

Differences in mental representations spontaneously adopted for reasoning

DENNIS E. EGAN and DOROTHEA D. GRIMES-FARROW
Bell Laboratories, Murray Hill, New Jersey 07974

Retrospective reporting was used to show that people spontaneously adopt different mental representations for reasoning. In two experiments, subjects solved a large number of three-term series reasoning problems and then gave retrospective reports (verbal protocols, drawings, and forced-choice strategy selections) about their solutions. A majority of subjects' reports could be classified reliably into two groups. One group (abstract directional thinkers) claimed to construct a mental ordering of the three geometric figures used as the terms in the problems. A second group (concrete properties thinkers) claimed to attribute physical properties to mental geometric objects. In both experiments, abstract directional thinkers made few errors and were sensitive to the number of pivot-first premises in a problem. Concrete properties thinkers made more errors and were sensitive to the use of inverse relations and the number of alternations between a relation and its inverse as a problem was read. Quantitative models of the two kinds of reasoning are presented. Implications are discussed concerning theories of reasoning, tests of reasoning, and the usefulness of retrospective reporting as a general method.

Reasoning is considered to be the epitome of human intellectual performance. Because of its obvious importance, reasoning has been studied experimentally using a number of tasks, and some sort of reasoning test is included in virtually every psychometric battery. Among the various reasoning tasks, perhaps none has been studied as frequently as the three-term series problem, also known as the transitive inference problem or linear syllogism (see Sternberg, 1980, for a review).

The exact form of the three-term series problem used in the present studies is best demonstrated by example: "Circle is lighter than square. Triangle is darker than square. Is triangle lighter than circle?" In these problems, the two initial statements, or premises, are composed of two terms and a relation. Each premise has the form: "Term is relation term." One term, the pivot or middle term of the linear order, occurs in both of the premises. In the problems used in the present studies, the question after the premises always had an unambiguous answer and always involved comparing the nonpivot terms.

The main question addressed by this research is whether different people spontaneously adopt different ways of mentally representing problems like the one given in the example. This question has both theoretical and practical importance. For theoretical purposes, it is important to know whether a single theory can possibly account for the performance of each individual

subject solving a reasoning problem. The vigorous debate concerning different representations of three-term series problems (Clark, 1969; Huttenlocher, 1968; Sternberg, 1980) and the apparent ease of influencing subjects to adopt different representations for linear orders (Mayer, 1979; Mynatt & Smith, 1977; Sternberg & Weil, 1980) suggest that individual subjects may spontaneously represent these problems in quite different ways. Moreover, recent studies (MacLeod, Hunt, & Mathews, 1978; Mathews, Hunt, & MacLeod, 1980) have demonstrated that different people adopt different representational strategies for the sentence-picture verification task, which is similar to deductive reasoning.

For practical purposes, it is important to understand what a score on a reasoning test means. One possibility is that a score is a measure of the efficiency of an examinee in executing the same sequence of mental operations as everyone else. This view would be a straightforward application of Sternberg's (1980) theory or any other single theory of reasoning to a testing situation. Another possibility is that a score at least partly reflects a strategic decision made by the examinee to adopt a particular scheme for representing problems. This possibility assumes not only that subjects adopt different representations, but also that some representations result in better performance than others.

The present studies used retrospective reports to identify individual differences in reasoning. Subjects solved reasoning problems and immediately afterward described their solutions. Subjects giving similar reports were classified together. Follow-up analyses then sought to validate this classification scheme by looking for certain patterns in the previously obtained performance data. Johnson (1978) described these steps of classifica-

These two experiments were summarized in a paper presented at the annual meeting of the Psychonomic Society in Phoenix, Arizona, November 8-10, 1979. The able assistance of Peter Meany is gratefully acknowledged. E. Z. Rothkopf gave many helpful comments on previous drafts of this paper. Requests for reprints should be sent to Dennis E. Egan, Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974.

tion and validation in studies identifying different concept learning strategies.

The problems used in these studies were designed to encourage the use of spatial representational schemes. The three terms were always the geometric figures circle, square, and triangle. The relationships used were either the positional pairs above-below, right-left, and front-back or the visual comparative pairs rougher-smoother, darker-lighter, and fatter-thinner.

The mode of presenting problems was thought to be an important variable because some experimental results suggest that reading and listening differentially interfere with the representation of spatial relationships (Brooks, 1967; Ormrod, 1979; Shaver, Pierson, & Lang, 1974). The first experiment included a listening (L) and a listening-plus-reading (LR) mode. The second experiment used only the L mode.

EXPERIMENT 1

Method

Relations and inverses. In this paper, an inverse is defined as that member of a pair of opposing relational words that leads to greater difficulty in these and many other experiments (e.g., Jones, 1970; see also Clark, 1969). For example, the word "above" leads to faster and more accurate performance than the word "below" in a large range of tasks. Consequently, "above" will be referred to as a relation, but "below" will be referred to as an inverse. Similarly, "fatter" is called a relation, but "thinner" is called an inverse. The terms of the problems will be referred to as follows: The "A term" is the initial term in the linear order (e.g., when using above-below, the A term would be the highest); the "B term" is the pivot, or middle, term; the "C term" is the end term of the linear order (lowest when using above-below).

Reasoning problems. A total of 192 problems consisting of 32 problems involving each of the six relation-inverse pairs

formed the basic set of materials. For each relation-inverse pair, 16 problem types (see Table 1) resulted from a 2⁴ factorial combination of the following variables: (1) whether the relation or inverse was used in the premise relating the A and B terms, (2) whether the relation or inverse was used in the premise relating the B and C terms, (3) whether the A-B premise preceded or followed the B-C premise, and (4) whether the relation or inverse was used in the problem question. One version of each problem type ended with the question "Is A relation (inverse) C?"; the other ended with "Is C relation (inverse) A?" By pooling the data from these two problems, each of the 16 problem types was balanced for correct response.

Modes of presentation. Following Brooks (1967), an L and an LR condition were employed. Two complete sets of problems were tape-recorded by a male speaker who began each problem by saying "Next," then read the problem, and then paused 5 sec before beginning the next problem. A typical problem took approximately 8 sec to read. In the LR condition, at the signal "Next," subjects turned over a card with the same problem that the speaker was saying written on it. Subjects were encouraged to read the card silently in pace with the speaker's voice. The problems were printed on 5 x 8 in. white cards. The three geometric terms in each problem were .25-in. figures outlined in black ink, and words were typed in IBM Orator type font. In the L condition, subjects turned over a blank card at each "Next" signal. In both conditions, subjects answered problems by using a pencil to cross out "Yes" or "No" or "?" next to the number of the problem on an answer sheet. Subjects were instructed to mark the "?" rather than guess randomly.

Design. Each subject solved a total of 384 problems, one complete set of problems in both the L and LR conditions. Each set was arranged in eight blocks of 24 problems. Presentation mode was alternated every four blocks, and the sequence of presentation modes was counterbalanced across subjects. Within each block, the 24 problems were selected and ordered randomly under the constraints that each of the six relation-inverse pairs occur four times and that no two successive problems involve the same relation-inverse pair.

Procedure. Subjects were tested individually for one session lasting approximately 2.5 h, including breaks permitted after any block of 24 problems. After some general directions, subjects

Table 1
Structure of Different Problem Types and Corresponding Error Rates in Two Experiments

Problem Type				Error Proportion			
				Experiment 1		Experiment 2	
Premise 1	Premise 2	Questions		AD (N = 7)	CP (N = 5)	AD (N = 42)	CP (N = 18)
A > B	B > C	A > C?	C > A?	.101	.275	.228	.194
B > C	A > B	A > C?	C > A?	.042	.308	.149	.190
A > B	C < B	A > C?	C > A?	.036	.425	.167	.292
C < B	A > B	A > C?	C > A?	.054	.392	.139	.250
B < A	B > C	A > C?	C > A?	.179	.467	.331	.347
B > C	B < A	A > C?	C > A?	.173	.492	.282	.292
B < A	C < B	A > C?	C > A?	.101	.492	.194	.278
C < B	B < A	A > C?	C > A?	.077	.367	.212	.268
A > B	B > C	A < C?	C < A?	.125	.317	.202	.255
B > C	A > B	A < C?	C < A?	.071	.325	.157	.287
A > B	C < B	A < C?	C < A?	.042	.375	.159	.232
C < B	A > B	A < C?	C < A?	.048	.292	.167	.301
B < A	B > C	A < C?	C < A?	.214	.467	.304	.375
B > C	B < A	A < C?	C < A?	.196	.383	.254	.301
B < A	C < B	A < C?	C < A?	.083	.350	.179	.278
C < B	B < A	A < C?	C < A?	.101	.350	.232	.319

Note—Relations (above, right of, in front of, rougher, darker, fatter) are indicated by ">," and inverses (below, left of, in back of, smoother, lighter, thinner) are indicated by "<." Error rates have been combined across the two possible questions for each problem type. AD = abstract directional; CP = concrete properties.

received six practice problems, three in the L condition and three in the LR condition. These problems used relations that did not occur in the rest of the experiment. Subjects were then told the relation-inverse pairs for the experimental problems and began the first block. After the final block of problems was completed, the experimenter interviewed each subject. The subjects were asked how they solved problems involving each relation-inverse pair. The interviews were tape-recorded and later were used to classify subjects into groups.

Subjects. Six male and six female high school students were paid to participate in the experiment.

Results

Retrospective reports. Subjects' reports revealed striking individual differences in the way problem representations were described, particularly those for the three visual comparative relation-inverse pairs, rougher-smoother, darker-lighter, and fatter-thinner. Some subjects claimed to always establish an order for the three geometric figures no matter what relation was used. For these subjects, "rougher," for example, would be identified with one end of a vertical or horizontal scale, as would "darker" and "fatter." Subjects described mentally arranging the geometric figures roughest to smoothest, darkest to lightest, and so forth. Other subjects claimed to attribute physical properties to the geometric figures in the case of the visual comparative relations. For example, these subjects described their representation of a rougher triangle as one having a roughly textured surface, or a lighter circle as being very bright, or a fatter square as being physically large. Distinctions between these subjects on the positional relations were not as clear (see General Discussion).

The rule for assigning subjects to groups was that if a subject described a representation that clearly involved physical properties for any relation, then he or she would be placed in a group labeled "concrete properties" thinkers. Subjects claiming to use directional representations for every relation were placed in a group labeled "abstract directional" thinkers.

This asymmetric rule was motivated by expediency. Classifying reasoners into many subgroups based on the exact representation of specific relations would require extremely large samples and complex analyses. Using this rule, all subjects could be classified into either the concrete properties group ($N = 5$: 3 females, 2 males) or the abstract directional group ($N = 7$: 3 females, 4 males).

Analysis of grouped data. Preliminary analyses of subjects' responses indicated that "?" responses accounted for slightly more than half of the errors in each group of subjects (52.6% for abstract directional subjects and 50.9% for concrete properties subjects). The relative proportion of "?" responses also was approximately equal for different presentation modes and blocks of practice. Consequently, "?" responses and wrong responses were combined for all remaining analyses.

A split-plot analysis of variance was used on the error data, and the method of least squares (Winer, 1971) was applied because of the unequal numbers of subjects in

the two groups. In this analysis, representation group was a between-subjects factor and presentation mode, relation, and problem type were within-subjects factors. The abstract directional group's error rate (10.3%) was significantly lower than that of the concrete properties group (38.0%) [$F(1,10) = 14.06$, $p < .01$]. The relation factor was significant [$F(5,50) = 3.26$, $p < .05$]. Error rates for the six relational pairs were as follows: above-below (17.2%), front-back (18.0%), fatter-thinner (20.4%), darker-lighter (24.1%), right-left (24.1%), and rougher-smoother (26.6%). There were also significant differences among the 16 problem types [$F(15,150) = 5.22$, $p < .001$].

For present purposes, the most important effect was the Representation Group by Problem Type interaction [$F(15,150) = 1.76$, $p < .05$]. This interaction suggests that people who gave different reports made different kinds of reasoning errors.

Validating abstract directional reports. Previous research suggests a direct test of the validity of reports from the abstract directional thinkers. These reports are consistent with spatial models of solving three-term series problems (DeSoto, London, & Handel, 1965; Huttenlocher, 1968; Sternberg, 1980; Trabasso & Riley, 1975) based on the "end anchoring effect." In theory, subjects using a directed spatial representation should find it easier to construct a mental array if the array is "end anchored," that is, constructed from the ends toward the middle. The implication for abstract directional thinkers is that the difficulty of solving a problem should be directly related to the number of premises in the problem that have the middle, or pivot, term stated first.

Each problem was coded by the number of premises having the pivot term first, and this variable was correlated with problem difficulty across the 16 types of problems. When error rates were combined across the six relations, the number of pivot-first premises in a problem accounted for the great majority of variance in problem difficulty for the abstract directional group ($r^2 = .822$, $p < .001$). Mean error rates increased monotonically over problems with zero, one, or two pivot-first premises (error rates were 4.5%, 8.8%, and 19.0%, respectively). This relationship was also statistically significant for each relation analyzed separately in the abstract directional group. On the other hand, the concrete properties group showed a much smaller, non-significant relationship between the number of pivot-first premises and error rate ($r^2 = .176$, $p > .10$). Mean error rates were not a monotonic function of the number of pivot-first premises (error rates were 37.1%, 34.8%, and 45.2%). The relationship between pivot-first premises and error rates did not approach significance for any individual relation in this group.

Validating concrete properties reports. Two potential sources of difficulty for concrete properties thinkers were suggested by the following rationale. Concrete

properties thinkers claimed to generate qualitatively different representations for a relation and its inverse. This qualitative difference for the concrete properties group suggests that those subjects may show a greater effect of using inverses than do the abstract directional subjects. Concrete properties thinkers may also have more difficulty processing changes in relations as they read the problems. That is, if a relation and its inverse have different concrete representations, it may prove difficult to alternately access the different representations.

The hypothesis that the number of changes between relation and inverse would influence error rates of the concrete properties thinkers received some support. The number of changes accounted for a significant portion of the problem variance for concrete properties thinkers ($r^2 = .255$, $p < .05$) but not for the abstract directional group ($r^2 = .049$, $p > .10$). Mean error rates increased monotonically with the number of alternations for concrete properties thinkers (error rates were 32.1%, 39.0%, and 41.9% for problems with zero, one, and two alternations, respectively), but not for abstract directional thinkers (corresponding error rates were 8.2%, 12.0%, and 11.8%). The hypothesis that the number of difficult relations or inverses would affect the error rates of the concrete properties group received no support. This variable was not significantly correlated with errors for either concrete properties thinkers ($r^2 = .026$) or abstract directional thinkers ($r^2 = .011$).

Discussion

Two conclusions follow from this experiment. One is that retrospective reports following a reasoning task can be used to classify subjects. The second conclusion is that the categories tentatively appear to have some validity. Reasoning errors by the group labeled abstract directional thinkers were influenced by a variable that is directional in character. The group labeled concrete properties thinkers was not as sensitive to the directional problem variable but did show some sensitivity to the number of alternations that occurred between a relation and its inverse in the statement of problems. The large overall difference in error rates favoring abstract directional thinkers may reflect the greater efficiency of generating mental arrays over generating concrete mental representations.

Since presentation mode produced no main effect or interaction, it is possible that subjects ignored the visual input in the LR condition. Only the L mode was used in Experiment 2.

EXPERIMENT 2

The second experiment was designed to replicate the first study with a larger number of subjects. Additionally, two supplementary techniques were used to collect retrospective reports. Subjects were asked to draw representations of problems and to choose between written

descriptions of the two types of problem representations found in Experiment 1. Agreement among various methods of reporting provides one measure of the reliability of retrospective reports.

Method

Design. Each subject solved one set of the 192 problems previously described. The problems were again organized in blocks of 24 and were administered in one of four different randomized orders.

Procedure. Eight groups were tested, ranging in size from 12 to 14 subjects/group. Each of the four presentation orders was given to two of the groups. The experiment was held in a classroom, and each group participated in a single session lasting approximately 1.5 h. The experiment began with subjects reading directions and then solving six practice problems. All problems were presented on a tape playback machine positioned so that all subjects easily could hear the problems. Following the practice problems, subjects began the first block of experimental problems. A 30-sec break was given after each block of 24 problems, and a break of approximately 5 min was given midway through the session.

Retrospective reports. Subjects' reports were collected in three phases immediately following the completion of the reasoning problems. First, subjects were asked to write a description of how they solved problems involving each of the relation-inverse pairs. The descriptions were essentially written versions of the oral retrospective reports collected in Experiment 1. Subjects were asked next to make a drawing of their mental representations for six different problems, one involving each of the relation-inverse pairs. Finally, written descriptions of the two ways of solving reasoning problems identified in Experiment 1 were handed out. Subjects were then asked to choose which of the descriptions more closely characterized their own way of solving the problems.

Subjects. Fifty male and 50 female high school students were paid to participate.

Results

Written descriptions. The written retrospective reports were classified by the two authors, working as independent judges without knowledge of the performance data. Each judge classified each subject according to the rule developed earlier. If the written description showed evidence of representing any problems by concrete properties, the subject was classified in one category (concrete properties thinker). Subjects claiming to use a directional representation for every relation were classified in another category (abstract directional thinker). If the written description was either unclear or described a representational scheme other than the two found in the first study, the subject was classified into a third category (other/not clear).

The two judges agreed on the categorization of 82% of the individual subjects. For another 17% of the cases, one judge classified a subject as one of the two types of thinkers identified, but the other judge classified the subject as "other/not clear." In 1 of the 100 cases analyzed, the judges directly disagreed, one judge classifying the subject as an abstract directional thinker and the other, as a concrete properties thinker. A reexamination of that subject's reports suggested that the subject was in fact combining the two representations. All further

analyses were carried out using the 60 subjects who were independently classified by the two judges as using one of the two representations identified earlier. There were 18 subjects (10 males, 8 females) classified as concrete properties thinkers and 42 subjects (19 males, 23 females) classified as abstract directional thinkers. Examples of the written reports given by the subjects are shown in Table 2.

Drawings. Subjects' drawings showed differences that corresponded to their written reports. Some subjects represented problems by drawing geometric figures with altered physical properties. For example, "rougher" might be represented as a spotted surface or a serrated outline, "darker" as shading, and "fatter" as greater physical size. Other subjects used standard geometric objects throughout their drawings but arranged the objects left to right or top to bottom on scales. The 60 subjects previously classified by consensus were again classified as either drawing only directional representations or drawing at least one representation with altered physical properties. For 56 of the 60 subjects (93.3%), the classification of the drawings corresponded to the previous classification of the written reports. Thus, these verbal and nonverbal modes of describing problem representations proved to be consistent.

Forced-choice responses. The problem solving descriptions chosen by subjects were analyzed in a way similar to the drawings. Three of the 60 subjects (1 concrete properties thinker and 2 abstract directional thinkers) claimed that neither description matched their actual problem solving process close enough to make a choice. A total of 51 of the 60 subjects (85%) chose descriptions consistent with the previous classifications based on written reports. While this method of forced choice shows substantial agreement with the written reports, the agreement might have been greater if the choice had

been limited to performance on problems involving visual comparative relations (see General Discussion).

Analysis of grouped data. As in Experiment 1, "?" responses accounted for approximately the same proportion of total errors in each group of subjects (40.3% for abstract directional thinkers and 38.0% for concrete properties thinkers). The reduced use of "?" and generally lower error rates in this study may be due to the fact that subjects did not have to manipulate stimulus cards as they did in the first study.

A split-plot analysis of variance using the method of least squares was applied to the error data. The pattern of results was similar to that found in Experiment 1. First, the abstract directional group's error rate (21.0%) was lower than that for the concrete properties group (27.9%), this time at a marginal level of significance [$F(1,58) = 3.60, .10 > p > .05$]. The difference in error rate between the two groups tended to increase from the first half (5.1%) to the second half (8.7%) of the session. The larger group difference later in practice is consistent with the larger difference between groups in Experiment 1, in which twice as many problems were used. Second, the relation factor was significant [$F(5,290) = 43.16, p < .001$]. Error rates for the pairs of relational words were as follows: above-below (13.0%), front-back (15.0%), right-left (24.3%), fatter-thinner (25.4%), darker-lighter (29.4%), and rougher-smoother (31.2%). Third, there were again significant differences among the 16 problem types [$F(15,870) = 11.72, p < .001$]. In this experiment, the Relation by Problem Type interaction also was significant [$F(75,4350) = 2.18, p < .001$].

Problem difficulty for different representations. The Representation Group by Problem Type interaction was significant [$F(15,870) = 2.06, p < .01$], so this interaction was analyzed as before, and the results are pic-

Table 2
Examples of Written Retrospective Reports

Subject	Abstract Directional Thinkers	Subject	Concrete Properties Thinkers
Problem Relation: Rougher-Smoother			
005	"Rather than imagining a rough/smooth figure, I put the figures in a horizontal line, in my mind, in the order of left/right rather than rough/smooth."	008	"I also drew a picture, and if something was rough—I would put craters in it in my mind—smooth was just plain white."
049	"I pictured the objects in my mind in a line of sequence."	098	"The picture came to mind of corners and smooth edges, then the question was solved."
Problem Relation: Darker-Lighter			
003	"I set up a scale with the lightest on the far right and darkest on the far left and placed the figures on their appropriate spots."	062	"In my mind, I 'colored in' the object that was darkest."
051	"Placed them in a line up and down, darkest being on top."	080	"I listened to the problem and tried to solve it mentally at times picturing the objects colored in or not."
Problem Relation: Fatter-Thinner			
086	"I also used a mental horizontal grid for this relation with the left side of the grid being the 'thin end' and the right side the 'fat end.'"	022	"This [fatter-thinner problem] was hard. I had to think of the shapes as squeezed or pulled."
099	"Put shapes in order from thinner to fatter."	100	"Made them [the figures] fatter and thinner in my head."

tured in Figure 1. The number of pivot-first premises in a problem influenced error rates of abstract directional thinkers more strongly ($r^2 = .718, p < .001$) than those of concrete properties thinkers ($r^2 = .203, .10 > p > .05$). For each of the six individual relations, the abstract directional thinkers were more sensitive than the concrete properties thinkers to this directional variable.

The effect of the number of pivot-first premises again was monotonically increasing for abstract directional thinkers, but nonmonotonic for concrete properties thinkers (see Figure 1). There was, however, a reliable difference for concrete properties thinkers between problems having two pivot-first premises and problems with either zero or one such premise [$t(14) = 2.89, p < .02$]. This result can be rationalized as follows. For all problems having two pivot-first premises (e.g., "Circle is lighter than square. Circle is darker than triangle."), a concrete properties thinker might have great difficulty if the B term (circle in the example) is represented first by the property suggested by a relation and then again by the property suggested by its inverse. Data from Experiment 1 showed the same effect [$t(14) = 2.90, p < .02$]. This effect of two pivot-first premises is incorporated in the model for concrete properties thinkers developed in the General Discussion section.

As in Experiment 1, the number of changes between a relation and inverse had a significant effect for the concrete properties group ($r^2 = .270, p < .05$), but not for the abstract directional group ($r^2 = .042, p > .10$). Additionally, in this experiment, the number of inverses had a significant effect for the concrete properties group ($r^2 = .263, p < .05$), but not for the abstract directional group ($r^2 = .005, p > .10$). The effects of these two problem factors are shown graphically in Figure 1. The concrete properties group shows a monotonic relationship between error rate and both the number of difficult relations and the number of changes in relation. The

abstract directional group shows less systematic relationships between error rates and each of these two factors.

GENERAL DISCUSSION

Classification of Reasoners

Two aspects of the method for classifying subjects require further explanation. First, subjects were placed in the concrete properties group if even one of their reports for the visual comparative relations suggested a concrete representation. In fact, a subject reporting a concrete representation for one relation was very likely to report using such a representation for more than one relation. For example, 17 of the 18 subjects identified as concrete properties thinkers in Experiment 2 gave reports having a concrete character for two or more relations.

Second, although subjects were classified on the basis of their reports about the visual comparative relations, the patterns of data within each group were similar for the positional and visual comparative relations. A possible explanation of this finding is that abstract directional and concrete properties thinkers use different representations for positional problems, but these representations are described in language that differs only in very subtle ways. For example, concrete properties thinkers might represent the spatial position of objects by means of a visual image, whereas abstract directional thinkers might represent spatial position by means of an abstract order-preserving scale. Despite their qualitative differences, these two forms of representation for positional problems could be described using very similar words (e.g., for an above-below problem: "I thought of the objects from top to bottom"). This sort of ambiguity is a difficulty inherent in the present method of retrospective reporting.

A detailed analysis of the actual words used by the

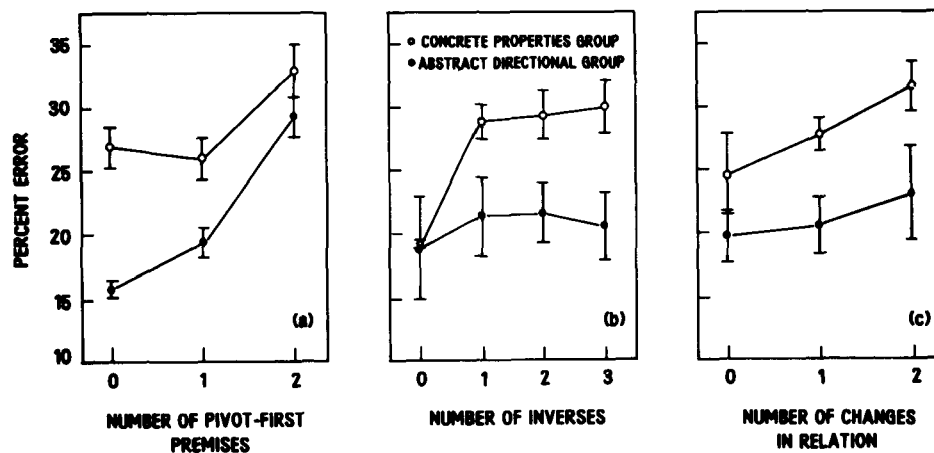


Figure 1. Error rates for concrete properties and abstract directional thinkers as a function of three problem factors: (a) the number of pivot-first premises, (b) the number of difficult relations or inverses, and (c) the number of changes in relation as the problem was read. Brackets represent ± 1 standard error computed over problems for each point.

two types of thinkers tends to support the conjecture that the different groups represented positional problems in different ways. Key words were identified that suggest either the use of an image (variants of the words "picture" and "draw") or the use of an abstract, order-preserving scale (variants of the words "put," "order," "line," and "horizontal/vertical"). Concrete properties thinkers tended to use more words suggesting an image than did abstract directional thinkers (means were 2.39 and 1.81 words/subject, respectively). On the other hand, abstract directional thinkers tended to use more words suggesting a scale than did concrete properties thinkers (means were 1.79 and .72 words/subject, respectively). This interaction of Representation Group by Word Type was reliable [$F(1,58) = 4.39, p < .05$], but it should be stressed that this analysis was entirely post hoc. Further research might obtain a clearer separation of representational strategies used on positional problems by probing with questions that involve the key words analyzed.

Two Processing Models

The abstract directional model. A process model for abstract directional thinkers was developed to account for their reports and pattern of errors. Abstract directional thinkers (see Table 3) are assumed to encode the first premise and establish a mental scale for a problem. Then, the two terms stated in the first premise are arranged on the scale, the grammatical subject being placed first. The second premise is then encoded, and the subject searches for the third, or missing, term. This search is easier if the third term is the grammatical subject rather than the object of the second premise (Huttenlocher, 1968). The value of the "search" parameter (0 or 1 respectively) reflects this difficulty, and accounts for the effect of starting the second premise with the pivot term. Next, the third term is positioned on the mental scale, and it is assumed that there are three distinct cases for this operation. The easiest case occurs when the third term is placed next in the sequence established by the first two terms. For example, if the first two terms are arranged smooth → rough, positioning the third term is easiest if it is the roughest. If the first two terms are placed rough → smooth, then positioning

the third term is easiest when it is the smoothest. Two more difficult cases exist and correspond to problems beginning with a pivot-first premise. In the easier of these cases, the third term does not fall next in sequence, but instead, it must be placed at the end of the scale associated with the relation. In the remaining, most difficult case, the third term again does not fall in sequence but must be positioned at the end of the scale associated with the inverse.

The rationale for the difference in difficulty among the three cases of positioning the third term should be the subject of further research. One possibility is that the first two terms may be reordered in the preferred direction, as suggested by Huttenlocher (1968), but only if the third term does not follow naturally in the sequence established by the first two terms. For present purposes, the three cases of positioning the third term are simply reflected in three values of the "position" parameter (0, 1, or 2, respectively).

The concrete properties model. The model for concrete properties thinkers suggests that these subjects generate and compare images of objects having the stated properties. For each premise, concrete properties thinkers (Table 4) are assumed to encode the premise and then generate an image pair in which the grammatical subject takes on the property stated in the premise and the grammatical object remains neutral. After two such pairs have been generated, the question is encoded, and then the two image pairs are scanned for the answer. Differences in difficulty among problems are assumed to arise from three sources. One kind of difficulty has to do with whether the relation or inverse is used in each premise. Using an inverse presumably makes the appropriate image pair more difficult to generate. For a given problem, the parameter "generate" takes on a value equal to the number of difficult images required (0, 1, or 2). The parameter "encode" reflects the difficulty of alternately accessing a relation and its inverse. The value of this parameter equals the number of alternations between a relation and inverse as a problem is read (0, 1, or 2). Finally, the parameter "scan" reflects the difficulty in dealing with images that are inconclusive. As noted previously, problems having two pivot-first premises result in the B term's taking on a property in one image pair

Table 3
Model for Abstract Directional Subjects

Process	Problem Factors	MP
(1) Encode Premise 1		
(2) Establish abstract scale		
(3) Arrange first two terms placing grammatical subject first on scale		
(4) Encode Premise 2		
(5) Find third term	Is third term grammatical subject or object?	Search
(6) Position third term	Does third term fall in "natural" next position?	Position
(7) Encode question		
(8) Scan the scale		
(9) Respond		

Note—MP = model parameter.

Table 4
Model for Concrete Properties Subjects

Process	Problem Factors	MP
(1) Encode Premise 1		
(2) Generate Image Pair 1 by assigning property to grammatical subject	Is difficult (inverse) relation used?	Generate
(3) Encode Premise 2	Is relation the same as that in Process 1?	Encode
(4) Generate Image Pair 2 by assigning property to grammatical subject	Is difficult (inverse) relation used?	Generate
(5) Encode question	Is relation the same as that in Process 3?	Encode
(6) Scan Images	Are the images conclusive?	Scan
(7) Respond		

Note—MP = model parameter.

and then taking on the inverse property in the other pair. Such problems produce inconclusive image pairs in which the A and C terms are both neutral. Confronted with this type of problem, concrete properties thinkers may guess or reformulate one of the premises to arrive at an answer. The scan parameter has the value 1 for such problems and 0 otherwise.

Testing models using multiple regression. The proportion of variance in problem difficulty uniquely associated with each parameter in each model was determined by stepwise multiple regression. The data are shown in Table 5. For both experiments, the abstract directional model was the better predictor of performance for abstract directional subjects and the concrete properties model was the better predictor for concrete properties subjects. If errors on the various problem types are combined across experiments, the abstract directional model accounts for 90.3% of the variance in problem difficulty for abstract directional thinkers (the concrete properties model accounts for 80.4% of the variance for this group). Both the search and position parameters account for significant and unique portions of variance in problem difficulty for abstract directional thinkers. The concrete

properties model accounts for 80.1% of the variance in problem difficulty for the combined concrete properties thinkers (the abstract directional model accounts for 38.7% of the variance for this group), and each of the parameters scan, generate, and encode accounts for significant and unique variance.

The estimated reliability¹ of the abstract directional data combined across experiments was .951, so the abstract directional model ($R^2 = .903$) accounted for .903/.951, or 95.0%, of the reliable variation in the problem type data for those subjects. The fit of the concrete properties model ($R^2 = .801$) actually slightly exceeded the theoretical upper limit of the reliability of the combined concrete properties data (estimated reliability was .727).

It is important to note that certain parameters of the two models are correlated in the 16 problem types used. The most important example of this confounding occurs for the parameter scan in the concrete properties model that is correlated with both the search and position parameters in the abstract directional model. These correlations account for the contribution of scan to variance in problem difficulty for abstract directional subjects. This interpretation of the contribution of scan is consistent with the facts that it is only the parameter in the concrete properties model that correlates with performance for abstract directional subjects and that the two-parameter abstract directional model accounts for more variance in that group than does the three-parameter concrete properties model.

Tests of parameters using analysis of variance. The effects of model parameters also were tested using analyses of variance. For this purpose, error rates on classes of problems reflecting the values of the model parameters were analyzed. The general result was that representation groups interacted, as expected, with three of the five theoretical parameters.

For the abstract directional model, mean error rates for each subject in the two experiments were obtained for each of the six search by position parameter value combinations. Abstract directional subjects showed a larger increment in error rate due to searching than concrete properties subjects did [$F(1,350) = 12.60$, $p < .001$]. The interaction of Representation Groups by Position did not approach significance ($F < 1.0$), although

Table 5
Proportion of Variance in Problem Difficulty
Attributable to Parameters of Two Models

Parameter	Experiment					
	1		2		1 + 2	
	A	C	A	C	A	C
Abstract Directional Model						
Search	.576†	.021	.668†	.105	.668†	.076
Position	.288†	.305*	.208†	.201	.235†	.311*
ΣR^2	.864†	.326	.876†	.306	.903†	.387*
Concrete Properties Model						
Scan	.794†	.372*	.724†	.374*	.768†	.474†
Generate	.001	.185*	.016	.166*	.011	.221†
Encode	.024	.079	.024	.087	.025	.106*
ΣR^2	.819†	.636†	.764†	.627†	.804†	.801†

Note—These proportions are increments in R^2 values due to each parameter. The order in which the parameters are given corresponds to the step at which they entered the regression equation for the group of subjects appropriate to a particular model. A = abstract group; C = concrete group.

* $p < .05$. † $p < .01$.

position had a strong main effect ($[F(2,350) = 24.93, p < .001]$). The lack of interaction may have been due to the fact that the position parameter was positively correlated with both the scan and generate parameters. Thus, apparent increases in mental positioning difficulty for concrete properties subjects may have been due to actual increases in the difficulty of scanning and generating images.

Analysis of the concrete properties model parameters was complicated by correlations within and across models. For this reason, the problems were divided into subsets. In the first subset (eight problems), the value of scan was 0, the value of encode was either 0 or 1, and the value of generate was either 0 or 2. This subset permitted an assessment of the encode and generate parameters independent of the influence of scan and search. The only remaining correlation was that between generate and position, and this correlation worked against finding the hypothesized interaction. Representation groups interacted in the expected way with both encode [$F(1,210) = 4.04, p < .05$] and generate [$F(1,210) = 5.73, p < .05$]: Concrete properties subjects showed larger increases in error rates than did abstract directional subjects when problems required more difficult encoding and image generation.

In the complementary subset of eight problems, we could analyze the scan parameter while controlling for encode and generate. For these problems, the effect of scan actually was greater for abstract directional subjects than for concrete properties subjects, as suggested by Table 5. Since scan was strongly correlated with both parameters of the abstract directional model (particularly in this subset), this finding is difficult to interpret.

Implications for Theories of Reasoning

These results suggest that no single theory of solving three-term series reasoning problems can account for the data of all subjects, unless that theory includes an account of why individual subjects adopt one representational scheme rather than another. In particular, different "componential" analyses of reasoning (Sternberg, 1980) may be appropriate for different groups of subjects. Differences in the way people make transitive inferences cannot be described simply by different parameters of a single information processing model.

The two categories of reasoners identified in these studies may think about problems in ways consistent with two strategies for representing linear orders proposed by Mynatt and Smith (1977). Mynatt and Smith suggested that some subjects represent linear orders by a "rehearsal" strategy, whereas others use an "imagery" strategy. Rehearsal may correspond to the representation employed by abstract directional thinkers. The imagery strategy may be closer to the representational scheme used by concrete properties thinkers. Mynatt and Smith set up different conditions to cue groups of people to use one or the other strategy. The present

studies may have identified the same strategies employed spontaneously by different people for a common set of reasoning problems.

There is a similarity between the factors found to influence the performance of the two groups of reasoners in these studies and the spatial and linguistic factors involved in previous debates over theories of solving three-term series problems. It is tempting to conclude that the heterogeneity of mental representations found in these studies "resolves" the theoretical debate about reasoning. Such a conclusion is almost certainly an oversimplification. The procedures used in the present studies have special characteristics that may not generalize to other procedures commonly employed. Among the critical procedural characteristics of the present studies are the following: (1) The dependent measure was accuracy under a deadline, (2) the terms and relations were selected for ease of spatial representation, (3) an entire problem was presented before a pause allowing for a response, (4) negatives were not used in any problem, (5) subjects were not allowed to write down anything but their answers, and (6) subjects were not discarded from the analyses because of high error rates. Altering one or more of these procedural characteristics may change either the nature of the retrospective reports obtained or the pattern of performance within a group of subjects giving the kind of reports found in these studies.

Implications for the Assessment of Reasoning

The present results suggest that scores on standard tests of reasoning may be much more complicated than an indication of the efficiency with which a particular examinee executes a sequence of mental operations common to all examinees. In the present studies, concrete properties thinkers typically scored lower than abstract directional thinkers on the test-like tasks used. This result implies that a score on a reasoning test may at least partially indicate what sort of mental representation an examinee adopted for an entire set of problems.

As French (Note 1) hypothesized and present results confirm, different problem solving styles in a population mean that item difficulty and the factor composition of reasoning tests may differ across groups of subjects. The most important remaining question is why individuals spontaneously adopt different mental representations on a reasoning test and whether this account of reasoning test scores permits a greater understanding of the correlates of such scores.

The Method of Retrospective Reporting

The present studies provide a direct assessment of the usefulness of retrospective reports on cognitive processes. Some previous researchers have regarded reports as being useless; others have assumed the validity of reports without an empirical test. Most commonly, retrospective reports have been used in a very informal way to suggest or corroborate hypotheses. In the present

studies, however, retrospective reports played a critical role as the basis for classifying subjects. The resulting classification scheme was reliable as assessed by inter-judge agreement and agreement with nonverbal methods of reporting. The classification of subjects based on retrospective reports also was valid. Subjects giving different retrospections exhibited different patterns of reasoning errors consistent with their retrospections.

In further research, the validity of retrospective reports for three-term series reasoning problems has been tested in other ways. Abstract directional thinkers have been found to score higher than concrete properties thinkers on tests of spatial (but not verbal) aptitude (Grimes, 1980; Egan, Note 2). Compared with abstract directional thinkers, concrete properties thinkers have proved to be more sensitive to the consistency between the appearance of terms as suggested by a problem's wording and the actual appearance of the terms drawn on stimulus cards (Grimes, 1980). Patterns of subjects' reasoning errors can be altered in predictable ways by directions to adopt the abstract directional or concrete properties strategies described in Tables 3 and 4 (Egan, Note 2).

General conditions for using the method of retrospective reporting remain to be established (see Nisbett & Wilson, 1977; Smith & Miller, 1978). Ericsson and Simon's (1980) recent model of verbal reporting suggests that the method used in the present studies had both desirable and undesirable aspects. On one hand, the present method probed for fairly specific memories, a desirable feature according to Ericsson and Simon (1980). In Experiment 2, this method of verbal reporting was followed by even more specific nonverbal probes, another desirable feature.

A possibly undesirable aspect of the present method was that the verbal reports were given retrospectively rather than concurrently with reasoning. The rationale was to collect reasoning performance data (in this case, error rates under a deadline) that could be used to validate different reports and to use conditions for the reasoning task similar to those employed by other investigators. Concurrent reporting probably would have affected the time course of reasoning if not the representational strategy selected (see Ericsson & Simon, 1980). Granting the potential drawback of memory failure, retrospective reporting in these studies proved to be a reliable and ultimately valid method of examining higher-level mental processing.

SUMMARY

These studies provide evidence for two informal observations about deductive reasoning that have previously lacked clear empirical support. One naive observation now supported by data is that different people seem to reason in qualitatively different ways. Even if the two models of reasoning developed here are not entirely correct, the more important result is that they are different.

These studies strongly suggest that the performance of different subjects cannot be characterized simply by different parameters of a single information processing model. The other naively compelling but formerly untested observation is that people seem to be able to report how they reason. The present results make it clear that verbal reports about reasoning carry information. Although no subject described a comprehensive model of reasoning, retrospective reports did contain enough information to classify subjects reliably, and the resulting classification scheme apparently was valid.

REFERENCE NOTES

1. French, J. W. *The relationship of problem-solving styles to the factor composition of tests* (ETS RB 63-15). Princeton, N.J.: Educational Testing Service, 1963.
2. Egan, D. E. *Modifying strategies for reasoning*. Manuscript submitted for publication, 1982.

REFERENCES

- BROOKS, L. R. The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, 1967, 19, 289-299.
- CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, 76, 387-404.
- DESOTO, C. B., LONDON, M., & HANDEL, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 1965, 2, 513-521.
- ERICSSON, K. A., & SIMON, H. A. Verbal reports as data. *Psychological Review*, 1980, 87, 215-251.
- GRIMES, D. D. *Individual differences in mental representations of three-term series problems*. Unpublished doctoral dissertation, Rutgers University, 1980.
- HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, 75, 550-560.
- JOHNSON, E. S. Validation of concept-learning strategies. *Journal of Experimental Psychology: General*, 1978, 107, 237-266.
- JONES, S. Visual and verbal processes in problem solving. *Cognitive Psychology*, 1970, 1, 201-214.
- MACLEOD, C. M., HUNT, E. B., & MATHEWS, N. N. Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior*, 1978, 5, 493-508.
- MATHEWS, N. N., HUNT, E. B., & MACLEOD, C. M. Strategy choice and strategy training in sentence-picture verification. *Journal of Verbal Learning and Verbal Behavior*, 1980, 19, 531-548.
- MAYER, R. E. Qualitatively different encoding strategies for linear reasoning premises: Evidence for single association and distance theories. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, 5, 1-10.
- MYNATT, B. T., & SMITH, K. H. Constructive processes in linear order problems revealed by sentence study times. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 357-374.
- NISBETT, R. E., & WILSON, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, 84, 231-259.
- ORMROD, J. E. Cognitive processes in the solution of three-term series problems. *American Journal of Psychology*, 1979, 92, 235-255.
- SHAVER, P., PIERSON, L., & LANG, S. Converging evidence for the functional significance of imagery in problem solving. *Cognition*, 1974, 3, 359-375.

- SMITH, E. R., & MILLER, F. D. Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review*, 1978, **85**, 355-362.
- STERNBERG, R. J. Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, 1980, **109**, 119-159.
- STERNBERG, R. J., & WEIL, E. M. An aptitude-strategy interaction in linear syllogistic reasoning. *Journal of Educational Psychology*, 1980, **72**, 226-239.
- TRABASSO, T., & RILEY, C. A. On the construction and use of representations involving linear order. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J.: Erlbaum, 1975.
- WINER, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

NOTES

1. The reliability of the data was estimated in the following way. For each group of subjects, error rates on the 16 problem types were correlated across Experiments 1 and 2. These correla-

tions ($r = .907$ for abstract directional subjects and $r = .571$ for concrete properties subjects) are estimates of the proportion of reliable variance within each experiment for each group of subjects. These estimates compare closely to the R^2 values for the fit of the abstract directional model to the abstract directional subjects in the two experiments ($R^2 = .864$ and $.876$) and the fit of the concrete properties model to concrete properties subjects ($R^2 = .636$ and $.627$). It was also possible to estimate the reliability of the combined data from Experiments 1 and 2 by applying the Spearman-Brown formula. The estimate for the combined abstract directional data was $r = .951$, and for concrete properties subjects, it was $r = .727$. These last estimates are comparable to the fit of the abstract directional model ($R^2 = .903$) and the concrete properties model ($R^2 = .801$) to the combined data of the respective groups. This analysis suggests that the fit of the models is somewhat limited by the reliability of the data, particularly in the case of concrete properties subjects.

(Received for publication April 15, 1981;
revision accepted February 4, 1982.)