JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 19, 531–548 (1980)

Strategy Choice and Strategy Training in Sentence–Picture Verification

NANCY N. MATHEWS AND EARL B. HUNT

University of Washington

AND

COLIN M. MACLEOD

Scarborough College, University of Toronto

On each of three days, high and low spatial ability subjects verified the truth of simple sentences as descriptions of simple pictures. A two-reaction time version of a sentence picture verification task was used. On the first day (Free condition) 32 subjects were allowed free choice of strategies. The data replicated a previous finding of individual differences in strategy usage, with a hypothesized pictorial strategy used by 11 subjects, and a linguistic strategy used by the rest. Training in a pictorial and a linguistic strategy was counterbalanced over the second and third days, and the subjects yielded data in each condition virtually identical to their Free condition counterparts. The psychometric data provided additional support for the strategy distinction.

The process of comprehending whether a linguistic statement corresponds to a visual scene has intrigued psychologists because it seems to be an important element in many kinds of mental activity. Somehow, the two forms of stimuli must be converted to internal representations that can be compared. Do all persons form and compare these representations in the same fashion, even when the comprehensive problem is so rudimentary that the role of prior knowledge and education is presumably negligible? If different comprehension strategies occur, can they be understood sufficiently to be taught? In attempting to answer these questions, we will also address the issue of an individual's cognitive flexibility in strategy usage.

One of the more popular paradigms for investigating the coordination of perception and language is the *sentence–picture verification* paradigm developed by Clark and Chase (1972). An observer is asked to verify or reject a simple linguistic statement such as PLUS IS ABOVE STAR, or STAR IS NOT BELOW PLUS, as being a description of an equally simple picture, either ($\overset{+}{*}$) or ($\overset{*}{+}$). Clark and Chase have found that it takes longer to verify complex statements, such as the second example just given, than it does to verify simple statements such as the first. Furthermore, the time required to verify the sentence is a regular function of its linguistic complexity. Several models have been proposed to account for these findings (for a review see Carpenter & Just, 1975), but no one of them has been generally accepted as being *the* appropriate description of information processing in sentence verification.

In an earlier study, we questioned whether any one "general" model could account for subject's behavior in sentence verification experiments (MacLeod, Hunt, & Mathews, 1978). In particular, we claimed that some individuals approach a

We are grateful to Marcy Lansman for her helpful criticism. This research was supported by National Institute of Mental Health Grant MH-21795. Requests for reprints may be sent to: Earl B. Hunt, Department of Psychology, NI-25, University of Washington, Seattle, Wash. 98195.

0022-5371/80/050531-18$02.00/0

sentence verification task by reading the sentence, forming an image of the expected picture, and then comparing the presented picture to their image. Other individuals, according to our analysis, read the sentence and retain it in memory in verbal form, then describe the picture to themselves in verbal terms, and compare the two verbal descriptions. If it is true that some nonzero proportion of subjects in an experiment follows each strategy, then any analysis of group data which fails to consider individual subject strategies is bound to be suspect. We showed that if subjects are not instructed in "how to do the task," then different subjects will indeed follow different strategies. Although our data, averaged over subjects, were in substantial agreement with the expectations of a psycholinguistic model of sentence verification (i.e., one essentially based on a specific version of the verbal strategy just described), we also found that there was a very wide range in the degree to which the model fit individual data. Correlations between observed and model-predicted effects ranged from +.99 to −.89, and the distribution of individual fits was not normal. Further analysis of the data provided evidence for the existence of the two qualitatively different strategies that we have described. Subjects were classified as appearing to be following a linguistic or a visual-imaginal model. The best predictor of the reaction times of subjects following the linguistic model was a test of verbal aptitude. For subjects appearing to follow a pictorial model, the best predictor was a test of spatial aptitude. The stability of the findings was shown by the fact that the two aptitude tests had been taken two or more years before the experiment.

Our previous work relied upon observation of the strategies chosen by uninstructed subjects. Taken at their face value our results suggest yet another classification of people by their chosen cognitive style. Before reaching such a conclusion, however, one ought to ask how inflexible a cognitive style is. In this paper, we extend our earlier results in three ways. First, we replicate those results showing that different subjects use qualitatively different ways of approaching the sentence verification task. Second, a training procedure is instituted to test the accuracy of the models as characterizations of subject behavior. Finally, we show that almost all subjects can adopt either strategy at will. In keeping with our earlier work, the effects of cognitive abilities such as spatial ability, on the process of strategy acquisition were also examined.

The general design of the experiment is outlined below. University students participated for 3 consecutive days in a sentence verification study. On the first day they used whatever strategy they chose to use, i.e., no particular instructions regarding strategy were given. Their data were then subjected to an internal analysis to determine what strategy they had decided upon. On the second day, half the subjects were given brief instructions in the use of a verbal strategy, and half were given brief instructions in a pictorial strategy. On the third day, these instructions were reversed. Again, on the second and third days, appropriate strategy use was verified by an internal analysis of the data. Finally, the effectiveness of strategy use was related to the subject's general ability to perform spatial reasoning tasks, as assessed by conventional psychometric measures.

METHOD

Subjects and psychometric measures. Thirty-two university students were paid $3.00 an hour to participate in the experiment. They were selected from a pool of volunteers who had taken the Washington Pre-College Test (WPC) in their junior year of high school. The WPC is similar in form to the Scholastic Aptitude Test (see Note 1).

Two of the measures of the WPC, a Verbal Composite Score (VC) and a Spatial Ability Score (SA), were used to select the subjects. The VC score ($M = 50$, $SD = 10$)

is a measure derived from the Spelling, English Usage, Reading Comprehension, and Vocabulary subtests of the WPC battery. The SA score ($M = 50$, $SD = 10$) reflects the ability to visualize how a two-dimensional figure would look in three dimensions if it were folded along certain lines.

The sample was restricted to subjects of average verbal ability, i.e., students whose VC scores fell within one standard deviation of the population mean. For the 32 subjects, the mean VC score was 51.3 and its standard deviation 3.2. Half of these subjects were chosen to have high spatial ability and half to have low spatial ability. Each subject's score had to be at least half a standard deviation from the population mean. The mean SA score for the 10 men and 6 women in the Low Spatial group was 39.8 ($SD = 3.4$). The seven men and nine women in the High Spatial group had a mean SA score of 63.6 ($SD = 3.5$).

As subjects had taken the WPC some years before the experiment, four current psychometric scores were collected. Subjects were administered Form A of the Nelson–Denny Reading Test (1960) which yields a comprehension score and a reading rate measure. They were also administered the Two-Dimensional and Three-Dimensional Spatial tests from the Multiple Aptitude Tests Battery (Segal & Raskin, Note 2).

Stimuli. Sixteen different sentence–picture pairs were formed from the four binary dimensions, (STAR, PLUS), (IS, IS NOT), (ABOVE, BELOW), and ($\overset{*}{+}$, $\overset{+}{*}$). In Table 1, the eight possible sentences are paired with one of the pictures, demonstrating half of the sentence–picture stimuli.

Apparatus. Subjects sat in individual sound-attenuated booths. The stimuli were presented on Tektronix 604 display scopes, and responses were made using on-line keyboards. The experimental facilities allowed up to seven subjects to be tested simultaneously and independently.

Procedure. Each subject came to the laboratory for 3 days. A daily session consisted of instructions, practice, and experimental trials on the sentence–picture verification task, in addition to psychometric testing. A session took approximately 1 hour.

Ignoring the variations of the stimuli for the moment, the experiment included the within-subjects variable of instruction type (free, linguistic, and pictorial) and two between-subjects variables, high versus low spatial skills and order of instruction presentation (linguistic–spatial or pictorial–linguistic).

Since the instructions played a crucial role in the experiment, the important parts of them will be included verbatim. On Day 1 all subjects received the same general in-

TABLE 1
THE SENTENCE–PICTURE STIMULUS PAIRS BY TRIAL TYPE AND NUMBER OF CONSTITUENT COMPARISONS

Trial type	Sentence	Picture	Number of Constituent Comparisons
True Affirmative (TA)	PLUS IS ABOVE STAR STAR IS BELOW PLUS	+ *	K
False Affirmative (FA)	STAR IS ABOVE PLUS PLUS IS BELOW STAR	+ *	$K + 1$
True Negative (TN)	STAR IS NOT ABOVE PLUS PLUS IS NOT BELOW STAR	+ *	$K + 5$
False Negative (FN)	PLUS IS NOT ABOVE STAR STAR IS NOT BELOW PLUS	+ *	$K + 4$

Note. The number of constituent comparisons and trial type difficulty (TA < FA < FN < TN) are as predicted by Carpenter and Just (1975).

structions. Because the subjects were free to choose any strategy they wished to perform the task, this session will henceforth be called the *Free* condition. The Free instructions were:

> You are going to be asked to make judgments about whether a simple picture is true in relation to a sentence. (Two examples were shown and explained.) Here's how the task will work. First, you will see the sentence for as long as you need. For example, STAR IS ABOVE PLUS may appear. When you are ready for the picture, press either button. Immediately after, a picture, either a plus above star or star above plus, will appear. Your task is to indicate whether this picture is *true* with relation to the sentence you just read. If it is, press the TRUE button; if not, press the FALSE button. Then the next sentence will appear, and so on. What we are interested in is *how long* you spend on reading the sentence and on making your True–False judgment for the picture. You should try to go as *quickly* as you can, *without* making *errors*.

On the second day, half of the subjects, eight high spatial and eight low spatial, received the *Linguistic* strategy instructions. (On the third day these subjects were given *Pictorial* instructions.) The Linguistic instructions were:

> The computer will be presenting the same sentence–picture verification tasks as yesterday. However, today you will be solving the problems with a specific strategy which I will now describe. You might call today's strategy a verbal or linguistic strategy. Here's how it will work. When the sentence appears, look at it just long enough to read it and still remember it when the picture appears . . . Then, when the picture appears, you will have a set of words for the sentence and some words for the picture like PLUS ABOVE STAR or STAR ABOVE PLUS. While the picture is on the screen, you decide whether the two sets of words mean the same thing or not. If they do, you respond TRUE and if they mean something different, you respond FALSE. One of the important parts of this strategy is that you should *not* be forming any visual images or pictures in your mind of $\overset{+}{*}$'s or $\overset{*}{+}$'s. Particularly, while the sentence is on the screen, do not take any time before hitting the key to recode or transform the sentence . . . Let me add that for some of you this will be a very different strategy than the one you've been using before. As a result, it may seem very difficult at first. *Please* don't lapse back into your old way of performing the task. This new strategy will get easier and easier with practice.

Also on the second day, the other half of the subjects, eight high spatial and eight low spatial, were given *Pictorial* strategy instructions. (These subjects were switched to Linguistic instructions on their third day.) The critical portion of the Pictorial strategy instructions was as follows:

> . . . You might call today's strategy a spatial, pictorial or visual imagery strategy. Here's how it will work. When the sentence appears, read the sentence and, *before hitting the key*, form a picture in your mind of what the sentence describes. In other words, *do not hit the key* until you have formed the appropriate visual image for the sentence, either $\overset{*}{+}$ or $\overset{+}{*}$. Once you have the image or picture in your mind that the sentence represents, *then* you can hit the key. When the computer presents its picture, all you have to do is compare the imaged-picture from the sentence with the computer's picture. If they are the same, you respond TRUE and if they are different, you respond FALSE.

In addition to the instructions specific to that session, the subjects were given a brief review and were urged to ask questions. Then they were reminded to "do both parts of the task as *quickly* as possible *without making errors*." Finally, the feedback procedure and the use of the response keys were explained. Subjects were told to use the index finger of their dominant hand for true responses and the index finger of their nondominant hand for false responses.

The number of blocks and the number of trials per block were identical for each of the sessions. Subjects first received two blocks of practice trials with 16 trials per block. During the practice trials, the experimenter observed the subjects. When it seemed appropriate (e.g., because the subject did not appear to be using the requested strategy), the instructions were reviewed for an individual. Two blocks of 64 trials were then administered, with a short break between blocks. After the first set of 128 trials (Set 1), the subjects were given a longer break. When they returned to the booths, they completed two more practice blocks of 16 trials and two more blocks of

64 trials (Set 2). Each of the practice blocks contained a random ordering of all the possible sentence–picture stimulus pairs. Within a block of 64 trials, each of the 16 different stimulus pairs had four repetitions randomly distributed throughout the block.

On each trial, a warning dot was presented for 500 milliseconds, followed by the stimulus sentence. When finished with the sentence, the subject pressed a key and the picture immediately replaced the sentence. In their turn, both the sentence and the picture were presented in the middle of the scope screen. The first reaction time recorded on a given trial will be called the *comprehension RT,* and was measured from the sentence onset to the initial key press. The second reaction time will be the *verification RT,* and was measured from the picture onset until the true or false key was pressed. Immediately after the subject's true/false response, a feedback message was displayed for 500 milliseconds. If the subject made an error, WRONG was presented on the screen. If the subject was correct, RIGHT was displayed along with the subject's verification RT. Five hundred milliseconds intervened between the offset of the feedback and the next trial's warning dot.

RESULTS

Psychometrics

Table 2 summarizes the mean psychometric scores of the High Spatial group, the Low Spatial group, and both groups combined. The psychometric tests which were administered at the time of the experiment concurred with the high-school WPC scores. The two groups were matched on the current verbal measures, yet differed in their spatial skills. The Low Spatial group's performance on the Nelson–Denny Reading Comprehension Test ($M = 54.1$, $SD = 7.1$) was nearly identical to that of the High Spatial group ($M = 54.9$, $SD = 7.6$). However, the High Spatial group scored significantly better than the Low Spatial group on both the Two-Dimensional ($t(30) = 2.98$, $p < .05$) and the Three-Dimensional Spatial Tests ($t(30) = 3.79$, $p < .001$).

The correlational relationships among the verbal and the spatial skills and the old and new measures are shown in Table 3. Both the sample's restricted range of verbal ability and the relatively long time period between psychometric sessions probably served to reduce the magnitude of the correlations, yet the verbal and comprehension tests were still significantly related to one another. There are significant positive correlations among all the spatial tests as well. This particular sample of subjects shows no relationship between their verbal and spatial skills; we have observed similar low relationships in college populations in other studies.

Reaction Times, Outliers, and Errors

When examining the nature of strategies that were used by the subjects either spon-

TABLE 2

MEAN PSYCHOMETRIC MEASURES OBTAINED FOR THE TWO GROUPS (WPC VERBAL AND SPATIAL WERE USED TO DEFINE THE GROUPS)

	WPC			Current			
	Verbal composite	Reading comprehension	Spatial** ability	*N*-D comprehension	Reading* rate	2-D* spatial	3-D** spatial
Overall	51.4	52.7	51.7	54.5	301	18.6	17.6
Low WPC spatial ability	50.9	53.6	39.8	54.1	338	17.0	15.9
High WPC spatial ability	51.8	51.8	63.5	54.9	263	20.3	19.3

Note Significance level for the difference between the two groups: * $p < .05$; ** $p < .001$

TABLE 3
PSYCHOMETRIC CORRELATIONS

	Reading comprehension	Spatial ability	*N*-D comprehension	Reading rate	2-D spatial	3-D spatial
Verbal composite	*.30*	.08	*.43*	−.02	.19	.20
Reading comprehension		−.07	*.42*	.08	−.04	−.03
Spatial ability			.01	*−.43*	*.50*	*.61*
N-D comprehension				.03	.07	.16
Reading rate					*−.32*	*−.34*
2-D spatial						*.38*

Note. Italicized correlations significant at $p < .05$ or greater.

taneously or through training, well-practiced subjects and stable data are essential. Thus, the analyses to be reported were based on the final set of 128 data trials in each instructional condition.[1] In a given condition, then, the first 192 trials are considered to be practice. Later in this paper, when the acquisition of the different strategies is examined, only the 64 labeled "practice" trials will be omitted from the analyses.

Analysis was restricted to the reaction times (RTs) that were typical of a particular subject for a particular condition and trial type. For each subject we computed means and standard deviations of the comprehension RT for affirmative and negative trials and of the verification RT for the four trial types, in each of the six sets of data (two sets in each of three instructional conditions). Any comprehension or verification RT further than three standard deviations from its respective mean was discarded. This procedure eliminated slightly less than 4% of the RTs, most of which were less than 200 or greater than 5000 milliseconds.

The overall experimental error rate was slightly less than 6%. The highest percentage of errors, 8%, occurred in Set 1 on the first day. By Set 2, the number of errors dropped significantly, and then remained stable in the subsequent conditions. The percentages of errors for the Set 2 data in the Free, Linguistic, and Spatial conditions were 5.2, 6.2, and 4.1, respectively.

Speed–accuracy trade-off problems do not seem to be present in the data, either at the level of subjects or at the level of trial types. Across subjects, the six correlations of errors with comprehension or verification RTs were all nonsignificant and ranged from −.01 to +.26. In other words, the faster subjects did not tend to make more errors. Both errors and RTs increased with the number of hypothesized operations in the constituent comparison model, yielding positive correlations of .45, .03, and .26 for the Free, Linguistic, and Spatial conditions, respectively.

Free Condition

Entire group performance. The subjects spent an average of 1908 milliseconds examining the sentence display. These Comprehension RTs were analyzed by a 2 × 2 × 2 ANOVA. The factors were spatial ability, order of instructions (as a control for the later conditions), and sentence complexity (affirmative versus negative). Sentence complexity produced the only significant value, $F(1,28) = 37.2$, $MS_e = 232016$, $p < .001$. Subjects took much longer to comprehend a negative sentence ($M = 2275$, $SD = 1276$) than an affirmative one ($M = 1541$, $SD = 752$), a standard finding.

The Verification RTs ($M = 963$, $SD =$

---

[1] Although not reported here, analyses were also conducted on the Set 1 data and on Sets 1 and 2 combined. In each case, the same general patterns of results emerged. Aside from diminishing practice effects, allowing more trials seemed to help the strategies in the free condition emerge more clearly. That is, three subjects' data no longer fell in the middle between the two strategy groups.

366) were analyzed using a 4-way ANOVA, the factors being spatial ability, instruction order, and the true/false, and affirmative/negative trial types. Three significant effects were found. False responses to the pictorial display were longer than true responses, $F(1,28) = 11.0, MS_e = 20251, p < .001$, and responses to negative sentences took longer than responses to affirmative sentences $F(1,28) = 31.2, MS_e = 94028, p < .001$. Finally the interaction of these two factors was significant, $F(1,28) = 5.2$, $MS_e = 18605$, $p < .05$. Carpenter and Just's (1975) constituent comparison model predicts that the means as a function of trial type will be ordered TA, FA, FN, and TN. The obtained data were 747, 885, 1133, and 1104 milliseconds respectively, with Carpenter and Just's linearity prediction accounting for 92.7% of the variance in the verification RT over the four trial types. The order of trial types is "not quite right," however, because the Carpenter and Just model predicts FN < TN. The residual 7.3% of the variance just approaches significance, $F(2,93) = 2.68$, $p < .10$.

Individual fits to a linguistic model. The effect most consistently predicted (and observed) by linguistic models is the difference in reaction times between negative and affirmative sentences. As in many other studies, negative trials took significantly longer than affirmative trials (Carpenter & Just, 1975). Certainly if the linguistic strategy we described were to be used, negative trials would take longer than affirmative trials.

It should be noted that there is a linguistic strategy that does destroy the negative–affirmative difference. Subjects could simply recode the negative sentences to affirmative sentences during the sentence comprehension period. Note that his strategy leaves affirmative sentences unchanged. Clark and Chase (1972) report that some subjects use this strategy.

With the above arguments in mind, a frequency distribution of the negative–affirmative differences for the Free verification RTs was plotted. The distribution of difference scores showed approximately one third of the scores grouped around zero. The remaining scores were clearly positive with a mean of 461 milliseconds. The two distinct clusters of subjects were confirmed by a variant of Fisher's clustering algorithm for one-dimensional data (Hartigan, 1975). We rank-ordered the negative–affirmative difference scores and then iteratively divided them into different pairs of subgroups until a t test between the groups was maximized. Table 4 shows the means and ranges of the negative–affirmative difference scores for the two groups. Clearly, the verification RTs of the Well Fit group reflect the linguistic complexity of the sentence.

The Poorly Fit group's data does not seem to be consistent with any linguistic strategy. It is obviously inconsistent with a strategy that does not involve recoding of negative sentences. It is also inconsistent with a strategy that is linguistic but does recode negative sentences, because the RTs for affirmative sentences, which are unchanged by a linguistic recoding, are much shorter in the Poorly Fit than in the Well Fit group.

Like MacLeod et al. (1978), we found evidence pointing to the use of at least two different strategies to solve the sentence verification task. However, the present division of the subjects into two groups was done somewhat differently than in the previous study. MacLeod et al. computed each individual's fit to a specific linguistic-based model, the constituent comparison model (Carpenter & Just, 1975), by computing the correlation between obtained RTs and the model-predicted pattern for each subject.

TABLE 4

MEAN NEGATIVE–AFFIRMATIVE DIFFERENCES IN VERIFICATION RT FOR THE FREE CONDITION AS A FUNCTION OF STRATEGY

Group	Number of subjects	Range of differences	Mean difference
Well fit	21	82 to 1059	461
Poorly fit	11	−71 to 44	−1

The previously described clustering algorithm was then applied to these correlation coefficients. If we split the Free condition data using either the method employed by MacLeod et al. (1978) *or* by simply using their correlational cut-off values for "well fit" and "poorly fit" to the constituent comparison model, we obtain the identical two groups defined by the negative–affirmative data. Table 5 shows the accuracy with which the Carpenter and Just model fits the data from each group. Thus, we have identified the two strategy groups using two different but, we believe, equally reasonable definitions of compatibility of a person's data to a linguistic model. In our earlier study we dealt with one specific model, the constituent comparison model, while in this study we have generalized to the set of all linguistic models.

Figure 1 displays the verification RTs for each group as a function of the number of constituent comparisons hypothesized by Carpenter and Just's (1975) linguistically based model. Two features are worthy of note. First, performance of the Poorly Fit subjects is much faster than that of the Well Fit subjects. Their mean verification RT is quicker on the average by 562 milliseconds. Second, the pattern of reaction times across trial types is markedly different for the two groups. The relationship between the constituent comparison model and the Well Fit data is very good. With a slope of 113.8 and an intercept of 764 milliseconds, the linear relationship accounts for 98% of the variance. It is obvious from Figure 1 that the model does *not* describe the Poorly Fit data at all. An identical picture was obtained in our previous study.

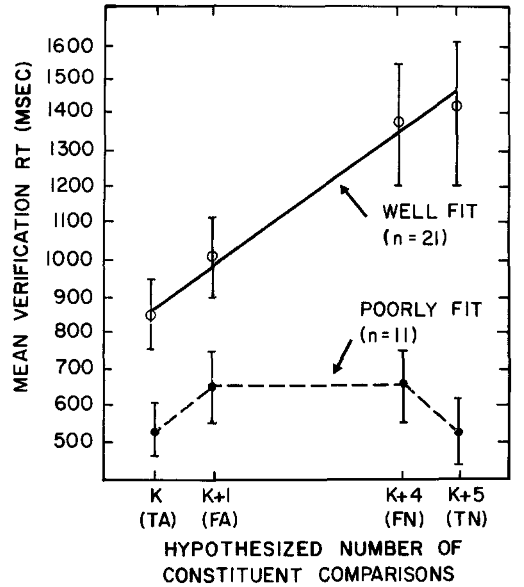

FIG. 1. Verification data from the two strategies chosen in the Free instructions condition. Mean verification RT is plotted as a function of the number of hypothesized constituent comparisons (trial type is shown in parentheses). Also included are the 95% confidence intervals and the best-fitting straight line for the Well Fit subjects only. (Intercept at 764 msec, slope at 114 msec.)

TABLE 5

INDIVIDUAL SUBJECT CORRELATIONS TO THE CONSTITUENT COMPARISON MODEL AS A FUNCTION OF GOODNESS OF FIT AND INSTRUCTIONS

Instructions	Number of subjects	Range of correlations	Median correlation
Free			
Well fit	21	.712 to .988	.938
Poorly fit	11	−.584 to .253	.021
Linguistic	32	.501 to .999	.945
Spatial	32	−.645 to .976	.275

Note. In the Free Instructional condition, subjects are split according to their choice of strategy.

Separate 2 × 2 ANOVAs, with factors of true/false and affirmative–negative, were run on each of the two groups' verification data. For the Well Fit groups, the familiar pattern of significant results occurred. Negative trials took significantly longer than affirmative trials, $F(1,20) = 64.2$, $MS_e = 70094$, $p < .001$, and false trials took slightly longer than true trials $F(1,20) = 3.6$, $MS_e = 26835$, $p < .07$. The true/false by affirmative–negative interaction was also significant, $F(1,20) = 6.7$, $MS_e = 22330$, $p < .05$. That is, negation on a true trial resulted in a slower reaction time than on a false trial, while on affirmative trials the converse was true. For the Poorly Fit group, only the true/false factor was significant, $F(1,20) = 45.3$, $MS_e = 3120$, $p < .001$.

The main and interaction effects related to linguistic complexity yielded F ratios of less than 1.

If our analysis is correct, the linguistic marking effect should appear only in the verification times of the Well Fit group. A comparison of trials for sentences containing ABOVE and BELOW showed a marking effect of 65 milliseconds ($t(20) = 2.95, p < .005$) in the Well Fit group, and 12 milliseconds ($t(10) = 1.62, p > .05$) in the Poorly Fit group.

If the Poorly Fit subjects are using a strategy which does not operate on linguistic information at the time of verification, what kind of strategy are they using? The very short verification RTs and the absence of an affirmative–negative or marking effect leads to the conclusion that these subjects processed the sentence information until it was amenable to a direct comparison with the pictorial display. If we assume for a moment that the subjects are using the pictorial strategy outlined in the introduction, then their comprehension RT should contain the additional time to convert the sentence representation into a pictorial representation. The Poorly Fit group's mean comprehension RT would have to be longer than that of the Well Fit group. In fact, this is true, as shown in panel (b) of Table 6. The overall mean comprehension and verification RTs show that the Poorly Fit subjects take almost a second longer to process the sentence than subjects in the Well Fit group, who are using a linguistic strategy. It would appear then, that the Poorly Fit group's data are in accord with the pictorial model.

TABLE 6

COMPARISON OF MEAN OVERALL COMPREHENSION AND VERIFICATION RT: SPONTANEOUS STRATEGY USAGE VERSUS INSTRUCTIONS IN STRATEGY USAGE

Group	Comprehension	Verification
(a)		
Well fit* ($n = 43$)	1652	1210
Poorly fit* ($n = 16$)	2579	651
(b)		
Well fit ($n = 21$)	1577	1156
Poorly fit ($n = 11$)	2524	594
(c)		
Linguistic instructions	961	1226
Spatial instructions	2515	559

* Data from MacLeod, Hunt, and Mathews (1978). These subjects had 160 less practice trials than the subjects in the Free condition (panel b).

The Free data separate neatly into two groups—the Well Fit group, whose reaction times can be described by a linguistic model, and the Poorly Fit group, whose data seem to fit a pictorial model. Since these models were postulated from the data of MacLeod et al., ideally the Free condition groups should resemble the groups of the previous study. The same patterns of reaction times across trial types, as shown in Figure 1 and the results of the 2 × 2 verification ANOVA, were obtained in the previous study. A comparison of panel (a) and panel (b) in Table 6 demonstrates almost no difference between the pattern of RTs for the two experiments.

MacLeod et al. reported that persons who chose the pictorial strategy had higher spatial ability scores than individuals who chose the linguistic strategy. The present study provided only very weak evidence in support of the observation. Six of the 16 subjects in the high spatial group chose the pictorial strategy, while 5 of the 16 in the low spatial group chose it. We point out, however, that this study was not designed to replicate the tendency to choose strategies that was observed earlier and, indeed, our subject sampling procedures may have made this unlikely. All subjects were required to have verbal ability scores within intermediate range. Since the average verbal ability of university students is quite high in an absolute sense (students are, after all, selected on largely verbal criteria), almost all our subjects may have been biased toward choice of verbal strategies. Naturally, this is speculation. A much larger study with a better defined population would be required to obtain an accurate picture of the prevalence of different strategies in different ability-level groups. While this might be a useful thing to do if one wished to study problem solving in groups of wide ranging ability, the data on training, that we report below, raises some question as to whether the question is terribly important with respect to the university student population.

Summary. Not only have we replicated the MacLeod et al. finding of reliable individual differences in strategy usage for solving the sentence verification task, we have also found the same set of strategies. In both experiments, two clearly different strategies were isolated. Although different methods were used to separate the strategies, the Well Fit and Poorly Fit groups of both experiments had virtually identical data. Ranges and means for the "fit" parameter to the constituent comparison model and for the negative–affirmative difference scores were similar. Furthermore the patterns of reaction times within a strategy group were the same across the experiments. Finally, comparison of the overall mean comprehension and verification RTs showed nearly the same levels of performance.

Instruction Effects

The instruction conditions provide a test of our hypothesized linguistic and pictorial models. If the models are to be supported, subjects trained in the linguistic and pictorial strategies must produce data which are indistinguishable from the Free condition Well Fit and Poorly Fit groups, respectively. The training conditions also examine the range and flexibility of subjects' comprehension processes. In the MacLeod et al. (1978) study, very few low spatial subjects used the pictorial strategy. The training procedure allows us to investigate the availability of the linguistic and spatial strategies to both high spatial and low spatial subjects. In addition, the subjects' ability to change their approach to the task at any given time will be examined. That is, can an individual be instructed to change from the linguistic strategy to a pictorial strategy and vice versa?

Figures 2 and 3 show comprehension and verification RTs as a function of instructions and of subject's spatial ability. Clearly, instructions had a major effect. The comparison of Figures 1 and 2 shows that the data for all subjects under a particular instructional condition are similar to the data obtained in the Free condition from subjects who were thought to have been using the instructed strategy on their own. Because the differences between conditions are large, and because variances as well as means are affected, the data within each instructional condition will be analyzed separately.[2]

Linguistic instructions. Spatial ability and order of instructions had no significant effect on either the comprehension or verification reaction times. High and low spatial subjects produced similar data, regardless of whether they had been previously trained on a different strategy. Giving instructions and practice on the linguistic strategy caused the comprehension RT to decrease and the verification RT to increase from the Free condition levels, which would be expected if the instructions elimi-

---

[2] For those who are not perturbed by the large violations of the homogeneity of variance assumption, a 5-way ANOVA was run on the verification RTs. The factors were spatial ability, order of instructions, instructions, true/false, and affirmative/negative. The main effects of spatial ability and presentation order were not significant at $\alpha = .05$. The other effects listed below were significant, even with $\alpha = .001$. The effect of instructions, of course, was highly singificant with an $F(2,56) = 52.6$. The remaining two main effects were also significant. The fact that the instructions interacted significantly with the affirmative/negative dimension and with the true/false by affirmative/negative interaction, argues for looking at the data in each instructional condition separately.

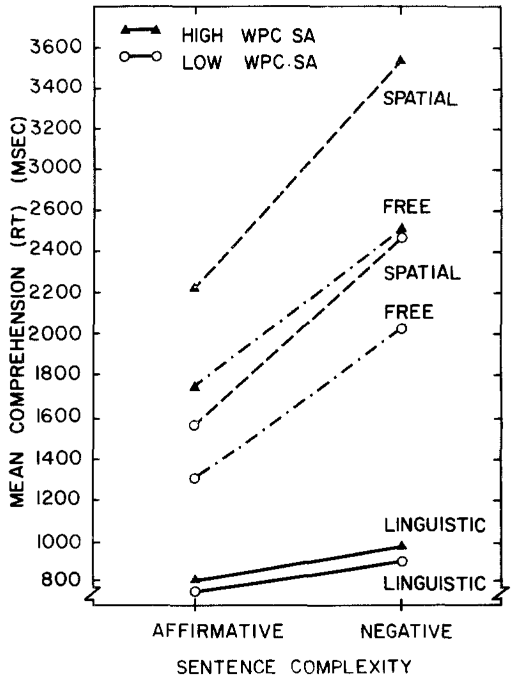

FIG. 2. Mean comprehension RTs for the high and low spatial groups as a function of instructions and sentence complexity (affirmative vs negative).

nated the use of a spatial strategy with its associated long comprehension times. For the comprehension RTs, sentence complexity produced the only significant effect, with negative sentences taking longer to comprehend than affirmatives, $F(1,28) = 47.9$, $MS_e = 8285$, $p < .001$. For the verification RTs, negatives required more time than affirmatives, $F,(1,28) = 117.0$, $MS_e = 63676$, $p < .001$, and falses took longer than trues, $F(1,28) = 12.6$, $MS_e = 7569$, $p < .001$. Once again the significant true/false by affirmative/negative interaction occurred, $F(1,28) = 13.2$, $MS_e = 25555$, $p < .001$, with TA trials faster than FA trials but TN trials slower than FN trials.

The group data were then fit to the linguistic constituent comparison model. The verification means for the four trial types were correlated with the number of constituent comparisons predicted by the model. An excellent fit was obtained, with 98.7% of the variance accounted for by the linear relationship between the obtained RTs and the predicted ordering. The best fitting line through the four verification means had a slope of 119.6 and an intercept of 795 milliseconds. At the level of individual subjects, fits to the model ranged from .501 to .999 with a median of .945. Both overall and at the level of individuals, instructions to perform a linguistic strategy on the verification task produced data well fit by a linguistic model.

Pictorial instructions. As in the Linguistic condition, the pictorial condition data showed no significant effects of instruction order. Figures 2 and 3 show that, with pictorial strategy instructions, both the high and low spatial groups shift their patterns of responding in the same fashion. The comprehension RT ($M = 2457$, $SD = 1108$) was much slower than in the Free and expecially in the Linguistic conditions. Conversely, the verification RT ($M = 524$, $SD = 155$) became *very* fast. On some trial types, the difference between the linguistic and pictorial verification reaction times is as much as 1 second.

Although the high and low spatial groups performed almost identically in the linguistic condition, they differed somewhat in the pictorial condition. High spatials spent significantly more time on the comprehension RT than did low spatials, $F(1,28) = 5.5$, $MS_e = 2140532$, $p < .03$, while they spent less time on the verification RT, $F(1,28) = 2.4$, $MS_e = 88691$, $p < .13$. Further discussion of the differences related to spatial ability will be deferred until the section on strategy acquisition and psychometric skills.

The pictorial strategy comprehension RT showed a significant sentence complexity effect, $F(1,28) = 70.3$, $MS_e \times 293392$, $p < .001$. The mean negative sentence RT was 3022 milliseconds and the mean affirmative sentence RT was 1886. Analysis of the verification RTs showed the usual true/false effect, $F(1,28) = 28.8$, $MS_e = 2750$, $p < .001$. Surprisingly, however, there was also a small but reliable affirmative/negative ef-

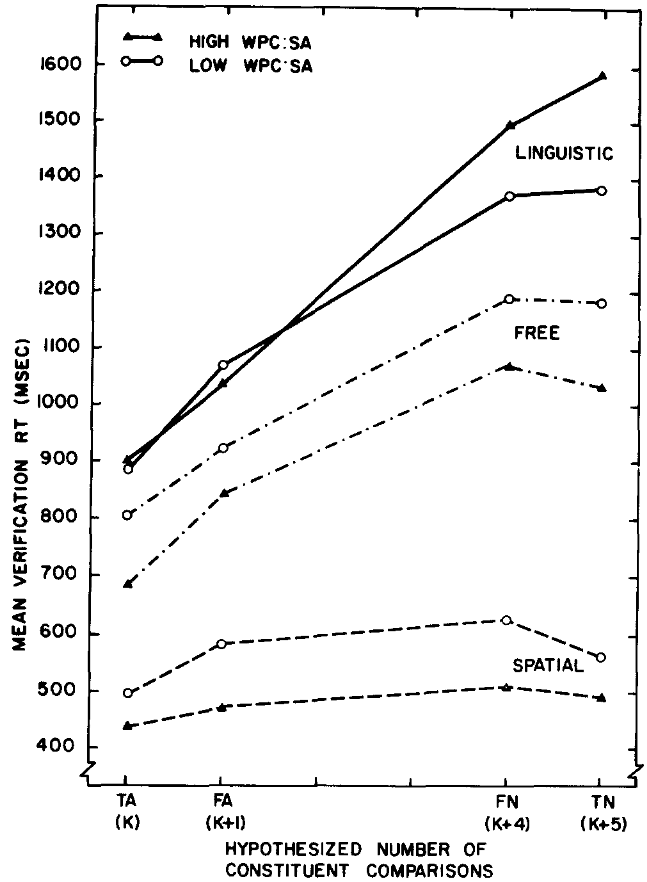

FIG. 3. Mean verification RTs for the high and low spatial groups, as a function of instruction and hypothesized number of constituent comparisons.

fect, $F(1,28) = 9.5$, $MS_e = 9158$, $p < .001$, with negative trials taking an average of 52 milliseconds longer to complete than affirmatives. Still, this is much smaller than the corresponding effect in the linguistic condition. Unlike the Free and Linguistic conditions, the true/false by affirmative/negative interaction was *not* significant ($F < 1$).

As a pictorial model would predict, there was no effect of the linguistic marked/unmarked dimension on verification RT. The marked/unmarked difference scores indicate that the subjects dealt with the dimension at the time of the comprehension RT, rather than at the time of the verification RT (82 versus 10 milliseconds). Note that this is a different pattern of results than those obtained in the linguistic condition where the mean marked/unmarked difference score was 11 milliseconds for the comprehension RT and 79 milliseconds for the verification RT. An analysis of variance on the difference scores confirmed this—with factors of spatial ability, comprehension/verification and linguistic/pictorial conditions, only the interaction of comprehension/verification RT with linguistic/pictorial instructions was significant, $F(1,30) = 6.0$, $MS_e = 25883$, $p < .02$.

Instructions to use a pictorial strategy produced very difficult data from those generated under the linguistic instructions. However, the pictorial verification RTs should not have had a significant affirmative/negative effect, if all the subjects had performed the strategy as directed. Were the instructions inadequate for training a subject to use the pictorial strategy consistently or had just a portion of the subjects failed? Examination of the individual's data found the latter to be true. Negative–affirmative difference scores were computed for each subject. While the median score was appropriately small, 18 milliseconds, the range was great. Difference scores ranged from −49 to a high of 387 milliseconds. Of the 32 subjects, 6 two high spatials and four low spatials (all six had three-dimensional spatial scores below the median), apparently did not fully acquire the pictorial strategy. They were the only subjects for whom the linguistic constituent comparison model could account for 50% or more of the variance in their verification RTs and whose negative–affirmative difference scores were greater than one standard deviation from the mean. Their mean negative–affirmative difference score was 236, with the smallest difference score being 118 milliseconds. With those 6 subjects removed, the remaining 26 subjects were found to have an appropriately negligible negative/affirmative effect (M = 8.8, SD = 29.6).

We found that all six of the subjects who had not performed the pictorial strategy after training had spontaneously chosen the linguistic strategy in the Free condition. Although we had counterbalanced order of instructions, it had not been feasible to balance subjects with respect to initial strategy choice. To assure ourselves that the initial strategy choice had not influenced the linguistic and pictorial results, we reexamined the comprehension and verification RTs of the linguistic and pictorial instruction conditions. The groups we contrasted were the 21 subjects who had initially chosen the linguistic strategy and the 11 subjects who had chosen the pictorial strategy in the Free condition. The conditionalized means are listed in Table 7. A set of t tests compared the two groups on the linguistic comprehension RTs, linguistic verification RTs, pictorial comprehension RTs, and pictorial verification RTs. None of the four t values even approached significance. Thus, on the average, initial strategy choice did not influence later training on the same or a different strategy. Although most subjects seem able to adopt either strategy, there do seem to be exceptions, at least for the small amount of training we did.

Relation to the other conditions. We have shown that, through training, subjects can learn and change strategies for solving a simple comprehension task. Their absolute levels and patterns of reaction times shift with instructions. How does performance with these instructed strategies compare with performance on the spontaneously selected strategies? The slope (113.8) and intercept (764.0) for the Well Fit group are almost the same as in the Linguistic instruction condition (slope = 119.5, intercept = 795.3). Fits to the Carpenter and Just constituent comparison model for the Free versus the Linguistic and Pictorial training conditions compare favorably in Table 5. Similarly, the ranges and means of the negative/affirmative difference scores for the Linguistic condition (range =

TABLE 7

MEAN COMPREHENSION AND VERIFICATION RTS FOR THE LINGUISTIC AND SPATIAL CONDITIONS AS A FUNCTION OF STRATEGY CHOICE IN THE FREE CONDITION

Group	Strategy Instructions	Comprehension RT	Verification RT
Free condition			
well fit	Linguistic	859	1224
(n = 21)	Spatial	2410	554
Free condition			
poorly fit	Linguistic	890	1183
(n = 11)	Spatial	2544	465

81–1161, $M = 484$) and Pictorial condition (range = -49–387, $M = 51$) are quite close to the values for the Well Fit and Poorly Fit groups listed in Table 4. Finally, consider the overall mean comprehension and verification RTs for this experiment and the MacLeod et al. experiment listed in Table 6. When the subjects in this experiment were given linguistic or pictorial instructions (panel c), their data resembled that of the subjects who spontaneously chose linguistic or pictorial strategies (panels a and b). These data compare very favorably, especially considering the differing amounts of practice in the two studies. The relatively shorter Linguistic comprehension RT is probably the result of the Linguistic instructions emphasis on *not* recoding (i.e., NOT BELOW into ABOVE).

Strategy Acquisition and Psychometrics

We have shown that most subjects can be trained to acquire a new strategy and to change strategies, regardless of whether they initially selected the strategy themselves. Now we will examine the relationship between psychometric measures and strategy performance in the sentence verification task. Since we severely restricted the range of verbal ability in our sample, those results will be mentioned only briefly. Spatial ability was specifically varied, and so can be looked at with regard to the ease of strategy acquisition and the quality of strategy performance.

Verbal skills. Even with the restricted range of verbal ability in our sample, we replicated the MacLeod et al. finding that performance on the verification task is related to traditional psychometric measures and taps some aspect of the comprehension process. Reading rate was highly related to the comprehension RTs in the Free and Linguistic conditions, $r = -.61$ and $-.51$, and to a lesser extent in the Pictorial condition, $r = -.31$. (Note that under pictorial instructions subjects must both read and form an image.) The negative correlations are expected, as faster reading rates are being paired with shorter comprehension RTs. Verification performance in the Free and Linguistic conditions was also related to the reading comprehension scores ($r = -.31$ and $r = -.32$, respectively). The better a subject's reading comprehension skills, the faster his verification reaction times when using a linguistic strategy.

Spatial skills. When we discussed the instruction effects, we found no differences between the high and low spatial subjects' performance in the Free or Linguistic conditions—neither errors nor reaction times differed significantly. Therefore this section will focus on the effects of spatial ability in the pictorial condition. First, we felt that the reaction time data from Set 2 should be reanalyzed to assure that the pattern of results obtained using the over 2-year-old WPC spatial measure would remain when the spatial groups were defined by a current spatial measure. We chose to use the Three-Dimensional Spatial Test of the Multiple Aptitude Test Battery, as it appeared to be the most sensitive score we had available. It should be noted that the classification of subjects as "High" and "Low" spatial by the current measures is not exactly the same as the classification using the measure obtained some years before, although the two measures are related (cf. Table 3).

The subjects' Three-Dimensional spatial scores were rank ordered and a median split was performed on the data. The 16 subjects whose scores fell above the median were treated as high spatial subjects and the remaining 16 were classified as low spatial. As with the WPC spatial groups, there were no differences between the high and low spatial groups in the Free and Linguistic conditions. In the Pictorial condition, the high versus low spatial differences were slightly larger than those obtained when the groups were based on the older WPC spatial score. As before, high spatials spent more time than low spatials (a difference of 817 milliseconds) on the Pictorial condition comprehension RT, $F(1,30) = 5.0$, $MS_e =$

2152342, $p < .05$. The 2 × 2 × 2 ANOVA (spatial ability × true/false × affirmative/negative) of the verification RTs in the Pictorial condition indicated that the low spatials were significantly slower than the high spatials at verifying the picture, requiring 134 milliseconds more on the average to make their judgment, $F(1.30) = 7.0$, $MS_e = 81798, p < .01$. The verification RT also showed a significant true/false effect, $F(1,30) = 26.3$, $MS_e = 3011, p < .001$ and an affirmative/negative effect, $F(1,30) = 9.3$, $MS_e = 9333$, $p < .001$. However, the affirmative/negative effect was qualified by a spatial ability by affirmative/negative interaction, $F(1,30)$, $= 5.2$, $MS_e = 9333, p < .05$. Low spatial subjects had a negative–affirmative difference score averaging 91.1 milliseconds, as opposed to the negligible 13.2 difference score of the high spatials. This suggests that some low spatial subjects had difficulty adopting a pictorial strategy, a topic to which we now turn.

There were significant differences in the way high and low spatial subjects performed the pictorial strategy. Low spatials spent significantly less time on the comprehension RT than did high spatials when they were supposed to be generating a pictorial image of the sentence. When it came time to verify the sentence against the picture, low spatials took significantly longer than the high spatials. One might hypothesize that the high spatial subjects are taking somewhat longer at the time of the sentence to generate a richer or more complete pictorial image of the sentence. When the actual picture appeared, it was easier for them to make the verification. The absence of a negative/affirmative effect in the verification data of the high spatials argues for their successful transformation of the sentence. The verification RTs of the low spatials continued to reflect the sentence's negative/affirmative dimension. This result may indicate that the low spatials either formed a hasty, impoverished visual image during the comprehension time, or that they gave up part way through the transformation. As previously noted, not all low spatial subjects seem to have been able to adopt a spatial strategy.

Acquisition. Thus far we have discussed the effects of spatial ability upon strategy performance when the subject has had ample time to acquire the strategy. The speed of strategy acquisition will now be examined. Each person had previously completed a set of 128 experimental trials, Set 1, before doing the 128 trials in Set 2.[3] Thus, if we include a subject's early performance, the Set 1 data, we can divide a strategy session into four blocks of 64 trials each. Strategy acquisition can then be studied from the beginning of a session across blocks. We hypothesized that low spatial subjects would have more difficulty than high spatial subjects acquiring the pictorial strategy, but not the linguistic strategy.

Separate 2 × 2 × 4 ANOVAs, with factors of spatial ability, instruction order, and blocks, were used to analyze the linguistic and pictorial verification RTs. In the Linguistic condition, there were no significant effects. Subjects did not change their performance significantly across blocks and the high and low spatial ability subjects did equally well. A quite different picture applied to acquisition of the pictorial strategy. This is shown in Figure 4. When the same ANOVA was applied to the pictorial verification data, the spatial groups differed significantly, $F(1,28) = 4.0$, $MS_e = 89974$, $p < .05$, performance improved across blocks, $F(3,84) = 20.3$, $MS_e = 2807$, $p < .001$, and the spatial ability by block interaction nearly reached significance, $F(3,84) = 2.65$, $MS_e = 2807, p < .06$. Performance of the high spatials was relatively constant across blocks. They seemed to have acquired the strategy immediately. However, the low spatials as a group

---

[3] We will omit the 32 practice trials at the beginning of each from consideration here. There is very little data on practice trials, and interruptions during practice often occurred, so that the data are probably not representative.

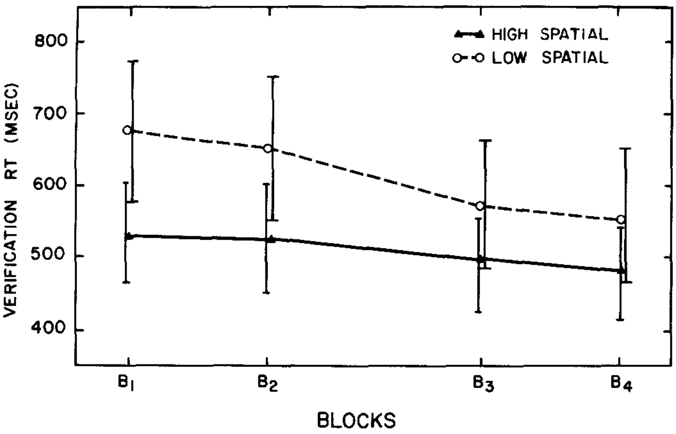

FIG. 4. Mean spatial verification RTs for the high and low spatial groups as a function of blocks (64 trials per block). Also included are the 95% confidence intervals.

showed significant improvement across blocks as they started to acquire the strategy. Analysis of the course of strategy acquisition at the level of individuals yielded the same result. A slope value for each subject was computed across the four blocks. Comparison of the two groups' slope values with a Mann–Whitney *U* statistic, also showed the speed of spatial strategy acquisition to be significantly different for the high and low spatial groups, $Z_U = 2.11, p < .05$.

The variability of a subject's reaction times from trial to trial also yields some information about how the subject is learning a new strategy. If the individual is alternating between strategies, the variability of his reaction times should be greater than if he had acquired the instructed strategy at the outset of the session and had continued to use only that strategy. The variances of both the comprehension and verification RTs were transformed using a logarithmic transformation to reduce any relationship between the means and variances (Scheffé, 1959). *t* tests were computed on the transformed variances in each of the instructional conditions, comparing the high and low spatial subjects. The only significant difference occurred in the Pictorial instruction condition. Verification RTs for low spatial subjects contained significantly more trial-to-trial variability than for high spatial subjects when performing the pictorial strategy, $t(15) = 2.98, p < .01$.

Summary. The linguistic strategy was easily acquired by all subjects. This is not surprising, as we intentionally restricted our selection of subjects to students with average verbal aptitude scores. Although almost all subjects could acquire the pictorial strategy, spatial skill was related to the ease in which the pictorial strategy was acquired and the quality of its performance. The subjects with higher spatial skills seemed to have acquired the pictorial strategy very quickly, for their verification RTs did not show any reliable improvement across the blocks of trials. In addition, high spatial subjects' verification RTs showed significantly less variability. With increased practice, the low spatials' verification RTs decreased, but did not reach the level of the high spatials. Moreover, the low spatials performed the strategy poorly, for their verification RTs continued to reflect the linguistic or affirmative/negative dimension of the sentences.

DISCUSSION

These results replicate our previous finding that individuals approach a simple sentence verification task in qualitatively different ways. We have extended our previous work by showing that, at least for college students, shifting from one strategy to another is rather easy. Although a few subjects initially had difficulty adopting a "pictorial strategy" on request, after a practice period they were able to do so. During our pilot work for this experiment we encountered two subjects who seemed unable, at first, to execute a pictorial strategy. After 2 days of training, even they were able to execute it.

We found no individuals who had trouble utilizing a verbal strategy. This is hardly surprising, as we intentionally restricted our sample to exclude persons of low verbal aptitude. Even if we had not done so, we do not think we would have found very many college students who have trouble using a verbal strategy, simply because of the high level of verbal aptitude required to function in an academic atmosphere. We shall return to this point shortly.

Our results suggest caution in interpreting experiments in which results averaged over individuals are offered as support for a general model of language use. The point that average data should always represent a fair picture of individual subjects is hardly a new one. A slightly more subtle point is that strategies of verbal comprehension may be chosen in response to what are called the "situational demands" or the "demand characteristics" (Orne, 1962). In our experiment these demands were explicit; we told our subjects what we wanted them to do, and they did it. We suspect that situational demands have been less blatant, but equally present, in other studies of language use. If the research participant is an "advanced undergraduate or graduate student at University X" and if it is well known that the experimenter is working on verbal reasoning, is it not possible that the participant will respond to an implicit demand to adopt a verbal strategy?

While our results have focused on the sentence verification tasks, we believe that there are other situations in which individual differences in choice of strategy will markedly affect the validity of models that fit the average data well. Glushko and Cooper (1979) have observed individual differences in the use of strategies in a figure comparison task. Sternberg (Note 3) has reported data similar to ours using a syllogistic reasoning task, and Smith (Note 4) has observed that strategy choice may affect the validity of models of linear ordering problems. Eley (Note 5) has, independently of our work, observed results similar to ours using a slightly different version of the sentence verification task.

Our results should not be regarded as arguments against the development and testing of information processing models of linguistic reasoning. Models of both linguistic and nonlinguistic information processing are valuable descriptive tools, provided that their scope is specified. Hunt (1978) has argued that in order to understand the intellectual performance of an individual, one must understand a person's primitive capacities for information processing, a person's knowledge (on which the information processing operations act), and the strategy by which a person orders particular information processing operations in order to solve the problem at hand. Our results should be seen in this light. We have demonstrated considerable flexibility in information processing in a tightly controlled task. We suspect that in more natural, less controlled situations people have even more flexibility in the way in which they can use language in thinking.

The factors that lead people to choose one strategy or another are of interest in their own right. In this and in our previously published work (as well as in some unpublished studies), we have found evidence that the linguistic strategy of sentence verification is dominant (though by no means universal) in a population of college students. This may be a function of the particular use that this population makes of the written language. In university studies, language is used as a vehicle for learning and a great deal of reading is done on topics that are not easily visualized. Sticht (1977), who studied the type of reading required in military and industrial settings, observed that much adult reading is oriented toward immediately perceivable, practical problems. Examples are traffic signs and instructions for assembling and operating machinery. It would be of considerable interest to know what situational characteristics call forth which types of strategy for coordinating linguistic and physical reality.

Just as situations differ in their demands, people differ in their capacities. What calls forth a particular strategy in individual A may not call forth the same strategy in individual B. Presumably the tendency to adopt particular strategies is related to other individual characteristics. To offer a plausible hypothesis, sex and age differences in spatial reasoning capacity have frequently been observed (Willerman, 1979; DeFries, Johnson, Kuse, McClearn, Polovina, Vandenberg, & Wilson, 1979). Are there parallel sex and age differences in the manner in which individuals deal with linguistic descriptions? What are the practical consequences of such differences for those who wish to transmit information to a particular population? Does the ability to shift strategies change over individuals in a predictable way? What are the implications of this for education and communication? The factors that control our notoriously flexible thought ought not to be overlooked in a search either for general laws or for typologies.

REFERENCES

CARPENTER, P. A., & JUST, M. A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review,* 1975, **82,** 45–73.

CLARK, H. H., & CHASE, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology,* 1972, **3,** 472–517.

DEFRIES, J., JOHNSON, R., KUSE, A., MCCLEARN, G., POLOVINA, U., VANDENBERG, S., & WILSON, J. Familial resemblance for specific cognitive abilities. *Behavior Genetics,* 1979, **9,** 23–43.

GLUSHKO, R. S., & COOPER, L. A. Spatial comprehension and comparison processes in verification tasks. *Cognitive Psychology,* 1978, **10,** 391–342.

HARTIGAN, J. *Clustering algorithms.* New York: Wiley, 1975.

HUNT, E. The mechanics of verbal ability. *Psychological Review,* 1978, **85,** 109–130.

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, **17,** 493–507.

NELSON, M. J., & DENNY, E. C. *The Nelson–Denny reading test.* Boston: Houghton–Mifflin, 1960.

ORNE, M. T. On the social psychology of the psychology experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17,** 776–783.

SCHEFFÉ, T. *The analysis of variance.* New York: Wiley, 1959.

STICHT, T. Comprehending reading at work. In M. A. Just & P. A. Carpenter (eds.), *Cognitive processes in comprehension.* Hillsdale, N.J.: Erlbaum, 1977.

WILLERMAN, L. *The psychology of individual and group differences.* San Francisco: W. H. Willerman, 1979.

REFERENCE NOTES

1. University of Washington Assessment Center. *Technical Manual: Washington Pre-College Test Battery.* Seattle, Wash. 1977.
2. SEGAL, D., & RASKIN, E. *Multiple aptitude tests.* Los Angeles, Calif.: California Test Bureau, 1959.
3. STERNBERG, R. Personal communication, 1979.
4. SMITH, K. Personal communication, 1979.
5. ELEY, M. C. *Image and propositional representation processes in sentence–picture verification tasks.* Manuscript submitted for publication, 1979.

(Received May 21, 1979)