“That man’s father is my father’s son”: The roles of structure, strategy, and working memory in solving convoluted verbal problems

PAUL J. CASEY

Charles Sturt University—Riverina, Wagga Wagga, New South Wales, Australia

Normative data ($N = 896$) collected on solutions to the traditional reasoning problem “Brothers and sisters have I none, / That man’s father is my father’s son” was the platform for a set of experiments to investigate the roles of problem structure, subject strategy, and working memory in error making in a class of verbal problems. The normative data revealed that the modal, but incorrect response to the brothers and sisters problem was “himself,” whereas the correct response of “son” was given by only 17% of subjects for the traditional version above, rising to 40% for a reordered novel version. This problem’s structure, identified as a *double-modifier structure*, was found to produce similar response patterns across a range of verbal problems. Results from (1) syntactical structure manipulations (e.g., “My father’s son is the father of that man”), and (2) efforts to teach subjects to use a linear strategy (a sequence of steps), interpreted within a framework of Just and Carpenter’s (1992) capacity theory of comprehension, suggested that working memory overload varies across within-structure modifications and subjects. Furthermore, this overload may be compounded by subjects’ choices of strategy.

Theories in several fields of complex thinking have relied heavily on variability in performance across conditions of particular classes of reasoning problems. The many studies on Wason’s deductive reasoning four-card selection task (Wason, 1966; Wason & Johnson-Laird, 1972) is prototypical of the use of data from a single problem type to develop theory.

The traditional *brothers and sisters* reasoning problem described below was the starting point for an experimental program to examine the roles of problem structure, subject strategy, and working memory in a given form of verbal reasoning problem.

You are approached by a stranger who shows you a photograph. The stranger, pointing to the photograph, says,
Brothers and sisters have I none,
That man’s father is my father’s son.

The task is to identify the person in the photograph.

The solution to this problem has been puzzled over for generations, with the majority of people giving an incorrect solution, and with many people refusing to accept what appears to be the logically correct answer, often preferring instead one particular incorrect solution. The

data from my 1988 1st-year psychology class ($N = 101$) reflect the difficulties all my classes have experienced with this problem, with 13% giving the correct solution (the son is in the photograph; the father is holding it), 77% incorrectly claiming that the person in the photograph is the stranger himself, and 10% giving another, incorrect solution.

As will become clearer in the course of this paper, the brothers and sisters problem essentially involves deciding which member of a linearly ordered set of elements (e.g., father, self, son) is the target element of the convoluted statement. As such it both contrasts with and is in the same domain as the three-term series problem type (or linear syllogism) that requires the subject to order a number of elements on some dimension (e.g., “If Mary is shorter than Jane, and Jane is shorter than Lucy, who is the shortest?”).

The three-term series problem was a focal problem in the Clark-Huttenlocher controversy over interpretative and constructive views of problem solving (see Ormrod, 1979, for a review of the literature). Clark (1969) proposed that in three-term problems the premises are first transformed linguistically, and then the final result is obtained from a condensation of the original statements. In contrast, Huttenlocher (1968) and others (e.g., Bransford, Barclay, & Franks, 1972; Potts & Scholz, 1975) argued that the subject constructs an image, going beyond the information given. The three-term series problem has also given rise to a number of mixture theories. For example, Shaver, Pierson, and Lang (1974) claimed that subjects use linguistic representations when first solving three-term problems but move to a spatial representation on later problems; Johnson-Laird (1972) proposed the opposite se-

I am indebted to the many people who assisted either as subjects or with the data collection—especially Wendy Matthews, Terry O’Connell, Greg O’Rourke, and Joanna Reynolds. I also wish to thank the many people who made helpful suggestions, especially Margaret Intons-Peterson and the referees, Eugene Winograd and a colloquium audience at Emory University, Dorrit Billman and a colloquium audience at Georgia Institute of Technology, and Mike Firment. Correspondence should be addressed to the author at the School of Humanities and Social Sciences, Charles Sturt University—Riverina, Box 588, Wagga Wagga, NSW 2650 Australia.

quence of representations; and Sternberg (1980) argued that subjects use both spatial and linguistic representations within the same set of problems. Data from subjects' attempts to solve the brothers and sisters type of problem may add to our knowledge of the strategies that subjects use to solve problems involving linearly ordered elements.

The brothers and sisters problem is also a complex, single-statement problem, and hence it has links with the literature on the comprehension of single statements. Just and Carpenter (1992) have described how the processing of complex embeddings in sentences is particularly demanding on working memory. For example, sentences in which the head noun is the object of a relative clause, such as "The reporter that the senator attacked admitted the error," produce error rates of approximately 15% in paraphrasing (Larkin & Burns, 1977). The difficulties experienced by subjects in attempting to solve the brothers and sisters problem may in part be related to working memory limitations associated with multiple embedded phrases.

In short, the brothers and sisters problem appears to be at the intersection of deductive reasoning and sentence comprehension, and data from subjects' efforts to solve this and equivalent problems may shed light on the general difficulty that subjects experience with word problems. For example, children in the United States perform 10%–30% worse on arithmetic word problems than on equivalent problems given in a numeric format (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980).

The major aims of this research, which involved 10 experiments, were as follows:

1. to produce normative data on solutions to different versions of the brothers and sisters problem;
2. to identify the sentence structure associated with incorrect responding in the brothers and sisters problem and similarly structured problems from other content domains;
3. to investigate the relationship between syntactical structure, subject strategy, and working memory in such problems;
4. to examine the outcomes from efforts to either induce or teach subjects to use a linear strategy;
5. and to relate the experimental outcomes to the literature on deductive reasoning and complex sentence comprehension.

NORMATIVE DATA COLLECTION

Experiment 1

The normative data collection included the manipulation of two variables—namely, phrase ordering of the problem statement and novelty of the version.

The rationale for including a revised phrase ordering was based on the observation that some people who respond correctly report the use of what I term a *linear strategy*, in which phrases are simplified and reorganized in a serial manner as follows. First, substitute "me" for "my father's son," then reverse the second line to produce "I am that man's father," and finally infer that the solution is "son."

Thus, if a subject uses a linear strategy, then one obstacle to obtaining the correct solution may be the number of substitutions and reversals that must be performed before the solution becomes clear. It was hypothesized that eliminating one reversal by stating the second line as, "My father's son is that man's father," would improve accuracy rate.

Furthermore, subjects might categorize the problem as "one of those trick problems" (Hinsley, Hayes, & Simon, 1977), and so be overly sensitive to some suspected deception. Hence, to test whether a novel version of the problem would yield a higher accuracy rate, another version was devised in which the second line read, "That woman's mother is my mother's daughter."

The major aims of the normative component of this research, then, were to (1) collect normative response data on the brothers and sisters problem, (2) test the hypotheses of improved performance from reversed and novel versions, and (3) examine subjects' solution strategies.

Method

Subjects. There were 896 subjects, consisting of 425 males and 471 females. The mean age was 35 years. Approximately 75 off-campus psychology students over the age of 21 years assisted with the data collection. The assistants were requested to sample from wide ranges of education levels, ages, occupations, and so forth. Within their own regions, each assistant was to collect data from 12 subjects. Subjects were drawn from over 200 towns and suburbs across Australia, with approximately 70% from the state of New South Wales. The occupations of the subjects varied widely. Most of the subjects had attained either secondary or tertiary education levels, with 31.2% having received some tertiary education.

Materials. Research assistants were issued booklets that described the procedure, together with forms for scoring demographic data, responses, and subjects' reported rationales for the solution given to the problem. There were four versions of the problem, the first line always being "Brothers and sisters have I none." The second line variations were as follows:

Traditional. "That man's father is my father's son."

Traditional reversed. "My father's son is that man's father."

Novel. "That woman's mother is my mother's daughter."

Novel reversed. "My mother's daughter is that woman's mother."

The context and wording of the traditional problem version were as follow, with appropriate adjustments for each group:

Suppose a stranger comes up to you and introduces himself as Jack. Jack shows you a picture (picture shown), and says to you:

Brothers and sisters have I none,
That man's father is my father's son.

Jack continues, "Who is the person in the picture?"

For the novel version, Jill was substituted for Jack, and a woman's picture was used.

Procedure. The subjects were pseudorandomly allocated to a condition, the restrictions being that each research assistant was requested to use the traditional phrase sequence for half the subjects and the reversed ordering for the other half, and also to attempt to balance gender across conditions.

Each subject was initially read the problem for the given condition. The subject was then asked whether or not he/she had heard the problem before, and if so, when, as well as any answer that he/she recalled. All the subjects were then requested to attempt to solve the problem without the aid of any writing materials. The subjects were requested to choose an answer from the set {grand-

father, father, himself, son, grandson} for the traditional version, or from the equivalent feminine set for the novel version. The subjects were also requested to give a confidence rating for their answers, ranging from 1 (*very high*) to 5 (*very low*). They were allowed 5 min and were advised that the problem could be repeated as often as the subject wished. After the subject had selected an answer, the subject was requested to explain the steps taken in the attempt to solve the problem.

Results

Forty-two percent ($n = 376$) of the subjects had heard a version of the problem before—44.6% and 39.2% of the subjects in the traditional and novel version conditions, respectively. Close to 9% ($n = 78$) of the subjects reported that they had chosen their answers solely on the basis of memory of a previous answer ("father"/"mother," 2; "himself"/"herself," 45; and "son"/"daughter," 31), and their data were dropped from further analyses. For the remaining subjects, the patterns of results were quite similar for those for whom the problem was familiar but could not recall a definite answer and for those for whom the problem was novel, and no further data were omitted prior to further analyses. Alpha was set at .05.

Order and version manipulations. The manipulations of the order of phrases in the problem and the version of the problem were both effective in significantly improving response accuracy. The percentages correct for each phrase ordering of each version of the problem are given in Table 1. The response accuracy data for order of phrasing and problem version were subjected to a log-linear analysis using a nonsaturated model (Knoke & Burke, 1980), with order of phrasing and version of the problem assumed to be independent. The model provided a good fit to the data [$\chi^2(1, N = 818) = .061$]. The ratios of the parameter estimates to their respective standard errors enabled a test of the null hypothesis that a particular parameter estimate, λ , was zero. The obtained z values for the independent variables order of phrasing and version of the problem were highly significant, being 4.15 and 3.20, respectively. The regression coefficients for order of phrasing and version of the problem were .17 and .13, respectively.

On the traditional version of the problem, 17.2% of subjects responded correctly with "son"; 60.8% responded "himself." Overall, 28.5% of the subjects gave the correct response, compared with 47.2% who gave the er-

roneous response "himself" or "herself." Most of the other 24.3% of incorrect responses were divided fairly evenly over the remaining response options, but a few subjects would not accept any of the offered alternatives. (One subject responded "Bruce Wayne," but it was later found that this subject incorrectly heard the second line as "Batman's father...")

Level of education was significantly related to response accuracy, with 33.3% of the university-educated subjects overall giving correct responses, as opposed to 26.3% of the nonuniversity-educated subjects [$\chi^2(1, N = 818) = 4.28$]. Confidence was positively and significantly related to accuracy ($r_{pb} = .30$). The mean confidence ratings for son, self, and other responses were 1.95, 2.54, and 3.25, respectively, with each mean significantly different from the other two. Neither age nor gender was related to response accuracy.

Reported reasoning processes. To examine the reported reasoning processes, 100 response forms were randomly selected. Of the 31 subjects who gave a correct response, 27 gave a linear analysis as described in the introduction; 1 guessed; and 3 could not give any explanation. Of the 69 subjects who responded erroneously (52 himself/herself), 14 guessed; 47 gave arguments with some semblance of logic but with at least one faulty assumption; and 8 gave seemingly unintelligible accounts.

Examination of the reported reasoning processes of subjects responding "himself"/"herself," together with further questioning, indicated that many such subjects appeared to be employing what I term an *equivalence* strategy. Thus, many of these subjects noted an equivalence between "that man's father" and "my father's son." Some then responded that "my father's son" is the reverse of "that man's father," so it has to be "himself." The observation is correct but the implication is wrong. Others argued, "I am my father's son and 'that man' is his father's son, so it has to be me." Each part of the argument is correct, but the whole argument is a subtle, incorrect restatement of the original problem.

It was noteworthy that, consistent with 21% of "himself"/"herself" respondents' giving the highest confidence rating (1), many subjects who responded incorrectly could not be convinced of the logic of the correct answer.

Discussion

The normative data were consistent with past observations and also with data from a later partial replication ($N = 744$). However, the effects of the experimental manipulations, while significant, were rather weak. Converting the χ^2 values for the Experiment 1 relationship between phrase order and accuracy [$\chi^2(1, N = 818) = 17.60$] and problem version and accuracy [$\chi^2(1, N = 818) = 10.52$] to the correlation ϕ yielded the low values of .15 and .11, respectively. Moreover, while the "himself" confidence ratings were significantly lower than the "son" confidence ratings [$t(569) = 6.71$], the degree of association between level of accuracy and confidence was rather low ($\omega^2 = .06$). Finally, the ϕ value for the rela-

Table 1
Percentages Correct in Brothers and Sisters Problem From Manipulating Both Phrase Order and Problem Version

Order	Version			
	Traditional*		Novel†	
	% Correct	N	% Correct	N
1	17.2	204	26.5	200
2‡	29.6	203	40.3	211

*"That man's father is my father's son."

†"That woman's mother is my mother's daughter."

‡E.g., "My father's son is that man's father."

tionship between level of education and response accuracy was a low .10. Thus, there is more to the difficulty of this problem than a sentence reversal or the familiarity level.

The oral presentation method was considered as possibly accounting in part for the low accuracy rate. However, results from previous testing in writing, and a written partial replication of Experiment 1 ($N = 93$) revealed no improvement.

The significant effect of novelty of version was consistent with response bias' playing some role in response variability. Moreover, the wording of line 1, "Brothers and sisters have I none," may promote an atmosphere effect of *self*. Indeed, several subjects who excelled at mathematics and computing reported that they knew the correct answer was "son" but were strongly drawn to "himself." (See Evans, 1982, and Woodworth & Sells, 1935, for a discussion of the atmosphere effect in syllogistic reasoning.)

However, the relatively high error rates across all conditions of Experiment 1 are consistent with the attribution of error making to problem structure. Experiment 2, the first of a set of experiments aimed at delineating the structure of the brothers and sisters problem, was a test of whether it was the structure of the problem or atmosphere that produced the high error rate.

STRUCTURE EXPERIMENTS

Experiment 2

The atmosphere versus structure issue was tested first by employing a problem similar in structure and content to the brothers and sisters problem, but in which the solution "himself" was impossible. If subjects chose "himself" solely on the basis of atmosphere, this new problem would force subjects to select an alternative answer. If most subjects now selected the correct response, it could be inferred that the "himself" response was in part a consequence of atmosphere. However, if most subjects chose an incorrect answer from the same level of the kinship hierarchy as "himself," (e.g., brother or cousin), and with a frequency similar to that for "himself" on the traditional version, the atmosphere effect hypothesis could not be accepted.

Next, a problem in which the correct answer was "himself" (e.g., "Brothers and uncles have I none, that man's father is my grandfather's son") would produce a very high *self* response rate under an atmosphere hypothesis, but a relatively low *self* response rate under a structure hypothesis. Experiment 2 was an examination of the atmosphere versus structure alternatives.

Method

Ninety subjects, drawn from university students and the general community, were presented with one of three problems, with 30 subjects pseudorandomly allocated to each problem. The problems were the following:

1. The traditional version of the brothers and sisters problem.
2. A problem similar in structure to the brothers and sisters problem, titled "Chipper's relative" (*The Puzzle Book*, 1982):

A raccoon named Chatter saw a friend named Chipper talking to a young raccoon. As Chatter came near the young raccoon trotted off. "Who was that little raccoon you were just talking to?" asked Chatter. "Why, that raccoon's father is my father's only son," answered Chipper.

Assuming all the raccoons were male, in what way were Chipper and the young raccoon related to one another?

3. The revised brothers and sisters problem:

Brothers and uncles have I none,
That man's father is my grandfather's son.

The three problems were presented in writing, and the subjects were allowed 10 min to work out a solution on paper. The Chipper problem was presented with a picture showing two raccoons talking to each other as they watched a third raccoon in the background running away. In other respects, the problem paralleled the brothers and sisters problem. The subjects were not constrained to any response set for any of the problems.

Results and Discussion

The results favored a structure rather than an atmosphere explanation. The modal response for each condition was incorrect, being (1) "himself," (2) at the "himself" level, and (3) "father," for Problems 1, 2, and 3, respectively, and in each case being at the level of the kinship hierarchy predicted by the structure hypothesis. The relationships between problem types and accuracy levels for the Chipper and traditional problems, and for the revised and traditional problems were non-significant [$\chi^2(1, N = 60) = .69$ and 1.20 , respectively]. The number of subjects for each problem responding correctly, giving the incorrect, modal response, and giving another incorrect response, are shown in Table 2.

The data so far have revealed that most subjects cannot solve the brothers and sisters problem correctly, and that the difficulty may lie in the structure of the problem. Indeed, a series of partial replications of the brothers and sisters problem that retained the structure but varied the relationships expressed in either the subject or the predicate (e.g., "That man's grandfather is my father's son"), with the first line also adjusted appropriately, yielded error rates similar to the error rates from the traditional problem. Furthermore, the results are similar to those from Experiment 1 when the subjects are first told the

Table 2
Numbers of Subjects ($N = 30$ per Group) Giving the Correct, Modal Incorrect, and Other Incorrect Responses for the Traditional Brothers and Sisters Problem, the Chipper Problem, and the Revised Brothers and Sisters Problem

Response	Problem Version		
	Traditional	Chipper	Revised
Correct	8	11	12
Modal, Incorrect	18	15	16
Other Incorrect	4	4	2
Total	30	30	30

Note—The modal response for the traditional and Chipper problems was "himself." However, for the revised problem, in which the correct answer was "himself," "father" was the modal response. Response frequencies rather than percentages are reported because of the relatively low N .

correct response (e.g., "son") and then required to choose from a set of convoluted statements (e.g., "That man's father is my father's grandson," "That man's father is my father's son") which statement yields this correct response. Subjects typically and erroneously choose a statement whose solution is a person one level below the target person in the kinship hierarchy. For example, subjects who are given "son" as the required correct solution typically select the statement "That man's father is my father's grandson" from the available set of statements.

The next task was to discover the precise problem structure associated with high levels of incorrect responding. This aim was taken up in Experiments 3 and 4.

Experiment 3

Experiment 3 was an attempt to discover the level of sentence structure at which successful management of the brothers and sisters problem breaks down by comparing results from employing the traditional, two-possessive-phrase problem structure with results from using a simpler structure containing only one possessive phrase.

Method

Forty-eight subjects, drawn from a range of occupations, all of whom had completed at least all years of high school, were pseudorandomly allocated to a level of one of the following conditions ($n = 16$ per level).

Condition 1. Baseline; the traditional brothers and sisters problem.

Condition 2. Reduced complexity of structure condition.

Level 1. The second line read, "That man is my father's son."

Level 2. The second line read, "That man's father is me."

The problems were presented in writing, the response set was constrained to {grandfather, father, himself, son, grandson}, and the subjects were allowed 10 min to reach a solution.

Results and Discussion

The results for each level in Condition 2 were very similar, so these within-condition data were combined for statistical analysis. In the baseline condition, 4/16 subjects were correct, with 9/16 responding "himself." However, 30/32 subjects were correct in Condition 2. Hence, reducing the level of complexity by eliminating a possessive phrase from either the subject or the predicate significantly improved performance [$\chi^2(1, N = 48) = 24.38, \phi = .71$].

The syntactical structure associated with difficulties in the brothers and sisters problem is emerging. The problem form is characterized by both the subject's and the object's consisting of a noun phrase, each of which is modified by a possessive phrase, with the problem's target being the noun in one of the possessive phrases. Experiment 3 demonstrated that the problem's difficulty disappeared with the removal of one of the modifying phrases.

The picture remains incomplete, however, until we test the hypothesis that the problem's target must be part of the subject's possessive phrase rather than the subject itself. Hence, Experiment 4 was an attempt to complete the identification of the structural characteristics of such problems by comparing subjects' performances on the traditional brothers and sisters problem type with performances

on a similar problem, still with two possessive phrases but with the problem target as the subject of the sentence.

Experiment 4

Method

Twenty-four subjects, drawn from across the community, were pseudorandomly allocated to one of two problem conditions ($n = 12$ per condition).

The problems for both Conditions 1 and 2 were presented in writing, and each contained the following common opening: "You enter a shop, and see a shop assistant with the name tag 'Bill' on his shirt. You say to Bill, 'I'd like to see your boss.'"

In Condition 1, the problem continued, "Bill replies, 'Well, you want my father's boss. My father's boss is my father's only son.'" However, in Condition 2, the problem continued, "Bill replies, 'My boss's father is my father's only son.'" The subjects in each condition were required to identify how Bill and the boss were related.

Results and Discussion

There was a significant relationship between response accuracy and condition [$\chi^2(1, N = 24) = 16.67$]. In Condition 1, the correct answer, "Bill is the boss," was given by 11 of the 12 subjects. In contrast, in Condition 2, the correct answer, "Bill's son," was given by only 1 subject, with all 11 error responses being "Bill."

Hence, error making seems associated with a sentence whose nouns are elements drawn from a set of linearly ordered objects (e.g., father, self, son), with (1) both the subject and the predicate complement each possessing a noun phrase and a possessive phrase, and (2) the problem target being the subject's possessive phrase. (The data from the Experiment 1 conditions in which the problem statement was reversed suggest that the problem target may also be the object's possessive phrase.) I now show that this problem structure, which I will term a *double-modifier* structure, provides difficulties for subjects across content domains.

GENERALIZING ACROSS PROBLEM DOMAINS

This section focuses on verbal versus algebraic performance on problems with the double-modifier structure, and also reports summary data from a number of experiments in which this structure was used in other domains.

Experiment 5

Previous research on arithmetic verbal problems (e.g., Carpenter et al., 1980; Cummins, Kintsch, Reusser, & Weimer, 1988) has shown that children encounter more difficulties with problems presented in words than in a numeric format. For example, Cummins et al. reported a problem that only 29% of first-grade children solved when it was in word form, but that 100% solved when it was in numeric form.

Many relatively simple numerical expressions, such as $5-1$, can also be converted to seemingly simple verbal expressions, such as "the number before 5." Thus, an algebraic sentence such as

$$x-1 = 5-1 \quad (1)$$

can be stated verbally as either,

The number before the number I'm thinking of is 5-1, (2)

or,

The number before the number I'm thinking of is the number before 5. (3)

Note that Sentence 3 contains prepositional phrases rather than possessive phrases because possessives such as "5's predecessor" sound unnatural. However, pilot work on the brothers and sisters problem with prepositional phrases indicated that the results were substantially the same as when the possessives were used.

If the findings from the brothers and sisters problem would more generally apply to verbal problems employing numbers, then the pattern of results from such problems should match the pattern of results from the brothers and sisters problem. Unlike in the brothers and sisters problem, however, two errors are likely, depending in part on the strategy used. Thus, subjects who use a linear strategy incorrectly may argue, "the number before 5 is 4, and the unknown number is the number before that, which is 3." However, subjects who either use an equivalence strategy incorrectly or forget the target may claim, "the solution to the left side equals the solution to the right side, which is 4." Hence, it was predicted that the most common errors would be 3 and 4.

In Experiment 5, subjects were tested on the three kinds of statements above. It was expected that subjects tested on the algebraic sentence (1) would perform better than subjects tested on the mixed mode sentence (2), and that the poorest performance would be that of subjects tested on the verbal problem (3).

Method

Seventy-two subjects, drawn from a range of occupations, were pseudorandomly allocated to one of the following conditions, 24 per condition; the subjects were required to solve the sentence peculiar to the condition. In Conditions 2 and 3, the problem began with, "I am thinking of a number," and concluded with, "What is the number I am thinking of?"

Condition 1. Mathematical symbols in an algebraic sentence.

$$x-1 = 5-1. \text{ Solve for } x.$$

Condition 2. Verbal and mathematical symbols.

The number before the number that I'm thinking of is 5-1.

Condition 3. Verbal presentation.

The number before the number that I'm thinking of is the number before 5.

The subjects were given the problems in writing and were allowed 5 min to complete the task.

Results and Discussion

As predicted, accuracy decreased as the problem became more verbal [$\chi^2(2, N = 72) = 30.70$]. The numbers of subjects correct out of 24 for Conditions 1, 2,

and 3 were 23, 14, and 4, respectively. In Condition 3, the accuracy rate was 17%, and the predicted errors of 4 (3 subjects) and 3 (9 subjects) made up 50% of the responses, a response pattern that matches the pattern from Experiment 1.

The results from several follow-up experiments confirmed the reliability of the results from the number problem experiment. First, three further number problems in the form of Condition 3 of this experiment ($n = 22$ per problem), but with the modifying-phrase prepositions "before, before" replaced, respectively, with "after, after," "after, before," and "before, after" (e.g., "The number before the number I'm thinking of is the number after 5"), yielded accuracy rates that did not differ significantly from that of the "before, before" problem. Likewise, four day-of-the-week problems ($n = 22$ per problem) matched in structure with the four number problems—for example, "The day before the day I'm thinking of is the day before tomorrow"—produced response patterns that were similar to those produced by the number problems. Tests with the chi-square statistic found that responses patterns did not differ either between or within the number and day problem types.

Similar results were found also for the day-of-the-week problem in the possessive form (e.g., "X's yesterday is yesterday's tomorrow"). Furthermore, results from problems from other domains whose members were ranked on some dimension (e.g., presidents, kings/queens, streets, and towns) were also similar to the results reported so far. Thus, the verbal problem "The president before the one we will study is the president before Nixon" will lead many people to respond with "Kennedy" and some to respond with "Johnson," rather than to respond correctly with "Nixon."

I now examine the interaction between subject strategy and structural variation within the general structure of the double-modifier form of problem.

STRUCTURAL VARIATION EXPERIMENTS

Within-structure manipulations that may be performed without affecting the overall double-modifier structure include the following:

Factor 1. Symmetry—that is, whether or not a middle term such as "father" is modified in both subject and predicate (e.g., "That man's father is my son's father," and "That man's father is my father's son").

Factor 2. Order—that is, the position of the unknown element, with the target being placed in either the subject phrase or the predicate complement (e.g., "That man's father is my father's son," and "My father's son is that man's father").

Factor 3. Modifying phrase type—that is, the use of possessive or prepositional phrases for the modifying phrases (e.g., "The father of that man is the son of my father").

Hence, there are at least eight ways of writing the brothers and sisters problem, 16 if the mixing of possessive and prepositional phrases is added as another level

of Factor 3 (e.g., "That man's father is the son of my father"). Note, however, that some problem statements in other domains cannot be expressed comfortably with possessive phrases. Thus, "Kennedy's successor" sounds natural, but "five's predecessor" does not.

A heuristic approach was adopted in exploring further which within-structure formats might be associated with relatively high accuracy rates by selectively manipulating the symmetry, order, and modifying-phrase-type factors. A heuristic approach was chosen because of the large subject pool that was required and because of the difficulties of interpretation that were associated with employing all levels of these three factors in the one experiment.

It is not clear from the data so far whether subject strategy is a general individual difference factor, or whether it is determined by the particular syntactical structure of the problem. There are some difficulties in determining how structure and strategy behave, partly because many subjects appear unable to articulate how they attempt to solve the double modifier problem. (See Johnson-Laird, 1983, p. 2, on the value of introspection.) The data so far reveal that the majority of correct respondents report what seems to be an intelligible linear strategy, while many protocols of incorrect respondents are difficult to categorize in terms of either strategy or source of error. The first of the structure-strategy experiments, Experiment 6, was based on some of the more intelligible protocols of error-making subjects; it was an investigation of the relationship between mode of problem representation and error making.

Experiment 6

Observations of subjects' effort to solve the brothers and sisters problem revealed that some subjects represented the problem diagrammatically. Moreover, some subjects drew a box on the middle left of the page, which they labeled "that man," and then drew an arrow upward, and wrote the word "father" above the arrowhead. Next, these subjects drew a box on the middle right of the page, which they labeled either "me" or "son," then drew another arrow upward, and again wrote the word "father" above this arrowhead. These subjects typically began their explanation with, "I am the son of my father, and 'that man' is the son of his father." The subjects who placed "me" in the box under "father" on the right side of the page then matched "me" with "that man" and so concluded that the answer was "himself." However, the subjects who used "son" in this box, following the same matching strategy, concluded that the answer was "son."

The possibility that accuracy rate was associated with mode of problem representation was tested by offering the subjects a choice from four possible solutions to the brothers and sisters problem, as described below.

Method

The subjects, 26 community college students over 21 years of age, were required to choose their preferred solution to the brothers and sisters problem from four alternative solutions prepared by the experimenter. Each solution consisted of either a straight verbal

or a diagrammatic plus verbal solution that led to the answer "son," or a straight verbal or a diagrammatic plus verbal solution that led to the response "himself." The diagrammatic solutions were as discussed above. One verbal-only solution consisted of the linear explanation that because "my father's son" is "me," and because the father of "that man" is "me," "that man" must be the son. The second verbal solution consisted of the argument that "I'm the son of my father, and that man is the son of his father, so it must be himself." The four solutions were offered in the context of four friends debating the correct solution to the problem. The subjects were requested to attempt to explain their choices and were given 10 min to complete the task.

Results and Discussion

The main results were (1) that response accuracy was unrelated to choice of mode of preferred problem representation [$\chi^2(1, N = 26) = 0.66$], with the overall accuracy level being 27%; (2) that most subjects (18/26) preferred the verbal mode of argument; and (3) that 19/26 subjects accepted a false equivalence argument. (The subjects' preference for an equivalence explanation does not necessarily mean that this is how these subjects arrived at the solution. It is possible that this explanation was chosen because it led to the subjects' preferred answer of "himself.")

Experiment 7 was a test of the possibility that a within-structure modification of the brothers and sisters problem that facilitated the use of an equivalence strategy would improve performance.

Experiment 7

The statement "That man's father is my father's son" (asymmetrical) can be restated as "That man's father is my son's father" (symmetrical) and retain the same structure and solution. The latter version, unlike the traditional version, permits an unambiguous factoring approach,¹ in which we express the problem as "father (that man) equals father (my son)," which allows the common term "father" to be canceled out, leaving "that man" equal to "my son." At the same time, there should be no disadvantage for linear strategists.

Results from pilot studies supported the hypothesis that accuracy rates would improve for symmetrical problem versions, with an overall accuracy level of 77.5% ($N = 66$) being obtained across a range of symmetrical problem statements from the kinship domain. Experiment 7 was intended to confirm the pilot study results, and also to test the effect of manipulating the kind of modifying phrase—prepositional or possessive.

Method

Each of 48 subjects, 24 female and 24 male volunteers from the community at large, varying in age (18–69 years) and occupation, was pseudorandomly allocated one of four versions of the brothers and sisters problem. The details of the problem were presented in writing, and the subjects were allowed 10 min to solve the problem. Two factors were manipulated in producing the four statements—namely, the level of symmetry and the level of modifying phrase. The key parts of the test statements were as follows:

1. "That man's father is my son's father" (symmetrical, possessive phrases).

2. "The father of that man is the father of my son" (symmetrical, prepositional phrases).
3. "That man's father is my father's son" (asymmetrical, possessive phrases).
4. The father of that man is the son of my father" (asymmetrical, prepositional phrases).

Results and Discussion

The level of modifying phrase had no effect on response accuracy, so the data were collapsed appropriately for further analysis. There was a significant relationship between response accuracy and level of symmetry, with 12 out of 24 subjects correct for the symmetrical statements and 3 out of 24 correct for the asymmetrical statements [$\chi^2(1, N = 48) = 7.38$].

The relatively high accuracy rate for the symmetrical versions is consistent with the hypothesis that symmetrical versions of the brothers and sisters problem will promote more accurate responding. However, inspection of the protocols from the 12 subjects who responded correctly in the symmetrical condition revealed that only 2 of them clearly used an equivalence strategy, whereas the other 10 seemed to use a linear strategy. Furthermore, the error-making subjects showed no evidence of having used an equivalence strategy; in the main, their reported strategies were incoherent.

The consistency with which subjects reported linear rather than equivalence strategies across the experiments up to this point cannot be ignored. It may help explain the relatively low accuracy rates (around 35%) for the day and number problems of Experiment 5 given in a symmetrical form (e.g., "The day before x is the day before tomorrow"), suggesting that symmetrical versions do not necessarily lead to similar improvements in performance across all problem domains. Hence, Experiment 8 was focused on the effects of attempting to facilitate the correct use of a linear strategy.

Experiment 8

If most subjects use some form of linear strategy, one method of improving group performance would be to structure the problem so that a linear strategy would be less error prone. The order factor seems relevant here. The brothers and sisters problem contains both left-branching and right-branching phrases, with the target element in the left-branching phrase. Left-branching sentences are more difficult to process than right-branching sentences because the left embedded phrase or clause disrupts the processing of the main clause, and information about the main clause must be retained while the left embedded phrase or clause is processed (Fodor, Bever, & Garrett, 1974). In the traditional brothers and sisters problem, linear strategists are initially unable to resolve the left-modifying phrase, but must retain the information about the subject and its modifying phrase until the object and its modifying phrase have been processed. Just as left-branching sentences cause more strain on working memory than right-branching sentences do, so too the positioning of the target element in the left-branching

phrase may place sufficient strain on working memory to lead to a breakdown in processing. Alternatively, subjects may hypothesize and incorrectly test a solution as part of their left-to-right processing. Indeed, several subjects who vocalized their processing made statements such as, "If it is 'himself,' then . . ."

The reversed order brothers and sisters problem condition of Experiment 1 ("My father's son is that man's father") produced a small but significant improvement in response accuracy. Subjects may still have had difficulties in making a correct predication, possibly because "that man," still adjacent to "my father's son," would be predicated incorrectly in cases of working memory overload.

Hence, in order to facilitate correct predication and to attempt to promote the use of a linear strategy, the problem form "My father's son is the father of that man," was tested as one of the conditions of Experiment 8.

In Experiment 8, both order and the form of the modifying phrase type (prepositional for subject phrase and possessive for object phrase, or vice versa) were manipulated as shown in the following Method section. It was expected that subjects would perform significantly better in the Order 2 condition. Moreover, an interaction was predicted such that Order 2 performance would be superior with the prepositional phrase in the predicate because this condition of all conditions should facilitate predication best.

Method

Each of 86 1st-year psychology students was pseudorandomly allocated one of the four brothers and sisters versions given below. The subjects were allowed 10 min to complete the task. The usual introduction to the problems was given in writing. The key lines of each problem were as follows:

Order 1

Condition 1. Subject as prepositional phrase.

The father of that man is my father's son.

Condition 2. Predicate as prepositional phrase.

That man's father is the son of my father.

Order 2

Condition 1. Subject as prepositional phrase.

The son of my father is that man's father.

Condition 2. Predicate as prepositional phrase.

My father's son is the father of that man.

Results and Discussion

For Order 1, 7/21 and 6/21 subjects were correct for Conditions 1 and 2, respectively, with 28 of the 42 subjects responding with "himself." For Order 2, 15/22 and 21/22 subjects were correct for Conditions 1 and 2, respectively. A log linear analysis, using a saturated model, revealed a significant effect for order ($z = -4.26$), and a significant interaction effect between order and modifying phrase type ($z = -1.86$).

The high level of accuracy for the Order 2, one prepositional and one possessive phrase condition, is replicable (e.g., Dempsey, 1991). Moreover, the replicability holds across content domains. Hence, one of the linear strate-

gists' difficulties with the traditional double modifier problem appears to have been the making of a correct predication, possibly associated with working memory overload produced by having the problem target in the subject phrase.

I now report outcomes from training studies, first to lead subjects to the correct solution of problems at the double-modifier level after earlier having obtained the correct solutions of problems with a single-modifier structure (Experiment 9), next to teach subjects a linear strategy working with examples in the double-modifier structure (Experiment 10), and finally to teach subjects first to restructure the problem and then to use a linear strategy (Experiment 10, Follow-Up Study).

TRAINING STUDIES

Experiment 9

A pilot study was used to check the potential for improved performance by drawing subjects' attention to salient features of the double-modifier structure. However, of 32 subjects given the traditional version of the brothers and sisters problem, 16 of whom were advised to solve "my father's son" first, and 16 of whom were advised to note that the target was "that man," only 5 correctly responded with "son," whereas 21 responded with "himself." Hence, Experiment 9 was intended to assess the effects of attempting to lead subjects to the correct version by first having them solve single-modifier versions of the brothers and sisters problem prior to attempting the more complex, traditional double-modifier version. (See Reed, Ackinclose, & Voss, 1990, and Ross & Kennedy, 1990, for recent research on the use of examples to aid problem solving.)

Method

Sixty-four subjects, drawn from university students and the general community, were pseudorandomly allocated to a level of one of the following conditions ($n = 16$ per level) of the brothers and sisters problem. Because the population was the same as in several previous experiments, no new baseline data were collected. Depending on the condition, the subjects first received either one or two simplified versions of the brothers and sisters problem, with the second line simplified as shown below, and were then tested with the traditional version. No feedback was given on the accuracy of the practice problem responses.

Condition 1. Practice with one simplified problem.

Level 1. "That man is my father's son."

Level 2. "That man's father is me."

Condition 2. Practice with two simplified problems.

Level 1. Practice 1. "That man is me."

Practice 2. "That man is my father's son."

Level 2. Practice 1. "That man is me."

Practice 2. "That man's father is me."

The problems were presented in writing, the response set was constrained to {grandfather, father, himself, son, grandson}, and the subjects were allowed 10 min to reach a solution. After presenting a solution, subjects were presented with a written tutorial explaining a verbal, linear strategy for obtaining the correct solution. Finally, the subjects who had responded incorrectly to the traditional problem and had been tutored on how to obtain "son" as

Table 3
Number of Subjects Correct, Responding "Himself," and Giving Other Responses for the Test Problem of Experiment 9

Response	Condition			
	1		2	
	Level 1	Level 2	Level 1	Level 2
Correct	7	6	5	4
Himself	8	8	8	12
Other	1	2	3	0
Total	16	16	16	16

the solution were asked their level of acceptance {full, part, or not at all} of the case for "son's" being the correct response to the traditional problem.

Results and Discussion

As in Experiment 3, reducing the complexity of the problem by one or more steps resulted in almost 100% accuracy on the practice problems. However, the prior working on less complex versions had no effect in improving test problem accuracy, with 56% of the subjects responding "himself," and 34% giving the correct response. Interestingly, 22 of the 42 subjects who had responded incorrectly to the test problem in Conditions 1 and 2, following presentation of a written, linear explanation of the problem's solution, would either only partly or not at all accept a solution other than "himself." The numbers of subjects correct, responding "himself," and giving other responses are shown in Table 3.

The inability of many subjects to make use of the seemingly apparent similarities and differences between the practice and test statements is consistent with working memory overload, and also consistent with the hypothesis that the more complex structure of the test problem may increase the likelihood of a subject's adopting an error-prone strategy. The tutorial's failure to persuade many error-making subjects of the implications of the similarities and differences between the practice and test problems indicates the power of the effects, as well as the difficulties in teaching people problem-solving techniques that strain working memory.

If there is a problem in moving from a simple to a more complex structure, subjects may acquire a reliable solution strategy better by training on double-modifier problems. Experiment 10 was an attempt to examine improvements in subjects' performance when subjects were trained with a linear strategy on double-modifier verbal problems.

Experiment 10

Method

Five 1st-year psychology tutorial groups were allocated to one of the following three conditions:

1. One group ($n = 20$) to the baseline condition.
2. One group ($n = 21$) to a training control condition.
3. Three groups ($n = 53$) to the training experimental condition.

There was no reason to suspect that the groups differed in any relevant manner.

Baseline subjects. The baseline subjects received a written version of the traditional brothers and sisters problem.

Training control subjects. The training control subjects received two written practice problems as well as the test problem.

Practice Problem 1

I am thinking of a number.

The number before the number I'm thinking of is 5 plus 1 minus 1.

What is the number I'm thinking of?

Practice Problem 2

I am thinking of a day of the week.

The day before the day I'm thinking of is the day before tomorrow.

What is the day I'm thinking of?

Problem 3: Test

The third problem was the brothers and sisters test problem.

The training control subjects attempted each problem in the preceding order and were not given any assistance.

Training experimental subjects. The training experimental subjects received the same problems as did the training control subjects but with the following procedural modifications. The subjects first completed Practice Problem 1 and then discussed their answers. Error-making subjects were tutored in a group on the rationale for the correct answer, and all subjects were instructed to attempt the remaining problems in the following manner: simplify the predicate, rewrite the problem, identify the subject, check for the target of the problem, and deduce the answer. Any difficulties encountered were discussed, and all subjects claimed that they were in agreement with the correct answer before the group moved to the next problem. The procedure for Practice Problem 1 was repeated for Practice Problem 2, prior to the subjects' attempting the test problem.

Results and Discussion

Training with problems in the double-modifier structure was effective in significantly improving performance on the brothers and sisters problem. Overall, there was a significant relationship between accuracy and level of experience with the problem structure [$\chi^2(2, N = 84) = 14.99$]. Baseline accuracy was similar to that in previous experiments, with 5 of the 20 subjects (25%) giving the correct answer ("son"). Nine of the 21 subjects (43%) in the training control group were correct, whereas 38 of the 53 (72%) training experimental group subjects responded correctly.

The medium-term effects of training were tested by having the trained subjects respond 1 week later to modified brothers and sisters problems, such as "That man's grandfather is my grandfather's son." The accuracy level was 69%, similar to the performance immediately following training. This accuracy level is similar to that obtained by Cheng, Holyoak, Nisbett, and Oliver (1986) for conditional reasoning tasks with a combination of rule and example training.

The training method above, which began with the subjects simplifying the predicate, was not effective for some 30% of subjects. Because Order 2 versions of the brothers and sisters problem yielded improvements in response accuracy, a follow-up training study was used to test the effect on response accuracy of having subjects first restructure the problem so that the target was placed in the predicate of the sentence.

Experiment 10, Follow-Up Study

Method

Eighty-one subjects, 1st-year psychology students, were tutored on solving the traditional brothers and sisters problem according to the following restructuring procedure. First, they were to circle the problem target. Second, if the target was in the first half of the sentence, they were to rewrite the sentence so that the target was in the second half of the sentence (e.g., "My father's son is that man's father"). Third, they simplified any phrase that could be reduced (e.g., "my father's son"), and they rewrote the sentence (e.g., "I am that man's father"). Fourth, they were to ignore any earlier hypothesis, focus solely on the simplified version, and deduce the solution.

Twenty-five subjects received a written test set of five double-modifier problems; a further 56 subjects received one extra problem in addition to the other five problems. There were three problems from the kinship domain (e.g., "That man's grandfather is my father's son"), and one each from three other domains. The subjects were given 20 min to complete the set.

Results and Discussion

Accuracy rates across the three kinship problems ranged from 67% to 70%. The other accuracy rates were 70% for a number problem, 73% for a town problem, and only 21% on a day problem, "Two days after the day that I'm thinking of is two days before tomorrow." Nine out of the 81 subjects made no errors, 22 subjects made only 1 error, and 11 subjects were correct only on either 1 or 0 test problems.

Hence, each double-modifier problem caused difficulties for some 30% of the subjects despite linear strategy training. It was not clear from the subjects' protocols whether the major difficulty occurred in restructuring the problem or in then solving the restructured problem, partly because many subjects did not show detailed work. However, the subjects who rewrote a simplified version of the problem statement usually managed to give the correct answer. Some of these subjects noted, however, that they were still confused.

The linear training was an attempt to drive a restructuring process prior to making any form of representation. To accept the restructuring rule, subjects must focus on the salient features of each sentence. If a subject fails to focus correctly, differing representations of each sentence may lead the subject to deny the semantic equivalence of the old and the restructured sentences and to doubt the credibility of the restructuring rule. In this regard, note the Experiment 9 data, which showed that many subjects could not see the semantic equivalence between "That man's father is me" and "That man's father is my father's son." Acceptance of a rule may also be related to the availability of sufficient working memory to allow a subject to see a particular relationship.

GENERAL DISCUSSION

Summary Update

The experiments have revealed that almost 100% of subjects are correct in identifying the problem target in

the sentences "That man's father is me" and "That man is my father's son," but as few as 20% are correct on the sentence "That man's father is my father's son." (Where appropriate, assume that all problems are introduced by the assertion "Brothers and sisters have I none.") Moreover, subjects erroneously tend to give the same response to "My father's boss is my father's only son" and "My boss's father is my father's only son." Furthermore, subjects are generally correct on the sentence "My father's son is the father of that man," but generally incorrect on the sentence "The father of that man is my father's son." These results hold for double-modifier problems across a range of content domains. Finally, training was partly successful in improving performance; but few subjects were correct on all test problems, and some subjects were incorrect on all test problems.

The data from the preceding experimental program are relevant to several cognitive issues, including (1) the role of working memory in sentence comprehension, (2) the relationship between problem structure, representation, and performance, and (3) whether mental models (Johnson-Laird, 1983) can account for differential performance across forms of double modifier problems.

Working Memory Capacity and Comprehension

Just and Carpenter (1992) have proposed that "the nature of a person's language comprehension depends on his or her working memory capacity" (p. 124). Language comprehension requires the storage of partial results, and when the processing demands are high, an individual with a relatively small working memory capacity may fail to retain some of these partial results. Analogously to the difficulty of left-branching sentences as opposed to right-branching sentences (Fodor et al., 1974), retention failure may be especially the case when the problem target is in a left-branching phrase, but the problem requires prior processing of information in a right-branching phrase. The evidence from the protocols of subjects who responded "himself" for the traditional brothers and sisters problem are consistent with the smaller working memory hypothesis, with many such subjects arguing that "'my father's son' is me, so the answer has to be 'himself.'" This sort of reasoning seems inexplicable in terms of any logic, but it is consistent with a working memory overload.

However, the data showed that whereas about 20%–40% of the population are likely to solve traditionally structured double-modifier problems prior to some form of training, the accuracy rate rises to around 70% with training. The effect of training raises two related questions. If incorrect responding is related to limitations on working memory capacity, then (1) why can some 30%–50% of initially incorrect subjects learn how to solve double-modifier problems, and (2) is it possible to teach the remaining subjects how to solve such problems? Partial answers to these questions may lie in the kind of problem representation (i.e., strategy) that subjects use.

Problem Structure and Problem Representation

Just as three-term series problems produce differential representations across subjects, spatial or linguistic, so too do double-modifier problems. A crucial question about each problem form is this: Does the mode of representation differentially affect performance? Like the three-term series problem, the double-modifier problem is open to two representational *modes*, verbal or spatial. However, unlike the three-term series problem, in the double-modifier problem each mode can have one of two representational *structures*, linear or equivalence. The representational form was not always clear from the subjects' protocols, but the data from Experiment 6 suggest that the mode of representation may be unrelated to response accuracy. However, the representational structure, linear or equivalence, may be related to response accuracy.

If a subject adopts a linear representational structure, the procedure is a sequence of predication and deduction (e.g., predicate "my father's son" of "father," and then deduce that "that man" is the son of "my father's son"). The position of the target element is important in such a strategy, the evidence from across experiments consistently revealing that having the target element in the subject phrase, a left-branching element, produces the highest error rates, arguably because of the working memory strain associated with the problem target's being in the left-branching element of the sentence.

However, subjects may adopt an equivalence representational structure, either initially, or as a backup strategy when working memory fails during the use of a linear strategy. The difficulty for subjects using an equivalence strategy seems to occur in the attempt to make a correct interpretation from the representation, with the same diagrammatic representation sometimes yielding different responses across subjects. For example, some error-making subjects in the brothers and sisters problem argued that "that man is the son of his father and I'm the son of my father, so the answer is me."

Interestingly, subjects have no difficulty with similarly structured algebraic problems. For example, 24 subjects correctly responded to all four of the following problems and had no preferred order of elements within the problems:

$$x - 1 = 5 - 1;$$

$$-1 + x = -1 + 16;$$

$$10 - 1 = x - 1;$$

$$-1 + 12 = -1 + x.$$

Such problems are scanned in a few saccades, and the matching of x with the correct value can be managed visually. An equivalence representation of double-modifier problems also permits a direct readout of the solution, but the subject may either misrepresent the problem or have problems with the "verbal arithmetic."

Optimal performance was obtained for the brothers and sisters version "My father's son is the father of that man."

The syntactical imbalance between the left and right side of the declarative sentence inhibits the use of an equivalence strategy (although it may still be used), and the placing of the target element at the end of the sentence enables the preliminary predication to be managed with relatively little strain on working memory.

Hence, the difficulties with the traditional double-modifier form of problem seem to be related to working memory limitations for the linear representation, and an incorrect procedure for the equivalence representation. The structure of the problem representation may interact with working memory capacity in such a way that an equivalence strategy may be more efficient for subjects with less working memory capacity. The representational structure chosen by subjects requires further investigation, with special care being required to obtain more complete working protocols than are currently being given.

Successful training in using a linear technique may require a minimum working memory capacity. The linear training, which stresses predication prior to deduction, enables subjects with sufficient working memory capacity to focus on the sentence predicate prior to deducing the problem target. Hence, previous failure in such subjects arguably was attributable to a lack of linguistic skill, a point that is in line with Clark's (1969) emphasis on linguistic processes in reasoning. However, the reporting by some trained subjects that they had a correct answer but did not know why it was correct suggests that adequate working memory space is not a prerequisite for acquiring the predication competency, but may be for comprehending the solution process.

Future experimentation needs to assess the working memory capacities of subjects tested on double-modifier problems and to look for a solution method that may lessen the strain on working memory. For example, efforts to teach an equivalence strategy might be productive with subjects with relatively small working memory capacity.

Mental Models

Johnson-Laird (1983) has proposed the notion of a mental model to help account for both human rationality and human error (Gardner, 1985). The mental models approach has been tested mainly in the area of syllogistic reasoning (e.g., Johnson-Laird, Byrne, & Tabossi, 1989), with experimental evidence supporting the prediction that inferences that require the construction of more than one model will be more difficult than those that require only one model. Johnson-Laird et al., working on a theory of multiply-quantified reasoning based on mental models, proposed that the first step toward a valid inference is the understanding of the premises—"that is, interpretative procedures construct a mental model representing their content" (p. 662).

To test the generality of the mental models approach, it would be of interest to attempt to apply this approach to double-modifier problems. Constructing a mental model

seems to be the initial and perhaps the major task in the attempt to solve a double-modifier problem.

Johnson-Laird (1983, chap. 7) distinguished between mental models and propositional representations. However, it is not immediately obvious whether subjects who use a linear strategy for double-modifier problems are working from a mental model or from a propositional representation. If subjects are generally correct on statements of the form (1) "My father's son is the father of that man," but incorrect on statements of the form (2) "The father of that man is my father's son" (Experiment 8), is the difference in performance attributable to differing mental models, to differing propositional representations, or to a failure to generate a mental model for Statement 2? Furthermore, protocols from subjects who used an equivalence strategy for the problem "That man's father is my father's son" showed that the same equivalence representation, arguably a mental model, led to correct responses from some subjects but incorrect responses from others, suggesting that the construction of a mental model may not be sufficient for a correct response in double-modifier problems.

In general, the double-modifier experimental findings and implications are consistent with Johnson-Laird's (1983) theorizing on mental models, and particularly with his observations on the importance of working memory (p. 124). However, some of the findings raise the questions of when a mental model is necessary for one to obtain a correct solution and of what skills are required to move from a mental model to a correct solution.

CONCLUSION

The traditional double-modifier problem form brings together aspects of reasoning, problem solving, and comprehension. It is an elementary problem form for some subjects, for others a seeming verbal illusion that is dispelled in part by training, but for some an insoluble problem form. Factors associated with performance variability include syntactical structure; possibly, representational structure; the amount of working memory available; and the level of linguistic skill.

Tasks for the future include making precise (1) the relationship between the representational structure of the double-modifier problem and ease of obtaining a correct solution, and (2) the relationship between working memory capacity and performance on double-modifier problems. Taking into account both (1) and (2), a further task is (3) to find a technique enabling all subjects to reach correct solutions for double-modifier problems.

The persistent confusion experienced by many people with the traditional brothers and sisters problem form is illustrated by the author's experiences in several seminars at which the preceding data were presented. On each occasion, at the conclusion of the seminar the author was

approached by participants who asked, "I'm just checking, the correct answer is 'himself,' isn't it?"

REFERENCES

- BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, **3**, 193-200.
- CARPENTER, T. P., CORBITT, M. K., KEPNER, H. S., LINDQUIST, M. M., & REYS, R. E. (1980). Solving verbal problems: Results and implications for national assessment. *Arithmetic Teacher*, **28**, 8-12.
- CHENG, P. W., HOLYOAK, K. J., NISBETT, R. E., & OLIVER, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, **18**, 293-328.
- CLARK, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, **76**, 387-404.
- CUMMINS, D. D., KINTSCH, W., REUSSER, K., & WEIMER, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, **20**, 405-438.
- DEMPSEY, J. (1991). *A rhyme to confound reason: Further exploration*. Unpublished manuscript, Charles Sturt University, School of Humanities and Social Sciences, Wagga Wagga, Australia.
- EVANS, J. ST. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.
- FODOR, J. A., BEVER, T. G., & GARRETT, M. (1974). *The psychology of language*. New York: McGraw-Hill.
- GARDNER, H. (1985). How rational a being? In H. Gardner (Ed.), *The mind's new science: A history of the cognitive revolution* (pp. 360-380). New York: Basic Books.
- HINSLEY, D., HAYES, J. R., & SIMON, H. A. (1977). From words to equations: Meaning and representation in algebra word-problems. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 89-108). Hillsdale, NJ: Erlbaum.
- HUTTENLOCHER, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, **75**, 550-560.
- JOHNSON-LAIRD, P. N. (1972). The three-term series problem. *Cognition*, **1**, 57-82.
- JOHNSON-LAIRD, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- JOHNSON-LAIRD, P. N., BYRNE, R. M. J., & TABOSSI, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, **96**, 658-673.
- JUST, M. A., & CARPENTER, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122-149.
- KNOKE, D., & BURKE, P. J. (1980). *Log-linear models*. Beverly Hills, CA: Sage.
- LARKIN, W., & BURNS, D. (1977). Sentence comprehension and memory for embedded structure. *Memory & Cognition*, **5**, 17-22.
- ORMROD, J. E. (1979). Cognitive processes in the solution of three-term series problems. *American Journal of Psychology*, **92**, 235-255.
- POTTS, G. R., & SCHOLZ, K. W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning & Verbal Behavior*, **14**, 439-452.
- The puzzle book: The 1982 childcraft annual*. (1982). Chicago: World Book.
- REED, S. K., ACKINCLOSE, C. C., & VOSS, A. A. (1990). Selecting analogous problems: Similarity versus inclusiveness. *Memory & Cognition*, **18**, 83-98.
- ROSS, B. H., & KENNEDY, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 42-55.
- SHAVER, P., PIERSON, L., & LANG, S. (1974). Converging evidence for the functional significance of imagery in problem solving. *Cognition*, **3**, 359-375.
- STERNBERG, R. J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, **109**, 119-159.
- WASON, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (Vol. 1, pp. 135-151). Harmondsworth, U.K.: Penguin.
- WASON, P. C., & JOHNSON-LAIRD, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.
- WOODWORTH, R. S., & SELLS, S. B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology*, **18**, 451-460.

NOTE

1. I am grateful to Margaret Intons-Peterson for suggesting that this approach be examined.

(Manuscript received October 31, 1990;
revision accepted for publication December 30, 1992.)