

Spatial Comprehension and Comparison Processes in Verification Tasks

ROBERT J. GLUSHKO

University of California, San Diego

AND

LYNN A. COOPER

Cornell University

Two experiments use the sentence-picture verification paradigm to study encoding and comparison processes with spatial information. Subjects decided whether a spatial description of a figure or a geometric figure matched a second figure. Three critical results (the effects of display complexity, the effects of lexical markedness, and the relative speeds of "same" and "different" responses) covaried across four experimental conditions. These results demonstrate that task-specific variables can be the primary determinants of how subjects verify sentences. When the two displays were presented successively and subjects took as much time as they needed to prepare for the test figure, verification time was not affected by the pictorial complexity of the test figure or by the markedness of the relational terms used in the descriptions, and "same" responses were faster than "different" responses. When subjects had less time to study the spatial description before the test picture appeared, the effects of complexity and lexical markedness on verification time increased and were largest when the two displays appeared simultaneously; concurrently, "differents" became faster than "sames." This pattern of results is not easily handled by current models for sentence-picture verification.

This research was funded primarily by National Science Foundation Grant BMS 75-15773 and National Institutes of Mental Health Small Grant MH 25722-01 to the second author. Experiment 2 was supported by National Science Foundation Grant BMS 76-15024 to David E. Rumelhart. Some pilot work by the first author was funded by National Science Foundation Grant GB-31971X to Roger Shepard at Stanford University. The first author held a National Science Foundation Graduate Fellowship while this paper was written. We thank our colleagues at UCSD and elsewhere whose discussion and critical comments shaped and improved this paper; in particular, we acknowledge the help of Jack Catlin, Jim Cunningham, Colin MacLeod, Jay McClelland, Jeff Miller, Allen Munro, Don Norman, Steve Palmer, Peter Podgorny, Dave Rumelhart, Arty Samuel, Roger Shepard, Al Stevens, and an anonymous reviewer. Requests for reprints may be sent to either author: Robert J. Glushko, Department of Psychology C-009, University of California, San Diego, La Jolla, CA 92093; Lynn A. Cooper, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853.

A major goal of cognitive psychologists is to understand the nature of the internal representations and processing operations that mediate the comprehension of spatial information. Many researchers use verification or comparison tasks to study how people understand simple declarative sentences which describe a spatial configuration.

There are two general experimental approaches which characterize much of this research on spatial comprehension. In studies using the psycholinguistic "sentence verification" paradigm the subject determines whether a sentence or a name correctly describes a spatial display (e.g., Carpenter & Just, 1975; Clark, Carpenter, & Just, 1973; Clark & Chase, 1972, 1974; Just & Carpenter, 1975; Seymour, 1973a, 1974a, d, 1975; Tversky, 1975). This task is typically used by psycholinguists to study the effects of variables such as syntactic complexity, negation, and lexical markedness (see Clark, 1969) on verifying spatial sentences. In "description matching" studies, experimenters primarily interested in memory processes compare matching a spatial description to a subsequently-presented visual figure with matching two sequentially-presented visual figures (e.g., Cohen, 1969; Nielsen & Smith, 1973; Santa, 1977; Seymour, 1973b, 1974b, c; Smith & Nielsen, 1970; Tversky, 1969). These researchers study the temporal properties of picture and sentence representations, and investigate the possibility of translating or recoding one representation into the other.¹

Researchers in spatial comprehension typically use the latency of a verification decision to infer both the nature of the representation underlying a spatial judgment and the nature of the comparison operations performed upon the representation. In the sentence verification paradigm, the response latency is the time to decide whether a sentence is "true" or "false" of a picture. In the description matching paradigm, this latency is the time to decide whether two displays are the "same" or "different."

Since the sentence verification and description matching paradigms resemble one another, researchers in each tradition have proposed models for their particular task and have suggested that these models might be extended to the other paradigm as a general model of spatial comprehension and comparison. The two most influential attempts at characterizing the comprehension of spatial information have resulted in the *propositional* and *dual code* models. The propositional model for spatial comprehension was developed for the sentence verification task, while the dual code model emerged from the description-matching paradigm. Unfortunately, the

¹ Thus, our analysis excludes versions of the sentence verification procedure which investigate purely syntactic effects on sentence-picture comparison (e.g., Glucksberg, Trabasso, & Wald, 1973; Olson & Filby, 1972; Wannemacher, 1974, 1976) and variants of the visual matching paradigm which do not study descriptions (e.g., Cooper, 1976; Paivio & Bleasdale, 1974).

typical patterns of results from the two approaches are somewhat inconsistent, and neither class of model seems capable of unifying the two lines of research.

We believe that previous investigators have underestimated the constraints imposed by the particular verification or matching task on the pattern of results, and thereby have overestimated the generality of the resulting models. We aim to explain some of the empirical and theoretical inconsistency in previous work by showing how the outcomes of spatial comprehension studies are greatly determined by the experimental tasks used. Therefore, rather than simply listing the conflicting results that have arisen from previous studies, we organize our presentation in terms of the characteristic outcomes from the different experimental procedures.

Propositional models were proposed in a number of sentence verification studies to account for a regular cluster of linguistic effects on verification latency. In general, descriptions that are linguistically complex (e.g., descriptions that contain negatives) take longer to verify than simple affirmative descriptions. The most reliable result is that sentences containing lexically-marked spatial relations like *BELOW* take longer to verify than sentences containing their lexically-unmarked counterparts like *ABOVE*.

Propositional models propose a single abstract propositional format for both sentence (description) and figure representations. This common representation consists of embedded relational predicates, each containing one or more arguments. For example, the propositional model presented by Clark and Chase (1972, 1974) holds that simple pictures of some shape "A" depicted above another shape "B" are encoded either as *ABOVE* (A,B) or as *BELOW* (B,A). Similarly, the two descriptions that A is *ABOVE* B and that B is *BELOW* A are represented as *ABOVE* (A,B) and *BELOW* (B,A), respectively. When sentences and pictures are compared, the verification latency depends on the number of operations needed to compare the corresponding constituents of the two representations. This constituent comparison process (Carpenter & Just, 1975) is serial rather than holistic, and the constituents are compared from the "inside out," that is, from the most embedded to the least embedded constituent of the representation.

Researchers using the description matching paradigm typically report two results that are at odds with the description complexity and lexical markedness effects from the sentence verification paradigm. First, the effects of description complexity often vary with the retention interval between the presentation of the description and the subsequent presentation of the test figure. For example, Nielsen and Smith (1973) report that as this description retention interval increases, matching latency in the initial description condition decreases and converges to that

for the initial figure condition. In addition, description matching studies often fail to find effects of lexical markedness on the time to make a sequential description–figure comparison (e.g., Seymour, 1974a, 1975).

Dual code models have been primarily concerned with the asymmetry in the effects of figure–figure and description–figure matches over time. The common feature of these models is that the proposed representations of figures and of spatial descriptions are qualitatively different. Nielsen and Smith's dual code model holds that a figure is represented as an integral image, while a description is initially encoded as a list of features with verbal properties. Their model proposes that the effects of description complexity disappear over time because subjects recode the verbal feature list into a more integral form if they expect an integrated test figure. Dual code models generally postulate two independent comparison processes, with a holistic operator underlying "same" responses and a serial feature comparison underlying responses of "different."

We report two experiments that employ features of both the description matching and sentence verification paradigms. As in the description matching paradigm, we compare reaction time for matching an initially-presented spatial description to reaction time for matching two sequentially-presented visual figures. As in the sentence verification paradigm, we systematically manipulate properties of the initially-presented descriptions.

Our purpose in reporting these experiments differs from that of many previous researchers. We do not test models in the traditional sense. We do not propose or support a particular model of spatial comprehension and comparison. Instead, we present our studies as *demonstrations* that certain experimental methods lead to results which characterize the task more than indicate representational or processing invariants.

EXPERIMENT 1

In the first experiment, we study the process of constructing an internal representation from a figure or from a verbal description of the figure. We generalized the spatial comparison task by using descriptions and figures at several levels of linguistic and pictorial complexity. Previous researchers have manipulated the linguistic complexity of spatial descriptions without varying the pictorial complexity of the test displays. For example, a sentence like STAR ISN'T BELOW PLUS is linguistically more complex than STAR IS ABOVE PLUS because it uses negation and a marked spatial relation, but both sentences describe displays of equal pictorial complexity.

In addition, we separated encoding and comparison time with a successive rather than a simultaneous presentation of the two displays.

After subjects were presented with either a visual figure or a verbal description, they are allowed as much time as necessary to "prepare" for the test figure. Following this preparation interval, a test figure was presented, and subjects were asked to judge as rapidly as possible whether the test figure was the same as or different from the initially-presented figure or description.

Method

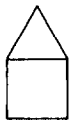
Subjects

Four subjects, all students and staff at the University of California, San Diego, participated in the experiment for four 2-hour sessions and a single 1-hour session.

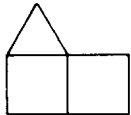
Stimuli

Subjects were presented with two different types of visual displays, geometric figures and verbal descriptions of geometric figures. Figure 1 shows sample figures and their corresponding descriptions. The figure displays could have three levels of complexity. That is, they could be composed of two, three, or four component parts. The component parts were always squares and equilateral triangles.

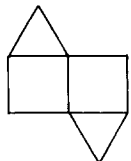
Figures. There were four different figures at each of the three levels of complexity. The four 2-component figures consisted of a square and a triangle, with the triangle above, below, or on the right or left sides of the square. The four 3-component figures consisted of two adjacent squares and a triangle above or below one of the squares. Finally, the four 4-component figures consisted of two adjacent squares and two triangles. For two of these figures, the



TRIANGLE ABOVE SQUARE



TRIANGLE ABOVE SQUARE 1
SQUARE 2 RIGHT SQUARE 1



TRIANGLE 1 ABOVE SQUARE 1
SQUARE 2 RIGHT SQUARE 1
TRIANGLE 2 BELOW SQUARE 2

FIG. 1. Sample figures and descriptions at the three levels of complexity used in Experiment 1.

triangles were either both above or both below the two squares. For the other two figures, the triangles were diagonally opposite one another, with one above and one below different squares.

Descriptions. The verbal descriptions were one, two, or three lines in length. Each 3-word line contained the name of a component figure, a spatial relation, and the name of a second component figure. The relational terms used were ABOVE, BELOW, RIGHT, and LEFT. These relations were used in their "natural" meanings (see Fig. 1).

Two or more descriptions for each figure were used in order to study the effects on preparation and comparison times of different relational terms and the order in which different component parts were introduced and arranged. Each of the four 2-component figures was described by a pair of 1-line descriptions using either the ABOVE/BELOW or RIGHT/LEFT spatial relations. Each of the 3-component figures had four 2-line descriptions and each 4-component figure had eight 3-line descriptions. In all, 56 different descriptions were used in Experiment 1.

Both figures and descriptions were projected from black-and-white 35 mm slides. Each of the two types of visual displays subtended approximately 8 degrees of visual angle, and the screen on which the displays were projected subtended approximately 16 degrees.

Procedure

On each experimental trial, two reaction times were recorded. The first reaction time (or preparation time, RT_1) consisted of the time needed to comprehend and encode the initial display. RT_1 was the time between the onset of the initially-presented figure or description and the subject's foot pedal press that denoted a request for presentation of the test figure. The foot pedal press resulted in termination of the first visual display and initiation of the test display sequence. The interval between the subject's preparation response (foot pedal press) and presentation of the test figure was either 0 or 3000 msec. The test figure appeared centered on the screen in the same spatial location as the first display.

The second reaction time (or comparison time, RT_2) consisted of the time needed to determine whether the test figure was the same as or different from the figure initially presented or described. This comparison time was measured from the onset of the test figure to the subject's indication of "same" or "different" by pressing one of two response buttons. After subjects responded, they received feedback from a light over the correct response button. A "same" test figure was identical in all respects to the one initially presented or described. "Different" figures were one of the other three figures from the same level of complexity as the "same" form. This constraint on the "different" set precluded responses based solely on the number of components in the test figure. In addition, the nature of the 4-component figures discouraged subjects from adopting a strategy of encoding a single feature from the initial display and responding "same" or "different" depending on whether or not the feature was present in the test figure.

The written instructions emphasized: (a) spending no more time than necessary to encode a representation from the first display that would yield rapid and accurate responses in the "same"-"different" recognition task; (b) judging only identical figures to be the "same" (structurally-identical figures presented in different orientations were to be judged as "different"); and (c) responding as quickly as possible while keeping errors to a minimum.

There were four experimental conditions: figure-figure (F-F) and description-figure (D-F) trials at each of the two interstimulus intervals, 0 and 3000 msec. In each experimental session, trials were blocked by each of the four conditions in 96-trial blocks, and all four conditions were run in each session for a total of 384 trials per session. The order of the four blocks was balanced within each session and randomly assigned between subjects, with the constraint that F-F and D-F blocks alternate.

Within each session and condition, trials were randomly selected from the three levels of figure or description complexity. For each F-F condition, the 384 trials from the four

experimental sessions were composed as follows. Each of the 12 different figures (four from each level of complexity) was presented 32 times as the initial display. On half of these trials, the test figure was the same as the initial figure; on the other half, a different figure appeared. In each description-figure condition, the composition of the 384 overall trials was more complex. Each of the eight different 1-line descriptions (two for each 2-component figure) appeared 16 times. Each of the 16 different 2-line descriptions (four for each 3-component figure) appeared eight times. Finally, each of the 32 different 3-line descriptions (eight for each 4-component figure) was presented four times.

The order of trials within blocks was randomly generated by a computer which also controlled the slide projectors and recorded the two reaction times on each trial. In a fifth session, trials on which errors had been made were retaken, embedded in other "filler" trials, in order to obtain a correct choice reaction time for each of the 1536 trials per subject required by the complete factorial design.

Results

Preparation Time (RT_1)

Figure 2 illustrates mean correct RT_1 , averaged over interstimulus interval, and plotted separately for the F-F and the D-F conditions.

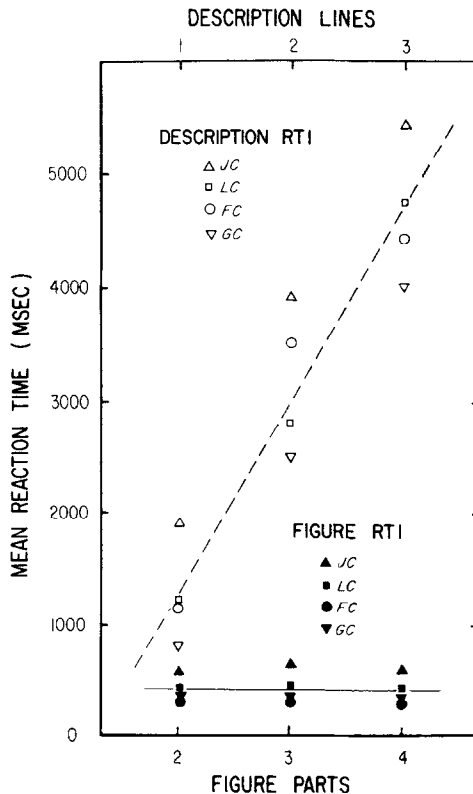


FIG. 2. Mean preparation or comprehension time (RT_1) as a function of complexity for the description and figure conditions of Experiment 1. Symbols are the means for individual subjects, with the best-fitting straight lines for the group data.

The most striking features of the data are the differences in the time required to prepare for a test figure in the two presentation conditions. In the F-F condition, approximately 400 msec are needed to encode an initially-presented visual figure, and this time is not affected by the complexity of the figure (the number of components or "figure parts;" see Fig. 1). In the D-F condition, the time needed to construct a memory representation from a description increases linearly with the complexity of the description (the number of sentences or "description lines;" see Fig. 1).

A 4-way analysis of variance was performed on the group RT_1 data; the factors were Subjects, Presentation Condition (D-F and F-F), Interstimulus Interval (0 and 3000 msec), and Complexity (the number of description lines and figure parts).² This analysis is shown in Table 1 and confirms the patterns shown in Fig. 2. Only three sources of variance not involving the Subjects factor were significant. These were Presentation Condition, Complexity, and the Presentation Condition \times Complexity interaction. Planned comparisons confirmed that preparation time for initially-presented figures was not affected by complexity and that the mean preparation times for descriptions of one, two, and three lines all differed from one another (the linear trend was significant [$F(1,3) = 124.86$]).

Relational term and description effects. In order to examine possible effects on preparation time of different relational terms and alternate descriptions of the same figure, separate analyses of variance were performed on the D-F RT_1 data for each level of description complexity. Previous investigators (e.g., Clark & Chase, 1972; Olson & Laxar, 1973) have reported that sentences using ABOVE or RIGHT are comprehended

TABLE 1
ANALYSIS OF VARIANCE FOR EXPERIMENT 1 (RT_1)

Source	Error term	MS error	F (df)	p
Subjects (S)	Within cell	.65	252.47 (3,6096)	.001
Condition (Con)	S \times Con	80.53	133.46 (1,3)	.01
Interstimulus interval (ISI)	S \times ISI	7.49	5.24 (1,3)	ns
Con \times ISI	S \times Con \times ISI	4.52	8.26 (1,3)	ns
Complexity (Com)	S \times Com	7.35	206.04 (2,6)	.001
Con \times Com	S \times Con \times Com	6.91	218.24 (2,6)	.001
ISI \times Com	S \times ISI \times Com	1.92	1.38 (1,3)	ns
Con \times ISI \times Com	S \times Con \times ISI \times Com	1.92	1.56 (2,6)	ns

² Analyses of variance in Experiments 1 and 2 considered only correct response times. All group analyses treated "Subjects" as a random factor. When values of the F statistic are reported, they are significant at the .05 level or better unless otherwise noted.

faster than semantically-equivalent sentences with their "marked" counterparts BELOW and LEFT. However, for our 1-line, 2-component descriptions, there was no overall effect of relational terms. In fact, there were no significant differences between any pair of relational term means. Sentences containing ABOVE (1171 msec) did not differ in RT_1 from those containing BELOW (1167 msec), and sentences containing RIGHT (1383 msec) did not differ from those containing LEFT (1309 msec). On the other hand, the large differences in preparation time among the four subjects make these comparisons somewhat insensitive (the 95% Scheffe confidence interval for the difference between two means is 309 msec),³ so our results are inconclusive about markedness effects on preparation time. Similarly, there were no overall effects of description types for 2-line, 3-component descriptions or for 3-line, 4-component descriptions.

Comparison Time (RT_2)

Figure 3a presents mean correct RT_2 averaged over subjects, interstimulus interval, and responses. Means for the F-F and D-F con-

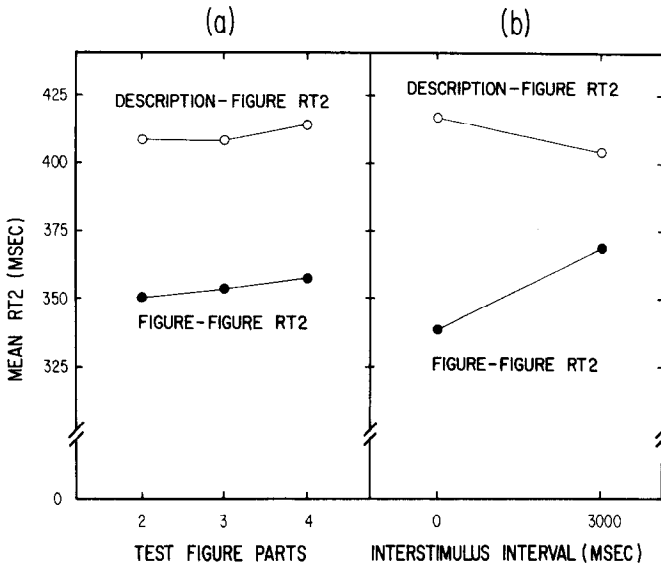


FIG. 3. (a) Mean correct comparison time (RT_2) as a function of test-figure complexity for the description-figure and figure-figure conditions of Experiment 1. (b) Mean correct RT_2 as a function of interstimulus interval for the description-figure and figure-figure conditions of Experiment 1.

³ The Scheffe method for making multiple comparisons (Scheffe, 1959) yields confidence intervals which simultaneously apply with a fixed Type 1 error probability to all possible contrasts among a set of means. The size of the confidence interval depends on the treatment error term from the analysis of variance. The Scheffe test is the most stringent method for making multiple comparisons and yields the widest confidence intervals.

ditions are plotted separately as a function of the complexity of the test figure or number of "test-figure parts." Figure 3a clearly shows that, while the F-F means are about 60 msec faster than the D-F means, neither reaction time function is affected by the complexity of the test figure. In addition, though not shown in Fig. 3a, "same" responses are reliably faster than "different" responses in both the F-F and D-F conditions. In figure-figure conditions, "sames" take 339 msec compared to 367 msec for "differents", and in description-figure conditions, "sames" take 390 msec with "differents" requiring 432 msec. In a 5-way analysis of variance on the group data, shown in Table 2, the main effect of Responses was highly significant and unambiguous since the Conditions \times Responses interaction was not significant.

The constant difference between the F-F and D-F matching times, the absence of an effect of test-figure complexity, and the greater speed of "same" responses were also apparent in the data of each of the four subjects. In Fig. 4a, the *difference* between D-F and F-F matching time is plotted as a function of the complexity of the test figure for each of the subjects separately. We introduce this difference score here because the most important measure for each subject is the *relative* speeds of the F-F and D-F conditions; the variation across subjects in the absolute speeds of

TABLE 2
ANALYSIS OF VARIANCE FOR EXPERIMENT 1 (RT_2)

Source	Error term	MS error	F (df)	p
Subjects (S)	Within cell	.005	163.24 (3,6048)	.001
Condition (Con)	S \times Con	.496	9.83 (1,3)	ns
Interstimulus interval (ISI)	S \times ISI	.206	.52 (1,3)	ns
Con \times ISI	S \times Con \times ISI	.062	11.68 (1,3)	.05
Complexity (Com)	S \times Com	.008	2.70 (2,6)	ns
Con \times Com	S \times Con \times Com	.007	.42 (2,6)	ns
ISI \times Com	S \times ISI \times Com	.004	1.58 (2,6)	ns
Con \times ISI \times Com	S \times Con \times ISI \times Com	.006	2.28 (2,6)	ns
Response (R)	S \times R	.047	36.32 (1,3)	.01
Con \times R	S \times Con \times R	.024	2.26 (1,3)	ns
ISI \times R	S \times ISI \times R	.009	4.15 (1,3)	ns
Con \times ISI \times R	S \times Con \times ISI \times R	.006	.43 (1,3)	ns
Com \times R	S \times Com \times R	.006	.22 (2,6)	ns
Con \times Com \times R	S \times Con \times Com \times R	.006	.20 (2,6)	ns
ISI \times Con \times R	S \times ISI \times Con \times R	.005	.23 (2,6)	ns
Con \times ISI \times Com \times R	S \times Con \times ISI \times Com \times R	.009	1.57 (2,6)	ns

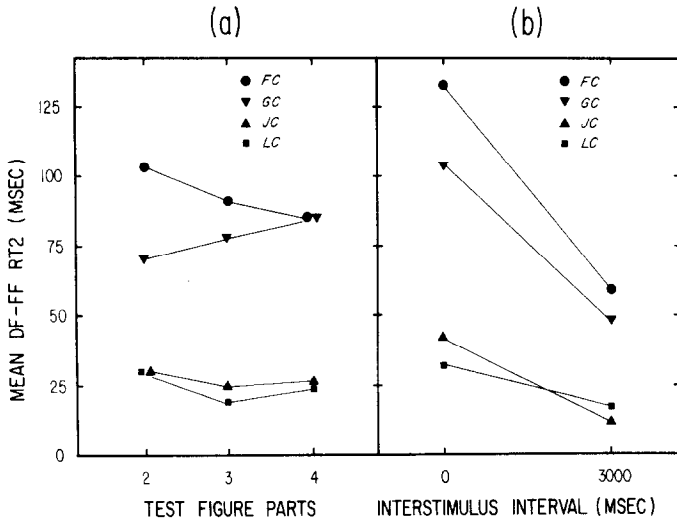


FIG. 4. (a) Mean *difference* between description-figure RT_2 and figure-figure RT_2 in Experiment 1 as a function of complexity, plotted separately for each of the four subjects. (b) Mean *difference* between description-figure RT_2 and figure-figure RT_2 in Experiment 1 as a function of interstimulus interval, plotted separately for each of the four subjects.

the two conditions tends to obscure the more meaningful trends in the data. The positive values of these differences indicate that D-F matching times were longer than F-F matching times for each subject. The relative flatness of the functions shown in Fig. 4a indicates that, for each subject, D-F and F-F reaction-time functions were nearly parallel. The difference between overall D-F and F-F matching times just failed to achieve statistical significance in the group analysis, but this difference between conditions was significant in each of the 4-way analyses of variance performed on the data of the individual subjects.

No complexity effects. The failure to find an effect of complexity in RT_2 is not due to a lack of power. The means for the 2-, 3-, and 4-component figures hardly differ (379, 380, and 385 msec, respectively), and with 2048 observations in each of these three means, the 95% Scheffe confidence interval for a difference between any two is at most 15 msec. Thus these differences are significantly smaller than any theoretically meaningful effect of complexity (e.g., Carpenter & Just, 1975, propose a 200-msec constituent comparison operation).

Effects of interstimulus interval. Figure 3b presents mean correct RT_2 as a function of interstimulus interval, plotted separately for the D-F and the F-F conditions. In the analysis of variance on the group data, the main effect of Interstimulus Interval was not significant, primarily because the Subjects \times Interstimulus Interval and Presentation Condition \times Interstimulus Interval interactions were both significant sources of variance.

In the individual subject analyses this Condition \times Interstimulus Interval interaction was significant in all four cases, but the nature of the interaction differed somewhat among the subjects. For all of the subjects, F-F matching times were considerably more rapid with the 0-msec interstimulus interval than with the 3000-msec interval. For two of the subjects, D-F matching times were relatively unaffected by the length of the interstimulus interval. However, for the other two subjects, D-F matching times were slower at the 0-msec interval than at the 3000-msec interval. The magnitude of the Presentation Condition \times Interstimulus Interval interaction for individual subjects can be seen in Fig. 4b, in which the *differences* between D-F and F-F reaction times are plotted as a function of interstimulus interval. Thus, the overall interaction shown in Fig. 3b represents a very reliable increase in F-F matching time with increasing interstimulus interval, and a somewhat uncertain effect in the opposite direction for D-F matches. A planned comparison confirmed that the overall interaction was located in the increase in F-F matching time [$F(1,3) = 16.61$].

Relational term and description effects. Separate analyses of variance on the group RT_2 data for each level of description complexity examined possible effects of using alternate spatial relations or syntactic structures on matching time. In none of these three analyses were the effects of Figures or Descriptions significant, nor did these factors interact with each other or with any other factor. In addition, we tested specific hypotheses of markedness effects with Scheffe contrasts over the eight 1-line description means, but found no significant differences. The two descriptions using ABOVE (399 msec) were verified no faster than the two with BELOW (401 msec), and those with RIGHT (413 msec) did not differ from those with LEFT (422 msec).

The theoretical importance of such effects of linguistic markedness on matching and verification times requires us to demonstrate that our failure to find them was not due to a lack of power. With 256 observations in each of the four relational term means, the 95% Scheffe confidence interval for the larger of the two critical differences (that between RIGHT and LEFT) is 35 msec. Markedness effects reported in 12 different experiments by previous investigators have averaged about 90 msec, and the smallest difference ever interpreted as theoretically meaningful was a 53-msec advantage of ABOVE over BELOW found by Seymour (1969).⁴ Thus the precision of the data here makes our power to find even the smallest "real" markedness effect considerably greater than .99.

⁴ We found eleven sentence verification experiments in which ABOVE was verified faster than BELOW. In chronological order with the ABOVE advantage in parentheses these are: Seymour, 1969 (53 msec); Chase and Clark, 1971 (2 experiments—75 and 83); Clark and Chase, 1972 (4 experiments—93, 117, 84, and 90); Clark and Chase, 1974 (2 experiments—137 and 136); Seymour, 1974a (91); Just and Carpenter, 1975 (56). We located two reports of RIGHT faster than LEFT: Olson and Laxar, 1973 (94); Just and Carpenter, 1975 (95).

Errors

Error rates were low, ranging from 2.6% to 6.4% for individual subjects, with an average rate of 3.7%. Errors occurred with equal frequency in the F-F and the D-F conditions, and were equally likely for displays at each level of complexity. However, errors were more frequent when the interstimulus interval was 0 msec (4.8%) than when it was 3000 msec (2.5%). This effect may be due to some difficulty in making two responses in rapid succession.

Discussion

Two central features of the results of Experiment 1 conflict with those reported in similar sentence verification and description matching experiments. This inconsistency poses problems for the general models of spatial comprehension and verification proposed in these other tasks. First, the comparison time functions for both F-F and D-F conditions shown in Fig. 3a are qualitatively alike in that neither function is affected by the complexity of the test figure or the length of its initial description. Second, neither the particular spatial relation (ABOVE, BELOW, RIGHT, or LEFT) used in a description line nor the order in which components were introduced influences D-F comparison time.

In particular, dual code models have been proposed to account for asymmetric effects of complexity on figure and description matching, and no such differences were found in the present experiment. Propositional models also predict effects of description complexity on comparison time. For example, Carpenter and Just's (1975) constituent comparison model (one of the propositional models for the process of comparing sentences and pictures) proposes a serial comparison of corresponding parts of the initial and test representations. Such a model should certainly predict an increase in RT_2 with complexity in our D-F condition, unless the time to find and compare each constituent can be extremely short or the size of each constituent is variable. Both of these alternatives, however, appear to be ruled out in Carpenter and Just's description of their model.

The absence of effects of relational terms and of alternate descriptions for a given figure on comparison time is also problematic for propositional models of spatial comprehension. For example, the serial comparison model of Clark and Chase (1972) is based in part on effects of lexical markedness of the particular spatial relations used to describe a figure.

The failure to find any effects of lexical markedness or description type on comparison time is consistent with the claim that the figures and descriptions are ultimately represented in the same way. Thus, while the representation may have contained information regarding the original form of the display, this information was not directly utilized in the comparison process. The absence of a complexity effect on comparison time in both the

initial figure and the initial description conditions supports the related claim that all parts of the representations are at once available to the comparison process or else retrieved and compared in parallel. Subjects may have tried to verify that the two displays were the same and then made a default response of "different" if this holistic or parallel comparison failed to produce a match. This analysis is consistent with the advantage of "same" matches over "different" matches.

However, this interpretation of Experiment 1 depends on the assumption that the internal representation in the comparison process preserves the complexity of the initial description or figure. Another possibility is that subjects construct a representation from a description that *initially* reflects its complexity (which would produce the increase in description RT_1 with complexity), but then abstract some simpler representation for verification that is independent of complexity. Specifically, it may be possible on some trials for subjects to make a correct "same" or "different" judgment on the basis of a single stimulus feature, such as the direction that a triangle points. While we acknowledge that the distractor items are sometimes discriminable from the correct figure by such a simple test, we believe that the inclusion of more complex figures for which such a one-feature strategy will not work (such as the 4-component displays) induces subjects to construct representations that preserve more complete information about the spatial structure of the figures. If a single-feature test were always used, errors would be more frequent for the 4-component test figures, but error rates were not affected by test figure complexity. Nevertheless, the best counter-argument against this single-feature interpretation of the results of Experiment 1 is to replicate them using a stimulus set for which a single-feature representation is inadequate in *all* cases. One purpose of Experiment 2 is to perform this validation.

In summary, our results are at odds with previous research on spatial comprehension. The data suggest that in our task, subjects construct spatial representations and use parallel or holistic verification operations. These results and interpretations are different from those proposed by propositional and dual code theorists, even though their experiments used paradigms quite similar to ours. We attribute much of this discrepancy to methodological differences between the procedures used in Experiment 1 and those used by other researchers. The two most salient methodological differences are (a) the use of subject-controlled instead of fixed duration displays, and (b) the use of successive instead of simultaneous presentations. The primary goal of Experiment 2 is to study spatial matching and verification processes in a number of experimental situations that span these different methodologies.

EXPERIMENT 2

In this experiment, we study four tasks along a continuum from simultaneous presentation of the description and figure displays, through

“deadline” conditions with successive displays, and finally to a subject-controlled condition where the onset of the test figure is completely determined by the subject. From the first experiment and from some logical considerations, we can generate a number of predictions about how encoding and comparison processes might change along with the structure of the verification task in Experiment 2.

In the subject-controlled condition, subjects presumably take the time to construct the optimal representation for the successive verification task. This representation allows subjects to decide rapidly whether the test figure is the one that they expect. Since the subject-controlled procedure separates the encoding of the description from the encoding of the test figure, this representation of the initial description logically must be capable of distinguishing the described figure from any member of the entire class of distractors in the experiment. The description representation that subjects constructed in Experiment 1 to expedite comparison with a pictorial display could be verified equally rapidly for figures of differing complexity without effects of the particular linguistic elements in the descriptions.

One reasonable tack for subjects in the deadline conditions would be to continue processing the description to generate this optimal representation even after the test figure is presented. On the other hand, once the test figure appears, subjects might use some information about its structure to facilitate the encoding of a representation from the description. Thus, in these conditions, and in the simultaneous condition as well, processing of the description might be contingent on particular features of the figure against which the description must be verified. The subject-controlled task is unique in that it alone requires the encoding of the description to be completed prior to the encoding of the test figure.

Contingent processing of the description in the deadline or simultaneous condition might have several effects. First, subjects might be able to detect a difference between the test figure and a partially-encoded description. Thus the representations of the description and the figure might be less complete in conditions other than the subject-controlled where there is no opportunity for contingent processing. If subjects construct the minimal representations of the two displays that will suffice on a given trial, then we might expect effects of display complexity on verification time. Finally, if some characteristics of the test figure are known while the description is being encoded, it may be desirable to preserve specific information about the description in its representation. This information could then be used in the verification procedure to direct scanning or testing operations toward features that are mentioned in the description. Effects of lexical markedness might be generated in the use of this specific lexical information.

In the simultaneous condition, subjects have no time to generate expectations about the test figure before it appears. Even more than in the deadline conditions, subjects may use features of the accompanying test

display to eliminate unnecessary encoding and comparison processing (i.e., to construct minimal representations that are optimal for particular test figures on a given trial). Nevertheless, most of the current models of spatial comprehension characterize simultaneous sentence-picture comparison as a sequence of *independent* stages consisting of (a) encoding the two displays, (b) comparing the two representations, and (c) executing a choice response (Clark & Chase, 1972, 1974; Carpenter & Just, 1975; Just & Carpenter, 1975). These models also assume that the comparison of the sentence and figure requires complete representations of each.

Method

Subjects

Four students at the University of California, San Diego were paid for completing two experimental sessions which together took 3 hrs. None of the subjects had been in Experiment 1.

Stimuli

We used 2- and 3-component figures with their corresponding 1- and 2-line verbal descriptions. There were eight different figures at each level of complexity. The component parts from which all displays were constructed were a square, a triangle, and a circle. We introduced the third component to generate a more complex set of figures than the 2-component sets used in Experiment 1. For any figure in the stimulus set, there exists another that has all but one component in the same location and orientation; alternatively, for any component in a figure, another figure exists in the set that has that component in the identical location and orientation. This stimulus set makes a single-component comparison strategy inadequate. This thereby implies that display complexity is incorporated in the comparison representations.

Two-component displays. The 2-component figures were the four arrangements of the triangle either above or below the square or the circle, the two vertical arrangements of the square and circle, and the two horizontal arrangements of the square and circle. Each of these eight figures had two 1-line descriptions using either the ABOVE/BELOW or the RIGHT/LEFT pairs of spatial relations.

Three-component displays. The 3-component figures were all composed of a square, a circle, and a triangle. All of these forms had the square and circle side-by-side with the triangle either above or below the square or circle. There were four such configurations with the circle on the right of the square and four with these positions reversed. Each of these eight figures had four 2-line descriptions. One of these descriptions used ABOVE and RIGHT, one used ABOVE and LEFT, one used BELOW and RIGHT, and the other used BELOW and LEFT.

Procedure

The four tasks in Experiment 2 have a number of features in common but are qualitatively different in other respects. The subject-controlled and the two deadline conditions all used the two-reaction-time procedure of Experiment 1. In all three tasks, the test figure appeared when the subject pressed a button to indicate that his comprehension of the spatial description was sufficient to enable a rapid and accurate "same" or "different" decision. However, in the two deadline conditions, the test figure came on after a certain amount of time had elapsed since the onset of the description even if the subject had not yet signaled the preparation was complete. Finally, in the simultaneous condition, the description and the test figure came on together, and only a single reaction time was recorded. We shall now discuss each of these conditions in more detail.

Subject-controlled condition. This task is primarily a replication of the description–figure (D–F) matching condition of Experiment 1. However, the addition of the third component in the figure set makes this condition a critical test of the interpretation from Experiment 1 that the description representation in the successive matching task incorporates spatial information in a form that preserves the complexity of the initial description.

6- and 2-sec deadline conditions. We included the two deadline conditions to study the spatial comparison activity in a situation intermediate between the subject-controlled condition and the simultaneous condition. From pilot work, we chose the 6-sec deadline to create a spatial verification situation where subjects were not likely to have always completed their preparation from the more complex descriptions when the test figure appeared. However, 6 sec is almost certainly enough time to prepare a suitable verification representation for the 1-line, 2-component descriptions. Similarly, we chose the 2-sec deadline to be close to the average time that subjects need to adequately understand the 1-line, 2-component descriptions.

The two deadline conditions are somewhat similar to the fixed duration presentations often used in description matching studies. In the fixed duration procedure, the initial figure or description display is presented to the subject for a fixed amount of time. While both deadline and fixed display procedures place an upper limit on preparation time, the deadline procedure allows subjects to initiate the test sequence *before* that limit is reached if their preparation is completed before the deadline.

Simultaneous condition. This task is the one usually used by researchers who study spatial comprehension and verification. The description and test figure appear together, and the subject's task is to decide whether the sentence describes the test figure. Thus only one reaction time can be recorded.

The descriptions and test figures were plotted on an oscilloscope display by a computer which also recorded the reaction times on each trial. The descriptions appeared at the top of the screen, and remained on when the test figures appeared below them (note that the description and figure came on together in the simultaneous condition). The entire screen subtended about 12 degrees of visual angle.

Each subject participated in all four experimental conditions. Trials were blocked by conditions, and subjects completed two of these blocks during each of the two experimental sessions. The order of these four blocks was balanced across subjects to control for practice effects.

Each condition included 192 trials. The first 64 trials in each block were treated as practice and not analyzed. The 128 experimental trials in each block consisted of 64 from each of the two levels of description and test figure complexity and were randomly ordered. Each of the 16 different 1-line descriptions (two for each of the eight 2-component figures) appeared four times. On half of these trials, the description was followed by the "same" figure, and on the other half, a "different" figure appeared. Each of the 32 different 2-line descriptions (four for each of the eight 3-component figures) appeared twice, once followed by a "same" figure and once by a "different" distractor. Thus each subject completed 512 experimental trials.

Results

The results of Experiment 2 are conceptually simple but complex in appearance. Therefore, we shall risk some redundancy to gain clarity by presenting the results from two different perspectives. First, we will present the results from the viewpoint of the general two-reaction-time task structure and review RT_1 and RT_2 across the four experimental conditions. Then, we will re-present some of the results separately for each of the four experimental conditions.

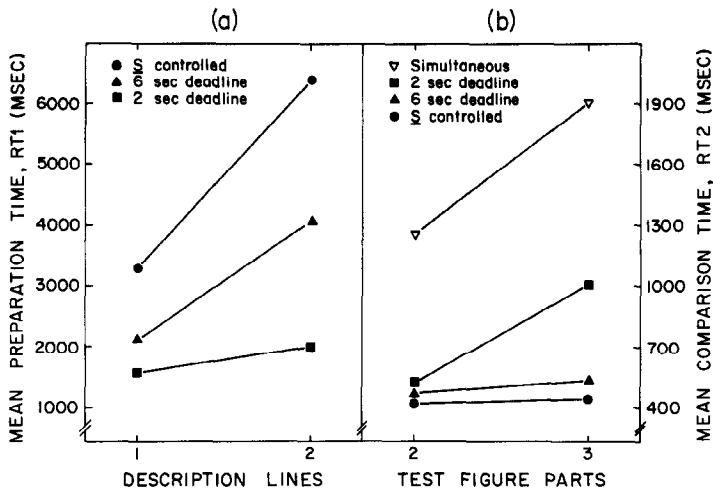


FIG. 5. (a) Mean correct preparation time (RT_1) as a function of description complexity for the subject-controlled and deadline conditions of Experiment 2. (b) Mean correct comparison time (RT_2) as a function of test figure complexity for the four conditions of Experiment 2.

The Spatial Comprehension Task: General Patterns

Complexity effects on preparation time. The left-hand panel of Fig. 5 presents mean correct RT_1 as a function of description complexity in the subject-controlled and deadline conditions. Three major features of the RT_1 results are apparent in this figure. First, in all three conditions, 2-line descriptions take longer to comprehend than 1-line descriptions. Second, the deadline conditions worked; the time taken to encode the descriptions in the subject-controlled conditions is greater than the time allowed in the deadline conditions. Finally, the difference in RT_1 between 1- and 2-line descriptions is largest in the subject-controlled condition, intermediate in the 6-sec deadline condition, and smallest in the 2-sec deadline condition.

A 3-way analysis of variance for RT_1 was performed using the means for the four relational terms at each level of description complexity as replications. The three factors were Subjects (4), Conditions (3—there is no RT_1 measure in the simultaneous condition), and Complexity (2 levels). For the 1-line descriptions, the replications were the means for descriptions using ABOVE, BELOW, RIGHT, and LEFT. For the 2-line descriptions, the replications were the means for the ABOVE and LEFT, ABOVE and RIGHT, BELOW and LEFT, and BELOW and RIGHT 2-term combinations. Table 3 shows the results of this analysis.

Every term in this analysis was a significant source of variance. Nevertheless, every effect is readily interpretable. The three main effects of Subjects, Conditions, and Complexity are significant because people differ in overall response speed, the tasks in Experiment 2 differ considerably, and because 1-line descriptions are easier to understand than

TABLE 3
ANALYSIS OF VARIANCE FOR EXPERIMENT 2 (RT_1)
(RELATIONAL TERM MEANS AS REPLICATIONS)

Source	Error term	MS error	F (<i>df</i>)	<i>p</i>
Subjects (S)	Within cell	.15	134.24 (3,72)	.001
Condition (Con)	S × Con	7.31	10.87 (2,6)	.05
Complexity (Com)	S × Com	2.56	29.06 (1,3)	.05
Con × Com	S × Con × Com	1.52	8.86 (2,6)	.05

2-line descriptions. The Conditions × Complexity interaction primarily results from the 2-sec deadline condition which greatly reduces the effect of description complexity by limiting 2-line descriptions to a deadline not much longer than the mean for 1-line descriptions.

Markedness effects on preparation time. Table 4 summarizes the effects of linguistic markedness in Experiment 2 by presenting means for the ABOVE/BELOW and RIGHT/LEFT relational terms for RT_1 and RT_2 in each condition. In RT_1 , the advantage of ABOVE over its marked counterpart BELOW is consistent, averaging 209 msec for the two deadline conditions and the subject-controlled condition. However, the high variability in RT_1 left all three ABOVE-BELOW differences just short of significance. The RIGHT-LEFT difference was even less reliable. While overall RIGHT was comprehended 194 msec faster than LEFT, this advantage was 748 msec in the subject-controlled condition and then reversed in favor of LEFT in the other conditions. LEFT was 6 msec faster than RIGHT in the 6-sec task and 171 msec faster in the 2-sec task.

Complexity effects on comparison time. Figure 5b presents mean correct RT_2 in each of the four conditions as a function of test figure complexity. If viewed in conjunction with the data for RT_1 in Fig. 5a, this figure illustrates the most important aspects of Experiment 2. *Subjects respond increasingly faster to the test figure as preparation time increases from zero (in the simultaneous condition) to 3 sec and more (in the subject-controlled condition).* Second, subjects generally respond faster to 2-component test figures than to 3-component test figures. However, this feature of the RT_2 data is dominated by the interaction between this effect of complexity and the effect of preparation condition. *The size of the complexity effect steadily shrinks as subjects take more time to prepare for the test figure, and the effect disappears completely when subjects take as much time as they need in the subject-controlled condition.*

These observations are supported by a 4-way analysis of variance of the RT_2 data which used the means for the four relational terms at each level of description complexity as replications. The four factors in the analysis were Subjects (4), Conditions (4), Complexity (2 levels), and Responses (2—Same and Different). Table 5 reports the results of this analysis. The three aspects of the RT_2 results apparent in Fig. 5b are captured by the main

TABLE 4
 LEXICAL MARKEDNESS EFFECTS IN EXPERIMENT 2

Condition	Spatial term	Mean (msec)	S.D. (msec)	Markedness effect	<i>t</i>	<i>df</i>	<i>p</i>
<i>RT₁</i>							
Subject-controlled	Above	3015	1577	341	1.33	190	.10
	Below	3356	1940				
	Right	3265	1748	748	1.35	62	.10
	Left	4013	2540				
6-Sec deadline	Above	1973	803	199	1.58	190	.10
	Below	2172	931				
	Right	2266	1154	-6	-.02	62	ns
	Left	2260	1007				
2-Sec deadline	Above	1496	392	87	1.57	190	.10
	Below	1583	372				
	Right	1685	384	-171	-1.76	62	.05
	Left	1514	379				
<i>RT₂</i>							
Subject-controlled	Above	426	102	7	.45	177	ns
	Below	433	106				
	Right	396	83	25	1.05	53	ns
	Left	421	90				
6-Sec deadline	Above	444	116	47	2.34	181	.01
	Below	491	152				
	Right	463	134	-8	-.26	58	ns
	Left	455	97				
2-Sec deadline	Above	510	149	31	1.28	171	.10
	Below	541	167				
	Right	540	192	-28	-.65	57	ns
	Left	512	124				
Simultaneous	Above	1159	261	130	2.77	176	.01
	Below	1289	353				
	Right	1239	349	169	1.38	53	.10
	Left	1408	525				

effects of Conditions and Complexity, and by the Conditions \times Complexity interaction.

Responses. Not shown in Fig. 5b but critical in the interpretation of Experiment 2 are the effects of responses. Table 6 presents mean times to respond "same" and "different" as a function of test figure complexity in each of the four tasks. In the overall analysis of *RT₂*, the average time to make a "same" response (844 msec) did not differ from the time to respond "different" (805 msec). However, this main effect is diluted by large 2-way interactions between Condition and Responses and between Complexity and Responses, and by the 3-way interaction between Conditions, Complexity, and Responses. We shall discuss these three interactions in turn.

TABLE 5
ANALYSIS OF VARIANCE FOR EXPERIMENT 2 (RT_2)
(RELATIONAL TERM MEANS AS REPLICATIONS)

Source	Error term	MS error	F (df)	p
Subjects (S)	Within cell	.014	28.31 (3,192)	.001
Condition (Con)	S × Con	.088	208.34 (3,9)	.001
Complexity (Com)	S × Com	.026	209.52 (1,3)	.001
Con × Com	S × Con × Com	.133	9.97 (3,9)	.01
Response (R)	S × R	.059	1.63 (1,3)	ns
Con × R	S × Con × R	.023	8.62 (3,9)	.01
Com × R	S × Com × R	.014	68.97 (1,3)	.01
Con × Com × R	S × Con × Com × R	.007	67.80 (3,9)	.001

The Conditions × Responses interaction reflects the finding that “same” responses are faster than “different” responses in the subject-controlled and the 6-sec deadline conditions, that the responses are equal in speed at the 2-sec deadline, and that “different” responses are faster than “same” responses when the two displays appear simultaneously. The Complexity × Responses interaction is due to the advantage for “sames” for 2-component test figures with the reverse advantage for “differents” for the 3-component test figures. The pattern of faster “different” responses for 3-component figures holds in every condition but the subject-controlled one, accounting for the 3-way Conditions × Complexity × Responses interaction. In the subject-controlled condition, “sames” are faster than “differents” at both levels of test figure complexity.

Markedness effects on comparison time. Table 4 shows the effects of

TABLE 6
MEAN RT_2 (msec) FOR EXPERIMENT 2

Condition	Complexity			
	2		3	
	Same	Diff	Same	Diff
Subject-controlled	412	439	431	455
6-Sec deadline	432	497	546	526
2-Sec deadline	481	570	1045	962
Simultaneous	1188	1312	2195	1613
MEAN	636	719	1051	891
		677		971

linguistic markedness on RT_2 by presenting means for the ABOVE/BELOW and RIGHT/LEFT pairs in the four conditions. The markedness effects change considerably in the different tasks. The ABOVE and RIGHT advantages are largest in the simultaneous condition; ABOVE is verified 130 msec faster than BELOW, and RIGHT is verified 169 msec faster than LEFT. The ABOVE advantage shrinks to 47 and 31 msec in the two deadline conditions, and finally to just 7 msec in the subject-controlled condition. The RIGHT-LEFT differences do not diminish according to such an orderly trend but the advantage of RIGHT is a nonsignificant 25 msec in the subject-controlled condition.

Errors. Subjects made errors on 8.4% of the trials in Experiment 2, with individual error rates ranging from 5.9% to 15.2%. Errors were most frequent in the 2-sec deadline condition and least common in the simultaneous condition. Errors were more likely for 3-component test figures than for 2-component test figures in every condition except the simultaneous task. Finally, errors were equally frequent on "same" and "different" trials in every condition.

Results of the Separate Conditions

We have now presented the results of Experiment 2 from a global perspective across the four tasks. The pattern of complexity effects, lexical markedness effects, and the relative speed of "same" and "different" responses changes in an orderly way as the task structure changes. This makes the local description of individual tasks considerably less important. Therefore, we shall just briefly consider each condition and review features of the results not adequately discussed in the preceding overview.

Subject-controlled condition. The results in the subject-controlled condition clearly replicate the description-figure (D-F) conditions of Experiment 1. Preparation time (RT_1) was strongly affected by the complexity of the verbal descriptions, with 2-line descriptions taking over 3 sec longer to understand than the 1-line descriptions. On the other hand, comparison time (RT_2) was not affected by test figure complexity. Three-component figures took only 18 msec longer to verify than 2-component figures. The failure to find a complexity effect of a theoretically-meaningful size (100–200 msec according to previous reports) is not due to a lack of power. In fact, the 18 msec difference that we did find approaches statistical significance [$t(469) = 1.82, p < .10$] because the standard error of the difference is only 10 msec.

Most importantly, the theoretically critical absence of effects of linguistic markedness on comparison time (RT_2) was replicated in Experiment 2. One-line descriptions with ABOVE (425 msec) were no faster than those with BELOW (433 msec), $t(177) = .45$, and descriptions with RIGHT (396 msec) were not faster than those with LEFT (421 msec), $t(53) = 1.05$. The precision of these data makes it clear that the failure to

find markedness effects is not due to lack of power, since the standard errors of these differences are just 16 and 24 msec, respectively.

6-sec deadline condition. Complexity affected both RT_1 and RT_2 in the 6-sec deadline condition. In RT_1 , 2-line descriptions were encoded over 2 sec faster than 2-line descriptions in the subject-controlled condition. This shortening of RT_1 for the more complex descriptions reflects the truncated preparation time distribution in a deadline condition. Any preparation activity that takes longer than 6 sec must carry over into RT_2 after the test figure is already present. Subjects had not finished their preparation by the 6-sec deadline on 15.6% of the 2-line descriptions and on 1.2% of the 1-line descriptions.

The differential carry-over from RT_1 in a deadline condition resulted in an effect of test figure complexity in RT_2 [$t(476) = 4.42, p < .001$]. Two-component test figures took 465 msec to verify, but the large number of trials on which preparation of 3-component representations was not complete at the test figure onset inflated verification time for the more complex figures to 536 msec.

To test our notion of "spillover," we separated trials on which the subject initiated the test display from those on which the deadline was reached before preparation was completed. RT_2 for those trials on which the subject was not prepared when the 6-sec deadline was reached was 864 msec, while RT_2 for subject-initiated 3-component test figures was nearly as rapid as RT_2 for 2-component figures, just 483 msec.

2-sec deadline condition. As in the 6-sec condition, complexity affected both RT_1 and RT_2 in the 2-sec condition. Preparation of verification representations for 2-line descriptions was completed before the deadline on just nine of the 256 trials (3.5%), while preparation from 1-line descriptions beat the deadline on 167 of the 256 trials (65.2%).

Three-component figures took almost twice as long to verify as 2-component figures. The 2-sec deadline so severely limits RT_1 preparation for the 2-line descriptions that the spillover into RT_2 produces a larger complexity effect there than in RT_1 . Again, however, on the small number of 3-component trials on which subjects were prepared for the test figure when the deadline was reached, the effect of test figure complexity is almost absent. For these nine trials, RT_2 for 3-component test figures was 540 msec, not much slower than the 514 msec that subjects took to verify 2-component figures when they had signaled preparation prior to the deadline. The complexity effect clearly results from the "unprepared" state; RT_2 for 3-component figures when subjects did not beat the deadline was 1023 msec.

Simultaneous condition. The complexity and markedness effects in the simultaneous condition replicate earlier sentence verification work using this procedure. When the description and the test figure appear simultaneously, the time to decide whether the description describes the

figure is longer for the more complex displays. One-line descriptions using ABOVE take 130 fewer msec to verify than those using BELOW. The standard error of this difference is 47 msec, making this markedness effect highly reliable [$t(176) = 2.77, p < .005$]. The advantage of RIGHT over LEFT was of comparable size, 169 msec, but the greater variability of this difference ($SE = 122$ msec) left it just short of significance [$t(53) = 1.38, p < .10$].

Discussion

We conducted Experiment 2 for two reasons. First, we wanted to unequivocally demonstrate that the description representations in our subject-controlled comparison task preserved spatial complexity without producing a complexity effect. More importantly, we sought to understand the effects of methodological variation in spatial matching and verification tasks and their implications for a general model of spatial comprehension.

Replication of Experiment 1

In the subject-controlled condition, 2-line descriptions took no longer to verify than 1-line descriptions, even though they initially took almost twice as long to comprehend. Since the three-primitive figure set that we used makes it likely that the verification representation incorporates the display complexity, one plausible interpretation is that subjects construct description representations which preserve information about the spatial arrangement of all the component parts. Moreover, this representation and its associated processes do not maintain information about the initial linguistic elements in the description, since RT_2 is not affected by the linguistic markedness of the relational terms. We thus conclude that this representation is more spatial than linguistic in terms of its functional properties in the comparison process.

Effects of Task Differences

The four conditions of Experiment 2 span the set of spatial matching and verification tasks by using both simultaneous and successive presentations of the descriptions and test figures and both subject-controlled and experimenter-controlled displays. There are three results in each condition that together characterize the information included in the internal representations and the nature of the comparison and verification processes that operate on the representations. These results are (1) the effects of description and test figure complexity on preparation time and comparison time, (2) the relative speed of "same" and "different" responses, and (3) the effects of linguistic markedness of the relational terms in the descriptions on comparison time. We can now begin to understand the effects of task differences by examining the changing pattern of these three outcomes in different tasks.

Complexity Effects

The effect of description complexity on RT_1 is trivial; subjects take longer to understand two sentences than one sentence. However, since subjects take about 3 sec to study each sentence in a description in the subject-controlled condition, the much smaller complexity increments in the deadline conditions require interpretation.

In the deadline conditions, subjects might try to use the same representations and processing strategies that they chose in the subject-controlled task. They might continue to process the description to construct a complete representation for verification even after the test figure was presented. This strategy would inflate RT_2 by the difference between the deadline RT_1 and the RT_1 in the subject-controlled condition (less the few hundred msec saved by not making the response to initiate the test figure). Figure 5 shows a tradeoff between preparation and comparison time, but the pattern is not as simple as the preceding strategy would predict. As preparation time is reduced, verification takes more and more time and is slower by increasing amounts for the more complex test figures. But large decreases in preparation time produce considerably smaller increases in comparison time.

A clear demonstration that the information processing changes considerably as the verification task varies comes from comparing the subject-controlled and the simultaneous conditions. If subjects performed the same processing of the description and test figure in these two tasks, then total processing time ($RT_1 + RT_2$) for the two tasks should only differ by the time needed to signal completed preparation in subject-controlled RT_1 . In addition, the difference between the more complex and less complex displays should be the same in both conditions. Neither of these predictions is confirmed by the data.

If encoding and comparison operations were serial and independent in the deadline and simultaneous tasks, then this RT_1-RT_2 tradeoff would be nearly additive. Instead, the marked nonlinearity suggests that rather than completing their "normal" preparation by comprehending the description in isolation, subjects begin processing the test figure and thereby eliminate the need to construct complete representations of the initial description. In general, complexity effects in RT_2 signify that subjects are constructing the minimal representations of the description and test figure that will suffice for a particular trial.

Response Effects

As subjects have less time to prepare for the test figure, one clear sign of the corresponding changes in their representations and comparison strategies is the transition from faster "same" responses to faster "different" responses. In the subject-controlled condition, sub-

jects are instructed to construct a representation of the description that allows them to decide rapidly whether the test figure is the one they expect. A reasonable interpretation of the faster "same" responses in this condition is that subjects attempt to verify that the test figure is the one described and respond "different" if this matching attempt fails. At the other end of the task spectrum, the simultaneous condition, subjects have no time to generate expectations about the test figure. The faster "different" responses in this condition imply that subjects process the description and test figures at the same time and respond "different" as soon as a mismatch is detected.

The transition from a strategy of verifying "sames" in the subject-controlled condition to detecting "differents" in the simultaneous task necessarily implies an interaction with description complexity. Detecting "differents" seems to be the strategy of last resort that subjects adopt when they have insufficient time to prepare completely for the test figure. Since 1-line descriptions take less time to comprehend than 2-line descriptions, they are less affected by deadlines. Thus, subjects spend more time in the "prepared" state for the simpler descriptions and "sames" remain faster than "differents" for 2-component figures in conditions that produce the reverse for the 3-component figures.

Lexical Markedness Effects

The time subjects took to comprehend the 1-line descriptions in the subject-controlled and deadline conditions was affected in almost every case by the markedness of the relational terms in the descriptions. While the sizes of these ABOVE and RIGHT advantages in RT_1 were comparable to those reported by other researchers in a number of paradigms, they do not enable us to understand the effects of task changes on representations and processing because they do not vary systematically across the preparation conditions.

The pattern of markedness effects on comparison time (RT_2), however, is highly informative. There is a clear trend for the effects of linguistic markedness to decrease as RT_2 gets faster; and RT_2 gets faster when subjects take more time to prepare for the test figure. In the subject-controlled condition, subjects take the time to construct representations with properties that expedite comparison with a pictorial display, and therefore abstract spatial information away from the particular form in which it is presented. Such abstract spatial information is desirable if no knowledge of the particular test figure is available while the verification representation is being constructed. On the other hand, if some characteristics of the test figure are known while the description is being processed, it may be desirable to preserve specific information about the description. Thus with short (or zero) advance preparation time, description representations inevitably contain more information about the

actual lexical items in the description, and markedness effects are generated in the use of this information.

GENERAL DISCUSSION

In Experiment 1 we extended the typical description matching and sentence verification experiments to more complex displays with subject-controlled presentations. Our results were not consistent with any of the models for spatial comprehension. We concluded that figure and description representations in this task preserved spatial information in a form which can be compared with a figure display in some holistic or parallel fashion. Nevertheless, we did *not* wish to propose this as a general model of how spatial information is represented and compared. Just as our results had been inconsistent with previous models, a "general" model from our particular task would be unable to accommodate the data which led to the other models. Models developed in limited experimental contexts become models of tasks, and proposing them as general models of spatial comprehension requires considerable extension of theory beyond the supporting data. Instead, we have studied a number of experimental situations which comprise a family of spatial comparison and verification tasks.

Three critical features in Experiment 2—complexity effects, the relative speeds of "same" and "different" responses, and lexical markedness effects—had changing outcomes across the four tasks which illustrated how the structure of the task determined the results. What may appear to be minor modifications of procedure from fixed to subject-controlled display durations or from simultaneous to successive presentations completely eliminate the usual large effects of complexity and the small but critical effects of linguistic markedness. Rather than indicating representational or processing invariants, configurations of these effects characterize specific tasks. We now discuss how these characteristic outcomes depend on the nature of the task.

Fixed-Duration vs Subject-Controlled Display Presentations

In most previous description matching studies, the initial descriptions or figures were presented for fixed periods of time, instead of allowing the subject to determine the optimal encoding time as in Experiment 1 and in the subject-controlled condition of Experiment 2. For example, Cohen (1969) presented the initial displays for 2 sec, Nielsen and Smith (1973) used 4 sec, Seymour (1974c) used 1 sec, and Santa (1977) used a display duration of 2 sec. In these experiments, description complexity and lexical markedness effects were found, and these effects were more pronounced when the interval between the offset of the initial display and the onset of the test display was relatively short.

We believe that the difference between these results and our own lies in

the use of fixed presentation times which differed between experiments. Subjects in previous experiments may have attempted to generate an integrated verification representation from the separately-described parts, but may not have had sufficient time to generate this optimal representation.

The subject-controlled procedure is particularly sensitive to individual differences in encoding and verification that are obscured when all subjects study the initial display for the same period of time. MacLeod, Hunt, and Mathews (in press) recently used a subject-controlled verification procedure to assess the relationship between performance on a spatial comprehension task and more traditional indices of verbal and spatial ability. They attempted to fit their results to the constituent comparison model of spatial comprehension proposed by Carpenter and Just (1975), but they found that the intersubject variability was too great to be captured by a single model. Two groups differed systematically on psychometric tests and differed greatly in their allocation of time between RT_1 and RT_2 in the 2-reaction-time procedure, resulting in widely disparate fits to the Carpenter and Just model.

The MacLeod *et al.* study demonstrates that the same verification task can be approached in radically different ways by different people. Their results support our claim that spatial comprehension is a complex activity that can take a variable amount of time from trial to trial. A fixed duration procedure transforms the distribution of RT_1 times into a single value and thereby discards data about individual differences and task constraints.

Successive vs Simultaneous Display Presentations

The other methodological dimension along which we varied the tasks in Experiment 2 is that between successive and simultaneous presentation of descriptions and figures. Most of the current models for sentence-picture verification were developed in experiments using the simultaneous procedure (e.g., Clark & Chase, 1972, 1974; Carpenter & Just, 1975; Just & Carpenter, 1975). These models propose a sequence of independent encoding, comparison, and response stages whose durations are typically estimated by using a set of reaction-time tasks which presumably differ only in the presence or absence of the to-be-estimated stages.

The Carpenter and Just (1975) model has recently been criticized on both empirical and theoretical grounds (Catlin & Jones, 1976; Tanenhaus, Carroll, & Bever, 1976; but see Carpenter & Just, 1976). More generally, however, all of the models developed using the subtract-and-estimate methodology make assumptions which may not be warranted when the sentence and picture displays are presented simultaneously. In particular, researchers who use experimental tasks in which the two displays are presented together, rather than successively, can only assume that the two displays are encoded independently in time. The nonlinear tradeoff

between RT_1 and RT_2 in Experiment 2 argues against the independence assumption in the simultaneous condition.

A second questionable assumption of most stage models is that the comparison of the two displays requires complete representations of each. Such complete encoding may not be necessary when the two displays appear together or when the alternative displays are known. In Experiment 2, the transition from faster "same" responses in the subject-controlled condition to an advantage for "different" responses in the simultaneous condition suggests that subjects in the latter task forego attempts to construct complete description representations to verify "sames" and instead use a strategy of detecting "differents" via contingent and partial processing of the test figure.

Concluding Remarks

In conclusion, our results demonstrate that spatial comprehension and comparison processes are strongly affected by the structure of the sentence verification or description matching task in which they are usually studied. Models constructed from the outcome structure of single tasks rather than from the pattern of results in a set of tasks will not be general but remain models of specific tasks.

Our central results (the changing configurations of complexity and markedness effects and the speed of "same" and "different" responses) place constraints on a general account of spatial comprehension. In subject-controlled verification tasks, the sentence and figure representations preserve spatial information in an abstract form that can be used by the associated comparison operations in a holistic or parallel fashion. Thus our results exclude as a general representation for spatial comprehension any data structure which can not use spatial relations as parallel access paths (such as strictly-serial list structures, serial production systems, or procedural systems without parallel computation). The changing configurations of representational properties in Experiment 2 also exclude iconic or otherwise uninterpreted image representations as well.

Nevertheless, our results do not discredit the earlier models as analyses of the properties and processing most useful in particular task environments. While subjects can generate many representations for a given display, a particular experimental task makes one kind optimal. The demands of the verification task (such as the number of alternatives, their complexity or their similarity to one another) greatly influence the usefulness of different kinds of information and affect the way in which this information is embodied in a representation. The fact that qualitatively different sorts of representations and processes may be required for different tasks simply underscores the flexibility of the human information processor in dealing with spatial information.

But the purpose of studying individual tasks should be to discover

constraints that operate across tasks; a general model must relate the results from several experimental approaches. Therefore, a unifying account should be concerned not so much with specific tasks as with the structures of tasks which make particular representations and processing most appropriate. In this framework, we conclude that the nature of spatial comparison processes is mutable, task-constrained, and *not* a basic unchanging aspect of human cognition.

REFERENCES

- Carpenter, P. A., & Just, M. A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 1975, **82**, 45-73.
- Carpenter, P. A., & Just, M. A. Models of sentence verification and linguistic comprehension. *Psychological Review*, 1976, **83**, 318-322.
- Catlin, J., & Jones, N. Verifying affirmative and negative sentences. *Psychological Review*, 1976, **83**, 497-501.
- Chase, W. G., & Clark, H. H. Semantics in the perception of verticality. *British Journal of Psychology*, 1971, **62**, 311-326.
- Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **74**, 387-404.
- Clark, H. H., Carpenter, P. A., & Just, M. A. On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.
- Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, **3**, 472-517.
- Clark, H. H., & Chase, W. G. Perceptual coding strategies in the formation and verification of descriptions. *Memory and Cognition*, 1974, **2**, 101-111.
- Cohen, G. Pattern recognition: Differences between matching patterns to patterns and matching descriptions to patterns. *Journal of Experimental Psychology*, 1969, **82**, 427-434.
- Cooper, L. A. Individual differences in visual comparison processes. *Perception and Psychophysics*, 1976, **19**, 433-444.
- Glucksberg, S., Trabasso, T., & Wald, J. Linguistic structures and mental operations. *Cognitive Psychology*, 1973, **5**, 338-370.
- Just, M. A., & Carpenter, P. A. The semantics of locative information in pictures and mental images. *British Journal of Psychology*, 1975, **66**, 427-441.
- MacLeod, C., Hunt, E., & Mathews, N. Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior*. In press.
- Nielsen, G., & Smith, E. Imaginal and verbal representations in short-term recognition of visual forms. *Journal of Experimental Psychology*, 1973, **101**, 375-378.
- Olson, D., & Filby, N. On the comprehension of active and passive sentences. *Cognitive Psychology*, 1972, **3**, 361-381.
- Olson, G., & Laxar, K. Asymmetries in processing the terms "right" and "left." *Journal of Experimental Psychology*, 1973, **100**, 284-290.
- Paivio, A., & Bleasdale, F. Visual short-term memory: A methodological caveat. *Canadian Journal of Psychology*, 1974, **28**, 24-31.
- Santa, J. L. Spatial transformations of words and pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 418-427.
- Scheffe, H. *The analysis of variance*. New York: Wiley, 1959.
- Seymour, P. Response latencies in judgments of spatial location. *British Journal of Psychology*, 1969, **60**, 31-39.

- Seymour, P. Judgments of verticality and response availability. *Bulletin of the Psychonomic Society*, 1973, **1**, 196–198. (a)
- Seymour, P. Semantic representation of shape names. *Quarterly Journal of Experimental Psychology*, 1973, **25**, 265–277. (b)
- Seymour, P. Asymmetries in judgments of verticality. *Journal of Experimental Psychology*, 1974, **102**, 447–455. (a)
- Seymour, P. Generation of a pictorial code. *Memory and Cognition*, 1974, **2**, 224–232. (b)
- Seymour, P. Pictorial coding of verbal descriptions. *Quarterly Journal of Experimental Psychology*, 1974, **26**, 39–51. (c)
- Seymour, P. Stroop interference with response, comparison, and encoding stages in a sentence-picture comparison task. *Memory and Cognition*, 1974, **2**, 19–26. (d)
- Seymour, P. Semantic equivalence of verbal and pictorial displays. In R. A. Kennedy & A. W. Wilkes (Eds.), *Studies in long term memory*. London: Wiley, 1975.
- Smith, E., & Nielsen, G. Representations and retrieval processes in short-term memory: Recognition and recall of faces. *Journal of Experimental Psychology*, 1970, **85**, 397–405.
- Tanenhaus, M. K., Carroll, J. M., & Bever, T. G. Sentence-picture verification models as theories of sentence comprehension: A critique of Carpenter and Just. *Psychological Review*, 1976, **83**, 310–317.
- Tversky, B. Pictorial and verbal encoding in a short-term memory task. *Perception and Psychophysics*, 1969, **6**, 225–233.
- Tversky, B. Pictorial encoding of sentences in sentence-picture comparison. *Quarterly Journal of Experimental Psychology*, 1975, **27**, 405–410.
- Wannemacher, J. T. Processing strategies in picture-sentence verification tasks. *Memory and Cognition*, 1974, **2**, 554–560.
- Wannemacher, J. T. Processing strategies in sentence comprehension. *Memory and Cognition*, 1976, **4**, 48–52.

(Accepted April 11, 1978)