIa* than *IVb*, and *IIIb* than *IVa*. The data in Huttenlocher (1968), the present Experiment II, and Clark (1969b) all confirm these predictions.

The critical comparison of Huttenlocher's spatial image theory and the present one, however, is found in the problems containing *isn't as good as* and *isn't as bad as*. By the former theory, a Type *I'b* problem, *B isn't as bad as C, and A isn't as bad as B*, like a Type *Ib* problem, should be easy. In *I'b*, the third term, *A*, to be placed relative to the two fixed terms *B* and *C*, is the subject of the second proposition, hence its placement is easy. But in *I'a*, *A isn't as bad as B, and B isn't as bad as C*, as in *Ia*, the third term *C* is difficult to place since it is the predicate of the second proposition. By the same analysis, *II'b* should be easier than *II'a*, *III'a* than *IV'b*, and *III'b* than *IV'a*. The present theory and analysis predict exactly the opposite. Compressing information from the first proposition for use in the second, as discussed above, should make *I'a* easier than *I'b*, and *II'a* easier than *II'b*. Also, by the principle of congruence, *IV'* should be easier than *III'*, overall, rather than the reverse. Each of the four possible comparisons in the results of Experiment II and Clark (1969b) support the present theory and run counter to the theory of constructing spatial images; the appropriate significance tests have been presented previously. Thus the latter theory is disconfirmed as a general explanation of reasoning in three-term series problems.

It does not follow from the disconfirmation of these two theories of spatial imagery, of course, that imagery does not occur in solving three-term series problems. It certainly does occur, although only 49% of the *Ss* in Clark (1969b) claimed that they used spatial imagery. The only firm conclusion we can draw at this time is that it has not been demonstrated that the use of spatial imagery differentially affects the solution of three-term series problems.

OTHER KINDS OF REASONING

To be of use, the present explanation for certain processes in deductive reasoning must have generality. It should not be restricted to three-term series problems alone or to problems containing comparative propositions. Several examples of another kind of reasoning problem will serve to illustrate the wider applicability of the theory proposed here.

The reasoning problem to be examined requires *Ss* to judge the truth or falsity of positive or negative statements. In Wason and Jones (1963), *Ss* were presented sentences like *29 is not an odd number* and were required to reply "true" or "false" while they were timed. In Gough (1965, 1966), *Ss* listened to sentences like *The boy didn't hit the girl*, examined a picture of either a boy hitting a girl or a girl hitting a boy, then pressed a "true" or "false" button while they were timed. The main result of interest is the interaction between the truth and positivity of the sentence: true positive and false negative sentences took less time, respectively, than false positive and true negative sentences.

When these tasks are viewed as reasoning problems, the present theory provides at least a partial explanation for the results. Wason and Jones' task can be thought of as consisting of a proposition and a question. The proposition is some previously known fact—for example, *29 is an odd number*—and the question is implicit in the presented sentence, *Is it true that 29 is not an odd number?* The underlying structure of the proposition is simply (*29 is odd*), whereas that of the question is something like (*is it true that (it is false that (29 is odd))*). The four truth and positivity questions, then, can be represented as: (a) true positive, (*is it true that (29 is odd)*); (b) false positive, (*is it true that (29 is even)*); (c) true negative, (*is it true that (it is false that (29 is even))*); and (d) false negative (*is it true that (it is false that (29 is odd))*). By the principle of congruence, *Ss* will answer more quickly when the functional relations underlying these questions are congruent with (*29 is odd*), the functional relations of the proposition. Con-

gruence is found for *a*, but not *b*, so true positives should take less time than false positives. This agrees with the results. Congruence is also found for *d*, but not *c*, so false negatives should take less time than true negatives. This is also supported by the results. In Wason and Jones' study, *S*s also made more errors on true negatives (*c*) than on anything else; they gave "false" so often presumably because a preliminary comparison showed that the functional relations of the proposition and question were different. Gough's (1965, 1966) tasks, analyzed in a similar way, further confirm these predictions; the present explanation, in fact, is essentially the same as one of the alternatives he offered for his results.

The principle of lexical marking, however, accounts for yet another part of Wason and Jones' (1963) and Gough's (1965, 1966) results. The question in the above analysis was always formulated as (*is it true that (such and such is so)*). It contained *true*, not *false*. The reason, of course, is that *true* is unmarked and *false* marked: *is it true?* implies no presuppositions about the answer—it could be either "true" or "false"—but *is it false?* implies that the answer is expected to be "false" (cf. also Fillenbaum, 1968). Thus, to answer false questions, *S* must reformulate his representation in memory to read (*it is false that (such and such is so)*), before he can give the answer, elliptically, as "false"; true questions are already in the correct form. This reformulation should take time, causing false questions to take more time overall than true questions. This prediction is confirmed in Wason (1961), Wason and Jones (1963), and Gough (1965, 1966).

CONCLUSION

In the past, deductive reasoning has often been studied as if it were an isolated process—even as if it were specific to a certain kind of task, such as the solution of three-term series problems. The processes described in the present paper, on the other hand, are quite general. They are not meant to explain the solution of two- and three-term series problems alone, but to account for cer-

tain linguistic processes in understanding statements and answering questions wherever they occur. The most important demonstration here has been that the principal difficulties inherent in many reasoning problems are not due to cognitive processes specific to these problems, but to the very language in which the problems are stated. Linguistic processes like these arise in every situation in which a problem is stated in linguistic terms.

REFERENCES

- BIERWISCH, M. Some semantic universals of German adjectivals. *Foundations of Language*, 1967, 3, 1-36.
- BURT, C. The development of reasoning in school children. *Journal of Experimental Pedagogy*, 1919, 5, 68-77, 121-127.
- CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge: M.I.T. Press, 1965.
- CLARK, H. H. The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, 1969, in press. (a)
- CLARK, H. H. The influence of language in solving three-term series problems. *Journal of Experimental Psychology*, 1969, in press. (b)
- 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., & STAFFORD, R. A. Memory for semantic features in the verb. *Journal of Experimental Psychology*, 1969, 80, 326-334.
- CLIFTON, C., KURCZ, I., & JENKINS, J. J. Grammatical relations as determinants of sentence similarity. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 112-117.
- CLIFTON, C., JR., & ODOM, P. Similarity relations among certain English sentence constructions. *Psychological Monographs*, 1966, 80(5, Whole No. 613).
- 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.
- DONALDSON, M. *A study of children's thinking*. London: Tavistock, 1963.
- DONALDSON, M., & WALES, R. J. On the acquisition of some relation terms. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, in press.
- DUTHIE, J. A further study of overlap error in three-term series problems. In M. Donaldson,

- A study of children's thinking.* London: Tavistock, 1963.
- FILLENBAUM, S. Recall for answers to "conductive" questions. *Language and Speech*, 1968, 11, 46-53.
- FODOR, J., & GARRETT, M. Some reflections on competence and performance. In J. Lyons & R. Wales (Eds.), *Psycholinguistics papers*. Edinburgh: Edinburgh University Press, 1966.
- 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.
- HALLIDAY, M. A. K. Notes on transitivity and theme in English: II. *Journal of Linguistics*, 1967, 3, 199-244.
- HANDEL, S., DESOTO, C., & LONDON, M. Reasoning and spatial representation. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 351-357.
- 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.
- HUTTENLOCHER, J., EISENBERG, K., & STRAUSS, S. Comprehension: Relation between perceived actor and logical subject. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 300-304.
- HUTTENLOCHER, J., & STRAUSS, S. Comprehension and a statement's relations to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 527-530.
- LEES, R. B. Grammatical analysis of the English comparative construction. *Word*, 1961, 17, 171-185.
- LURIA, A. R. *Human brain and physiological processes*. New York: Harper & Row, 1966.
- LYONS, J. *Structural semantics*. Oxford: Blackwell, 1963.
- LYONS, J. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press, 1968.
- MARSHALL, J. Psychological aspects of semantic structure. In R. Meetham (Ed.), *Encyclopedia of linguistics, information and control*. London: Pergamon 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.
- MILLER, G. A., & MCKEAN, K. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology*, 1964, 16, 297-308.
- PIAGET, J. Une forme verbale de la comparaison chez l'enfant. *Archives de Psychologie*, 1921, 18, 141-172.
- PIAGET, J. *Judgment and reasoning in the child*. London: Kegan, Paul, 1928.
- POSTAL, P. Underlying and superficial linguistic structure. *Harvard Educational Review*, 1964, 34, 246-266.
- 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.
- SMITH, C. S. A class of complex modifiers. *Language*, 1961, 37, 342-365.
- VENDLER, Z. *Adjectives and nominalizations*. The Hague: Mouton, 1968.
- WASON, P. Response to affirmative and negative binary statements. *British Journal of Psychology*, 1961, 52, 133-142.
- WASON, P., & JONES, S. Negatives: Denotation and connotation. *British Journal of Psychology*, 1963, 54, 299-307.
- WATT, W. C. On two hypotheses concerning psycholinguistics. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, in press.

(Received August 5, 1968)