

## SEMANTICS IN THE PERCEPTION OF VERTICALITY

BY WILLIAM G. CHASE AND HERBERT H. CLARK

*Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pa., U.S.A., and  
Department of Psychology, Stanford University, Stanford, Calif., U.S.A.*

Seymour has shown recently that people take less time to judge that the word *above* correctly describes the spatial position of a small circle drawn ABOVE a large reference square than they do for the word *below* and the circle BELOW the square. Seymour has attributed this asymmetry to the tendency for people to invariably scan a picture from top to bottom. In the present study, the first experiment confirms Seymour's results, but the next three demonstrate that the asymmetry Seymour found cannot be accounted for by an attentional-scanning process. Instead, it is proposed that people interpret the words *above* and *below* as abstract symbols at a first stage of processing, interpret the pictures ABOVE and BELOW as abstract symbols at an independent second stage, and compare these two sets of symbols at an independent third stage. In support of this model, the results show, for example, that *above* is interpreted about 80 msec. faster than *below* at the first stage quite independently of what happens at the second and third stages. The asymmetry Seymour found is therefore attributable to the difference in the interpretation times of *above* and *below* at the first stage of processing.

Seymour (1969) has recently reported a remarkable asymmetry in people's judgements of spatial location. He presented subjects with either the printed word *above* or the printed word *below* simultaneously with a configuration of a small circle either ABOVE or BELOW a large centrally located reference square. Seymour measured how long it took the subjects to judge whether or not the word correctly described the spatial location of the circle. The important result was that subjects required much less time to verify that *above* (the word) matched ABOVE (the circle's location) than that *below* matched BELOW. This finding implies that subjects do not use a simple search-and-confirm strategy of decoding the word and then looking at the critical part of the display to see if the circle is there.

Seymour attributed these results to the way the subjects scanned the display. The theory he proposed assumed that subjects first generate a model (some internal representation) of the display from the word *above* or *below* and then scan the model and display simultaneously from the top down until they locate the circle in both the model and the display. The critical part of the theory is the top-down scanning hypothesis. The assumption is that subjects invariably scan from top to bottom and, as a consequence, they are able to verify the critical upper part of the ABOVE display more quickly than the critical lower part of the BELOW display.

The purpose of the present study is to consider an alternative to the scanning hypothesis, one that is based on the abstract interpretations of the word and display. The alternative is meant to correct two weaknesses in Seymour's account. First, it seems unlikely that scanning is invariably top-down; rather, the direction of scanning will depend on specific requirements of the task. Second, Seymour's comparison operation seems too concrete for this task. His notion that subjects generate a concrete internal model from the more abstract meanings of *above* and *below* is much like Posner's (1969) concept of generation, the 'subjects' ability to go from a general code to one of greater specificity'. According to Posner, for example, subjects

can generate a visual image from the name of a letter and then use the image to compare against an actual displayed letter. But in a task like Seymour's it seems unlikely that subjects rely on this general-to-specific generation of a model, since they should be able to carry out the task even when it is impossible to scan. For example, when the top half of the display is always concealed, there is still enough information visible to decide whether *above* or *below* matches ABOVE or BELOW. But in this example, it is the interpretation of the display, not the display itself, that is required to carry out the task. That is, we suggest that subjects are more likely to generate an abstract representation out of the more complex information-rich display.

In the alternative we consider, the subjects first interpret, or represent mentally, both word and configuration as abstract symbols. The point of this process is that the subjects need to represent both the word and the display in the same format so that they can then test to see whether the two representations are the same or different. The critical feature of this account is that the mental representation of the word *above* is simpler than the word *below*, so subjects are able to encode *above* faster than *below*. The relative ease of encoding the ABOVE and BELOW displays depends critically on the visual attributes of each display. Furthermore, the encoding of the word and the encoding of the picture are independent of each other, and both are independent of the comparison process that follows. Evidence for this account will be considered more fully in the discussion that follows Expt. IV.

The present study consists of four experiments. The first is a replication of Seymour's results on the apparatus in our laboratory. The second is a repeat of the first with arrows in place of the words *above* and *below*; the aim is to demonstrate that a top-down scanning strategy is not obligatory. The third experiment, very similar to the first two, utilizes the words *present* and *absent*; the purpose here is to show a similar asymmetry between *present*-PRESENT and *absent*-ABSENT that cannot be due to a top-down scanning strategy. The last experiment again uses *above* and *below* but with displays that cannot be scanned from the top down; the purpose is to show that the same asymmetry occurs without scanning and so the *semantic* interpretations of *above* and *below* are crucial to Seymour's results. A discussion of possible explanations of these data follows the last experiment.

## EXPERIMENT I

### *Method*

The four types of displays prepared for Expt. I each consisted of a square with a smaller circle directly above or below the square; the figures were drawn with India ink and the circle was filled in to make a solid disk. At the centre of the square there was either the word *above* or the word *below* typed in Elite type. The displays, viewed through a Polymetric two-field tachistoscope at a distance of 18 in., subtended a visual angle of 4°. The four display types will be referred to as *above*-ABOVE, *below*-BELOW, *above*-BELOW, and *below*-ABOVE, depending on whether the word was *above* or *below* and whether the circle was ABOVE or BELOW the square. The first two displays - *above*-ABOVE and *below*-BELOW - will be called the 'true' conditions, since the word is a true description of the configuration; the last two displays will be called the 'false' conditions. The four displays are illustrated at the top of Table 4.

A deck of 28 stimulus cards was drawn up such that each of the four displays occurred seven times; one set of four displays was used for lead-in trials before a block and six sets were used to make the block of 24 trials. Each of 12 subjects was given four blocks of trials for a total of

*Semantics in the perception of verticality*

96 trials plus 16 lead-in trials; the stimulus cards were in a different random order for each subject and each block of trials.

On each trial the subject was presented one of the four displays and required to push either a 'true' or a 'false' button as quickly as possible, and his latency was measured to the nearest 0.01 sec. Half the subjects pushed 'true' with their right thumb and half pushed 'true' with their left thumb. The subject was instructed to respond as quickly as possible without error. After each trial the subject was informed of his latency if he was correct, and only that an error had occurred if he was incorrect.

To initiate the next trial, the subject pushed a 'start' button (situated midway between the 'true' and 'false' buttons) and 1 sec. later - time enough for the subject to place his thumbs in the proper position - the next display appeared. The subjects were 12 Carnegie-Mellon University students fulfilling a course requirement for introductory psychology.

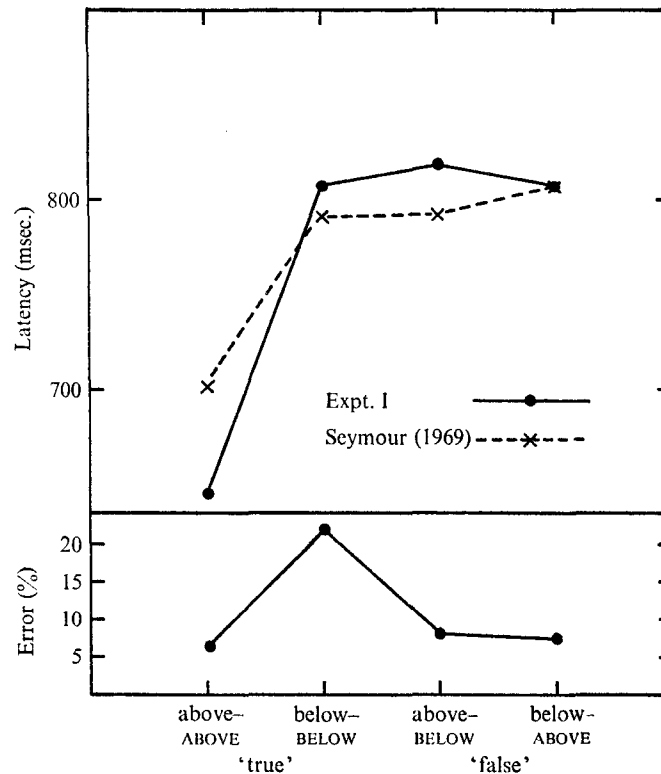


Fig. 1. Mean latencies from Expt. I and Seymour's (1969) results as a function of the words *above* and *below* and the displays ABOVE and BELOW. The means are averaged over 12 subjects, four blocks of trials and six or fewer correct responses per block.

*Results*

The overall mean latencies of the four conditions, shown in Fig. 1, replicated quite nicely the pattern of latencies found by Seymour, also shown in Fig. 1. The basic latency analysis was carried out on the mean latencies of the six or fewer correct responses for each condition for each block of each subject; thus there were 16 mean latencies for each of 12 subjects, or 192 means. An analysis of variance showed the following significant effects: (1) displays with *above* were verified 75 msec. faster than those with *below* ( $F = 11.2$ ; d.f. = 1, 11;  $MS_e = 23,900$ ;  $P < 0.01$ );

(2) displays with the circle ABOVE were verified 85 msec. faster than those with the circle BELOW ( $F = 17.8$ ; d.f. = 1, 11;  $MS_e = 19,700$ ;  $P < 0.005$ ); (3) the 'true' displays were verified 85 msec. faster than the 'false' displays, i.e. there was a significant interaction between the *above-below* variable and the ABOVE-BELOW variable ( $F = 15.2$ ; d.f. = 1, 11;  $MS_e = 23,200$ ;  $P < 0.005$ ); (4) subjects became faster as the experiment progressed ( $F = 9.88$ ; d.f. = 3, 33;  $MS_e = 10,600$ ;  $P < 0.001$ ). No interactions between blocks and the four display conditions approached significance.

Unlike Seymour's data, the present results showed more errors in one condition than in the others. An analysis of variance showed significantly more errors on the *below-BELOW* condition than on the other three ( $F = 25.2$ ; d.f. = 1, 33;  $MS_e = 73.5$ ;  $P < 0.001$ ). This contrast accounted for 99 per cent of the variance between conditions and the other three conditions were not significantly different from each other.

#### EXPERIMENT II

The previous result, then, confirms Seymour's important finding that *above-ABOVE* takes considerably less time to verify than *below-BELOW*. If it is true, as Seymour suggests, that this result is due to an invariant top-down scanning process, then the same result should occur when *above* is replaced by the UP arrow ↑ and *below* by the DOWN arrow ↓, since the UP and DOWN arrows 'mean' the same thing in this task as *above* and *below*.

Expt. II was identical in all respects to Expt. I, except that small UP and DOWN arrows were used in place of the words *above* and *below*. The four conditions will be referred to as UP-ABOVE, DOWN-BELOW, UP-BELOW and DOWN-ABOVE, replacing *above-ABOVE*, *below-BELOW*, *above-BELOW*, and *below-ABOVE* of Expt. I, respectively. The subjects were 12 more students from the same source as those in Expt. I.

Table 1. *Mean latencies and mean percentages of errors of Expt. II*

| Condition  | Response | Latency (msec.) | Percentage errors |
|------------|----------|-----------------|-------------------|
| UP-ABOVE   | 'true'   | 597             | 4.8               |
| DOWN-BELOW | 'true'   | 591             | 5.2               |
| UP-BELOW   | 'false'  | 615             | 7.1               |
| DOWN-ABOVE | 'false'  | 628             | 4.8               |

#### Results

The pattern of latencies in Expt. II was clearly symmetrical, in contrast to the previous results. The mean latencies and error percentages from Expt. II, calculated in analyses identical to those of Expt. I, are shown in Table 1. The analysis of variance produced the following results: (1) the UP displays were verified a small non-significant 4 msec. faster than the DOWN displays; (2) the ABOVE displays were verified a small non-significant 9 msec. slower than the BELOW displays; (3) the 'true' displays were verified a marginally significant 28 msec. more quickly than the 'false' displays ( $F = 4.81$ ; d.f. = 1, 11;  $MS_e = 7620$ ;  $P \approx 0.05$ ); (4) there was a slight (30 msec.) non-significant decrease in latencies as the experiment progressed. These results are obviously quite different from those in Expt. I. This is especially true of results 1 and 2, which show negligibly small differences (4 and -9 msec.)

*Semantics in the perception of verticality*

315

in this experiment in contrast to the large significant differences (75 and 85 msec.) in Expt. I. The prediction from the scanning hypothesis that UP and DOWN arrows should produce the same effects as *above* and *below* is therefore fully disconfirmed.

The percentages of errors in Expt. II also patterned quite differently from those in Expt. I. Unlike Expt. I errors, there was no significant difference among the four conditions.

*Discussion*

Just how did the Expt. II subjects go about verifying the displays? From the data it appears that they followed the simple strategy of (1) deciding which direction the arrow was pointing and (2) looking at the region indicated by that arrow to see if the circle was there or not. If the circle was there, they pushed the 'true' button; if not, they pushed the 'false' button. In scanning the display they apparently did not scan from the top of the display but rather looked immediately at the critical area indicated by the arrow. This simple search-and-confirm operation is just the kind that Seymour said was impossible. Since subjects did not find it necessary to scan from top to bottom in Expt. II, it is highly unlikely that they did so in Expt. I either. The explanation of the results of Expt. I must lie elsewhere, namely in the semantic interpretation of the words *above* and *below*.

EXPERIMENT III

This experiment was designed to show that the asymmetrical results of Expt. I are not unique to judgements of verticality. It can be argued that the mental operations for coding *present-absent* are similar in certain respects to those for *above-below*, and that a similar asymmetry will occur when judgements of presence or absence of the circle are made. In particular, it is predicted that the judgement *present-PRESENT* will be simpler and take less time than *absent-ABSENT*. The predicted result cannot be attributed to some top-down scanning process since information about presence or absence can occur in only one spatial location.

*Method*

In Expt. III, subjects were required to decide whether the word *present* or *absent* was 'true' or 'false' with respect to a circle under the four conditions, *present-PRESENT*, *absent-ABSENT*, *present-ABSENT*, and *absent-PRESENT*. In displays like those in Expt. I the word *present* or *absent* was typed within the reference square, and the circle was PRESENT or ABSENT above the square on two blocks of trials and below the square on the other two blocks. These two conditions are referred to as top-visible and bottom-visible, respectively. It should be made clear that when the circle is ABSENT the stimuli in the top-visible and bottom-visible conditions are identical - although, of course, their effect could be different by virtue of the other stimuli in these conditions.

Expt. III was identical in other respects to Expt. I, e.g. in its use of practice and experimental trials, blocks and number of trials, and so on. But to balance the effects of viewing conditions, subjects were divided into two groups. One group received the viewing conditions in the following order on the four blocks of trials: top-visible, bottom-visible, bottom-visible, top-visible. The other group received the complement of this ordering. Within each group, half the subjects used the right button for 'true' and half used the left button. The subjects were again 12 students drawn from the same source as those in Expt. I.

*Results*

According to the overall mean latencies and error percentages in Table 2, the two main results of Expt. III are (1) that *present*-PRESENT and *absent*-ABSENT are asymmetrical in much the same way that *above*-ABOVE and *below*-BELOW are asymmetrical, and (2) that it makes little difference whether the circle is expected above or below the square. As for the second finding, an analysis of variance of the 192 mean latencies showed that the 20 msec. difference between the top- and bottom-visible conditions was negligible ( $F < 1$ ). The analysis showed only four significant effects: (1) the *present* displays were verified 145 msec. faster than the *absent* displays ( $F = 14.4$ ; d.f. = 1, 11;  $MS_e = 69,900$ ;  $P < 0.005$ ); (2) the displays with the circle PRESENT were verified 74 msec. faster than those with the circle ABSENT ( $F = 12.1$ ; d.f. = 1, 11;  $MS_e = 21,500$ ;  $P < 0.01$ ); (3) 'true' displays were verified 53 msec. faster than 'false' displays ( $F = 9.28$ ; d.f. = 1, 11;  $MS_e = 14,500$ ;  $P < 0.025$ ); and (4) latencies again became shorter as the experimental session progressed ( $F = 45.2$ ; d.f. = 1, 11;  $MS_e = 31,600$ ;  $P < 0.001$ ). The error percentages showed much the same pattern as the mean latencies.

Table 2. *Mean latencies and mean percentages of errors for Expt. III*

| Condition         | Response | Condition   |                | Mean |
|-------------------|----------|-------------|----------------|------|
|                   |          | Top-visible | Bottom-visible |      |
| Latencies (msec.) |          |             |                |      |
| present-PRESENT   | 'true'   | 738         | 755            | 746  |
| absent-ABSENT     | 'true'   | 978         | 952            | 965  |
| present-ABSENT    | 'false'  | 846         | 900            | 873  |
| absent-PRESENT    | 'false'  | 927         | 961            | 944  |
| Average           |          | 872         | 892            | 882  |
| Error percentages |          |             |                |      |
| present-PRESENT   | 'true'   | 2.1         | 4.2            | 3.1  |
| absent-ABSENT     | 'true'   | 16.0        | 15.3           | 15.6 |
| present-ABSENT    | 'false'  | 4.9         | 8.3            | 6.6  |
| absent-PRESENT    | 'false'  | 4.9         | 6.3            | 5.6  |
| Average           |          | 7.0         | 8.5            | 7.7  |

*Discussion*

Expt. III, then, demonstrates that *present* and *absent* are asymmetrical just as *above* and *below* were, in spite of the fact that *present* and *absent* require no scanning. The information about the presence or absence of the circle was available at only one spatial location on any one block of trials. Furthermore, since the latencies were virtually identical for the top- and bottom-visible conditions, the conclusion is strengthened that subjects do not invariably scan displays from top down. The *present-absent* results, then, must be accounted for by underlying mental representations, not by a scanning mechanism, and evidence is mounting that the same is true for *above-below*.

EXPERIMENT IV

One might argue that Expt. II was a 'purely' perceptual task, containing, as it did, only arrows, squares and circles, and that because no linguistic interpretation of the words *above* and *below* was needed, the task somehow allowed subjects to bypass the requirement that they scan configurations from top to bottom. Posner & Mitchell (1967), for example, were able to identify different levels of processing involved in match-mismatch judgements for letters of the alphabet; the comparison could be carried out on the physical forms of the letters, on the letter names, or on some classification rule (consonant-vowel), depending upon the depth of processing required. One might argue, then, that the words *above* and *below* themselves constrain the subjects to scan configurations as Seymour proposed. This is not unreasonable since *above* and *below* both refer to verticality and hence they might force subjects, in some unknown way, to scan from the top down. Note, however, that this very restricted hypothesis is not the one proposed by Seymour, since he appears to posit the scanning strategy no matter what the task. Expt. IV was designed to eliminate even this last thread of hope for the scanning hypothesis.

The strategy of Expt. IV - similar to that of Expt. III - was to give subjects displays which they could not verify by scanning; instead they were always forced to inspect only one part of the display and make inferences about another part. If the *above*-ABOVE *below*-BELOW asymmetry still persists with these displays, the scanning hypothesis can be ruled out.

*Method*

Expt. I was repeated, but the spatial location either above or below the square was masked. Thus when the top-visible mask was used, the subjects could see the circle when it was above the square, but nothing when it was below the square; in the latter case, they had to infer that the circle was below the square by noting that the circle was not above the square. The analogous situation held, of course, with the bottom-visible mask. In Expt. IV, then, top-down and search-and-confirm strategies were of no use since within any block of trials only one part of the display contained information.

Expt. IV was identical in all respects to Expt. III, except that Expt. IV used the same four displays as Expt. I. Again top-visible and bottom-visible conditions were given to each subject in a counterbalanced ABBA design. The subjects were 12 students drawn from the same source as those in Expts. I, II and III.

*Results*

The important finding of Expt. IV was that the pattern of results for the top-visible condition closely paralleled that of Expt. I, but the pattern was quite different for the bottom-visible condition. The overall mean latencies for the eight conditions of Expt. IV are shown in Fig. 2 along with the mean latencies for Expt. I, labelled 'both-visible'. The comparison of Expts. I and IV can be seen more clearly in Fig. 3, which shows differences between the both-visible condition (Expt. I) and the top-visible and bottom-visible conditions. A preliminary analysis showed no differences between the top-visible and both-visible conditions; the 94 msec. overall difference between groups was not significant. A comparison between bottom-visible and both-visible conditions, however, showed a significant overall difference of 158 msec. between the groups ( $F = 9.08$ ; d.f. = 1, 22;  $MS_e = 65,530$ ;  $P < 0.01$ ).

More importantly, there was a strong interaction of viewing conditions with the ABOVE-BELOW variable ( $F = 17.1$ ; d.f. = 1, 22;  $MS_e = 6251$ ;  $P < 0.001$ ). This result can be more easily seen in Fig. 3, where two conditions are seen to be considerably slower than the other six. These are the two conditions in which the subject had to infer that the circle was ABOVE. Whereas this inference is apparently quite difficult, the inference that the circle is BELOW requires no more time in the top-visible condition than in the both-visible condition.

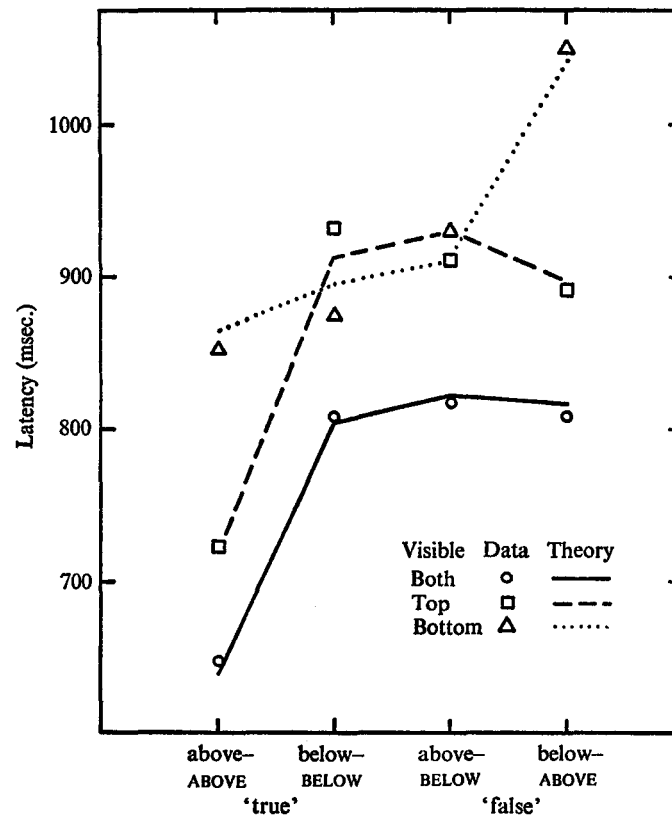


Fig. 2. Mean latencies for Expt. I, labelled 'both-visible', and the top- and bottom-visible conditions of Expt. IV. Also shown are the predicted latencies derived from the model of Clark & Chase (in preparation).

The principal difference among the both-, top- and bottom-visible conditions was to be found in the differences between the ABOVE and BELOW displays. The differences between *above* and *below* displays, and between 'true' and 'false' displays, remained unaffected by the visibility conditions. The analysis of variance for Expt. IV showed the following significant results. (1) The displays with *above* were verified 83 msec. faster, on the average, than those with *below* ( $F = 8.56$ ; d.f. = 1, 11;  $MS_e = 39,500$ ;  $P < 0.025$ ); this difference is close to the 75 msec. found in the both-visible condition of Expt. I. Furthermore, the interaction between the *above*-*below* factor and the top-*v.* bottom-visible factor did not approach significance. (2) The 'true' displays were verified 100 msec. faster, on the average, than the 'false' displays ( $F = 19.6$ ;

*Semantics in the perception of verticality*

d.f. = 1, 11;  $MS_e = 25,000$ ;  $P < 0.001$ ); this difference is close to the 85 msec. found in the both-visible condition. Again, there was no interaction of 'true'-'false' with the visibility conditions. (3) There was a strong interaction between ABOVE-BELOW and the viewing conditions such that the ABOVE displays were verified 115 msec. faster in the top-visible condition, but 49 msec. slower in the bottom-visible condition ( $F = 23.8$ ; d.f. = 1, 11;  $MS_e = 13,400$ ;  $P < 0.001$ ). The ABOVE displays were verified 86 msec. faster than the BELOW displays in the both-visible condition. (4) The overall mean for the top-visible display was 63 msec. faster than

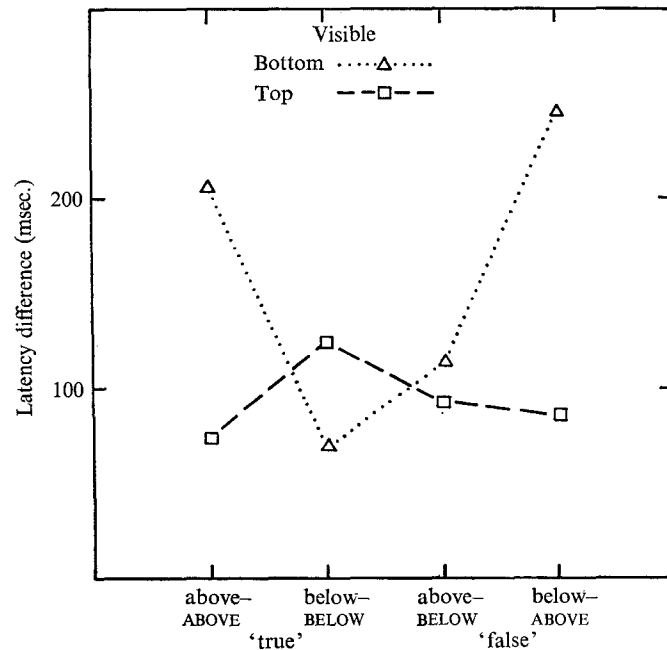


Fig. 3. The difference between the mean latencies of top- and bottom-visible conditions and that of the both-visible condition.

Table 3. Mean percentages of errors in Expt. IV

| Condition   | Response | Condition   |                | Average |
|-------------|----------|-------------|----------------|---------|
|             |          | Top-visible | Bottom-visible |         |
| above-ABOVE | 'true'   | 13.2        | 7.6            | 10.4    |
| below-BELOW | 'true'   | 16.0        | 13.2           | 14.6    |
| above-BELOW | 'false'  | 2.1         | 5.6            | 3.8     |
| below-ABOVE | 'false'  | 4.9         | 11.8           | 8.4     |
| Means       |          | 9.0         | 9.6            | 9.3     |

the mean of the bottom-visible display ( $F = 7.45$ ; d.f. = 1, 11;  $MS_e = 25,900$ ;  $P < 0.025$ ). (5) Finally, there was again an overall decrease in latencies with practice ( $F = 155$ ; d.f. = 1, 11;  $MS_e = 53,900$ ;  $P < 0.001$ ). The error percentages were more variable than the mean latencies, but they again showed the highest error rates in the below-BELOW condition (Table 3).

*Discussion*

The significance of Expt. IV is (1) that it suggests that abstract mental representations are used in making spatial judgements and (2) that it eliminates all hope for the scanning hypothesis.

Probably the most important finding was that the *above* displays were about 80 msec. faster than the *below* displays no matter whether the whole display or just the top or bottom was visible. Similarly, the 'true' displays were verified about 90 msec. faster than the 'false' displays regardless of the three visibility conditions. The only latencies affected by the three visibility conditions were the differences between the ABOVE and BELOW displays. This finding suggests that the mental processes that result in the *above-below* and 'true'-'false' differences can be separated out from those that result in the ABOVE-BELOW difference. More will be said about this later.

A subsidiary finding was that the pattern of latencies in the top-visible condition was approximately the same as in the both-visible condition, but the pattern in the bottom-visible condition was considerably different from the other two. This finding suggests that subjects employ approximately the same processes in the both- and top-visible conditions, but quite different processes in the bottom-visible condition. Put more specifically, the process of seeing that a circle is BELOW and the process of inferring that a circle is BELOW must be quite similar ones, but the process of seeing that a circle is ABOVE and the process of inferring that a circle is ABOVE must be quite different from each other. More will be said later about this too.

As for the scanning hypothesis, the results of Expt. IV finally rule out the hypothesis completely. According to the subsidiary finding, the both- and top-visible conditions resulted in the same patterns of latencies in spite of the fact that a scanning strategy was impossible in the top-visible condition. Therefore, since the *above-ABOVE below-BELOW* asymmetry can occur in the absence of scanning, the asymmetry cannot be accounted for by the scanning hypothesis.

The results also rule out the possibility of explaining the *above-ABOVE below-BELOW* asymmetry solely by an attention hypothesis. According to this view, people normally attend to the top of the picture, and the BELOW displays are processed by inference. Thus, *above-ABOVE* is faster than *below-BELOW* because, in the *below-BELOW* case, people have to make a time-consuming inference that the circle is BELOW; as a consequence, masking the bottom of the displays should have no effect since people infer that the circle is BELOW anyway. This explanation fails, however, to account for the bottom-visible latencies. Under this assumption, the bottom-visible latencies should show the same pattern as the top-visible latencies, once *above* and *below*, and ABOVE and BELOW, have been interchanged, because now people attend directly to the BELOW displays but make the time-consuming inference for the ABOVE displays. But the pattern in the results is nothing like this expected pattern. To take an example, in the bottom-visible condition, *below-BELOW* should be faster than *above-ABOVE*, since *below-BELOW* can be verified directly, whereas *above-ABOVE* requires an inference. Furthermore, this *below-BELOW above-ABOVE* difference should be the same size as the *above-ABOVE below-BELOW* difference in the top-visible condition. But in fact, in the bottom-visible condition, *above-ABOVE* was

still faster than *below*-BELOW, completely contrary to this expectation. Results like these, then, simply cannot be accounted for by an attentional-scanning strategy.

The findings of Expt. IV therefore suggest that another type of theory altogether is required to account for the verification process in Expts. I and IV. We will now present a theory that begins with abstract interpretations of *above*, *below*, ABOVE, and BELOW. It is based on a considerable amount of subsequent work on the *above*-*below* asymmetry; the present results are clearly too meagre to support an elaborate model by themselves. We will describe the parts of the theory that apply to the present experiments; the full theory and its justification can be found in Clark & Chase (in preparation).

#### A THEORY OF VERIFICATION OF SPATIAL LOCATION

The proposed explanation for Expts. I and IV is based on the premise that the time between stimulus and response is composed of a series of additive stages in the sense described by Sternberg (1969). In the first of four gross additive stages, the subject encodes the word *above* or *below*; in the second, he encodes the location of the circle; in the third, he compares his mental representations of the word and location; and in the fourth, he responds. The explanation of the *above*-*below* asymmetry lies completely in the representations that *above* and *below* take in the first stage.

At the first stage, the Word Encoding stage, it is assumed that words or sentences are encoded approximately in terms of their deep structure propositions (cf. Clark, 1969), which are represented as sets of symbols and rules. In this experiment we assumed that *above* and *below* actually stand for the full sentences *The circle is above the square* and *The circle is below the square*, respectively. These two sentences, in turn, contain very simple deep structure propositions that may be symbolized as (*circle above square*) and (*circle below square*), respectively. But these representations are not detailed enough, for *above* and *below* have even more abstract representations. Linguistically, *above* and *below* are closely related. Both *above* and *below* specify vertical comparison, although *above* specifies comparison in one direction and *below* specifies comparison in the opposite direction. Thus, *above* and *below* could be specified in a featural notation in which all features but one are identical for the two, and the one feature that differs specifies direction of comparison. The solution, however, is not that simple, for the word *above* is the normal or simpler way to encode two vertically arranged objects, while *below* is the more complex. To understand this, we must consider two linguistic facts about English: (1) that *A is above B* and *B is below A* are not synonymous, and (2) that verticality is always described in English in an asymmetrical fashion.

The first linguistic fact is that *A is above B* and *B is below A*, although they appear to be synonymous, differ at a critical point. The first sentence describes the position of *A* with respect to the position of *B*, whereas the second one does just the opposite. *A is above B*, for example, presupposes that the position of *B* is already known to the listener, and it asserts that *A* has a particular position with respect to *B*. Thus it is unacceptable to answer the question *Where is B?* with the answer *A is above B* (with normal intonation), since the presupposition of the answer (that the position

of *B* is known) conflicts directly with the presupposition of the question (that the position of *B* is *not* known). Therefore, *A is above B* is used to describe a picture in which *B* is the point of reference, whereas *B is below A* is used where *A* is the point of reference.

The second linguistic fact is that English normally describes verticality such that the point of reference is at the bottom of the described dimension. The only English adjectives used exclusively for describing vertical position are *high*, *low*, *tall* and *short*. It is already well known (cf. Clark, 1969) that *high* and *low* presume the semantically prior dimension of *height* (not *lowness*), and that they mean, roughly, 'of much height' and 'of little height', respectively. But *height* is always measured upward from a point of reference at the bottom, no matter whether much or little height is being measured. The analogous situation holds for *tall* and *short* and their semantically underlying dimension *height* (tallness). So in spite of their other differences, *high*, *low*, *tall* and *short* all presuppose a point of reference at the bottom of what is being measured or described.

The sentence *A is above B* also has the property that the point of reference, *B*, is at the bottom of what is being described; *B is below A* does not have this property. This suggests that the sentence *A is above B* is the normal way to describe a picture of an *A* above a *B*. Since verticality automatically brings with it the assumption that the point of reference is at the bottom, we have hypothesized that the semantic representation of *above* is formed in one step (since it conforms to this assumption), whereas that of *below* requires one additional step specifically to alter this automatic assumption (cf. Clark & Chase, in preparation, for details). In short, the *above* and *below* in (*circle above square*) and (*circle below square*) should be thought of as amalgamations of more primitive features, with the representation of *below* requiring one more step to form than the representation of *above*; likewise, *circle* and *square* represent more detailed semantic specifications of the abstract concepts of 'circle' and 'square'.

These considerations have one important consequence on the Word Encoding stage. Since *above* is represented more simply than *below*, the *above* displays should be represented more quickly than the *below* displays. The amount of time by which *above* is represented faster than *below* is given by the parameter *a* in this model.

At the second stage, the Picture Encoding stage, it is assumed that the ABOVE and BELOW locations of the circle are encoded, respectively, as (*circle above square*) and (*circle below square*). But the three visibility conditions vary in how easy it is for the subject to set up the two representations. Apparently, subjects normally attend to the upper location of the circle, hence they normally (or often) only *infer* (*circle below square*) in the BELOW displays. When either the bottom or top of the display is hidden from view, of course, the subject has to infer the hidden location. In short, (*circle above square*) should be encoded some amount  $b_1$  faster than (*circle below square*) in the both-visible condition, some amount  $b_2$  faster in the top-visible condition, but some amount  $b_3$  slower in the bottom-visible condition.

At the third stage, the Comparison stage, it is assumed that the subject (in these experiments at least) must carry out two time-consuming mental operations. The purpose of these operations is to keep track of an indexical 'truth value' of the word *vis-à-vis* the picture and to produce its correct value in the end as *true* or *false*

*Semantics in the perception of verticality*

(i.e. match or mismatch). The truth value is assumed to be *true* at the beginning, and the finding of a mismatch between the word and picture representations changes that *true* to *false*. Thus Operation 1 compares the representation of the word with that of the picture. If the two representations match completely, then the pre-supposed truth value *true* is retained. If the two representations do not match, Operation 1a changes the truth value from *true* to *false*.

The consequence of Operations 1 and 1a of the Comparison stage is to cause the two 'false' conditions to be slower than the two 'true' conditions. The first time-consuming operation (Operation 1) is required by all four displays – *above*-ABOVE, *below*-BELOW, *above*-BELOW and *below*-ABOVE – so it adds an equal increment of time to each of these conditions. But the second operation (Operation 1a) is required only of the two 'false' displays, *above*-BELOW and *below*-ABOVE, so it adds an increment only to these two displays. This increment of time is given in the model by the parameter *c*. Thus this model of the Comparison stage predicts that the two 'true' displays should be verified a fixed time increment *c* faster than the two 'false' displays.

Table 4. *The four displays and the theoretical comparison process for each*

|                        |                |                |                |                |
|------------------------|----------------|----------------|----------------|----------------|
|                        | ●              |                |                | ●              |
| The displays           | □<br>above     | □<br>below     | □<br>above     | □<br>below     |
|                        |                | ●              | ●              |                |
| Description of display | above-ABOVE    | below-BELOW    | above-BELOW    | below-ABOVE    |
| Stages                 |                |                |                |                |
| Word encoding          | (cir above sq) | (cir below sq) | (cir above sq) | (cir below sq) |
| Picture encoding       | (cir above sq) | (cir below sq) | (cir below sq) | (cir above sq) |
| Comparison             |                |                |                |                |
| Initial truth value    | true           | true           | true           | true           |
| Operation 1            | match          | match          | mismatch       | mismatch       |
| Operation 1a           | —              | —              | true → false   | true → false   |
| Response               | 'true'         | 'true'         | 'false'        | 'false'        |

Finally, it is assumed that the Response stage merely takes the final truth value of the Comparison stage – *true* or *false* – and converts it into a push of the correct 'true' or 'false' button. This process has no differential consequences for the latencies of the four displays. The four stages are shown in abbreviated form for the four displays in Table 4.

The proposed model, then, makes use of five parameters (*a*, *b*<sub>1</sub>, *b*<sub>2</sub>, *b*<sub>3</sub> and *c*) to predict the differences in latencies among the four displays within the three visibility conditions. From the appropriate subtractions in the data at the top of Table 5, *a* is estimated to be 81 msec.; *b*<sub>1</sub>, 86 msec.; *b*<sub>2</sub>, 115 msec.; *b*<sub>3</sub>, 49 msec.; and *c*, 96 msec. These parameters, along with the 'base' times, *t*<sub>1</sub>, *t*<sub>2</sub> and *t*<sub>3</sub>, can be combined appro-

propriately, as shown in the middle of Table 5, to form the predicted times at the bottom of Table 5. The predicted latencies are very close to the actual latencies, as can be seen by referring back to the two sets of times in Fig. 2. To evaluate the theory, the 12 d.f. associated with the data of Table 5 are partitioned into the 3 d.f. for estimating the 'base' times and the 9 d.f. within viewing conditions. Of the 9 d.f. within viewing conditions, the model uses 5 d.f. to estimate its parameters, and 4 d.f. are left for residual variance or lack of fit. The model accounts for over 98 per cent of the variance among the 9 d.f. within conditions; the residual variance does not approach significance when tested against the pooled error term. The lack of fit is best evaluated by calculating the root mean squared deviation (RMSD) and then comparing this RMSD with the smallest parameter estimate of the model (Sternberg, 1969). In this case, the RMSD of 13 msec. with 4 d.f. indicates a good fit. Thus the present model is accurate in predicting the verification latencies in Expts. I and IV.

Table 5. *Actual latencies, theoretical model and predicted latencies for Expts. I and IV*

| Display             | Visibility condition |                 |                     |
|---------------------|----------------------|-----------------|---------------------|
|                     | Both-visible         | Top-visible     | Bottom-visible      |
| Actual latencies    |                      |                 |                     |
| above-ABOVE         | 647                  | 722             | 852                 |
| below-BELOW         | 807                  | 932             | 875                 |
| above-BELOW         | 818                  | 911             | 931                 |
| below-ABOVE         | 807                  | 892             | 1052                |
| Theoretical model   |                      |                 |                     |
| above-ABOVE         | $t_1$                | $t_2$           | $t_3 + b_3$         |
| below-BELOW         | $t_1 + a + b_1$      | $t_2 + a + b_2$ | $t_3 + a$           |
| above-BELOW         | $t_1 + b_1 + c$      | $t_2 + b_2 + c$ | $t_3 + c$           |
| below-ABOVE         | $t_1 + a + c$        | $t_2 + a + c$   | $t_3 + a + b_3 + c$ |
| Predicted latencies |                      |                 |                     |
| above-ABOVE         | 639                  | 718             | 864                 |
| below-BELOW         | 805                  | 913             | 896                 |
| above-BELOW         | 821                  | 929             | 911                 |
| below-ABOVE         | 816                  | 895             | 1041                |

It would have been more elegant if  $b_1$ ,  $b_2$  and  $b_3$  had been equal, indicating that the subjects were simply inferring BELOW in the both- and top-visible conditions and ABOVE in the bottom-visible condition and that this inference took the same amount of time in each instance. Under the assumption that the  $b$ 's are equal, in fact, the model ( $b = 84$  msec.) still accounts for 95 per cent of the variance with only 3 d.f., and the RMSD of 23 msec. with 6 d.f. might be considered a marginal fit. However, there are two points which deviate (significantly) by about 50 msec. from the model, and it can only be concluded that the problem is due to different Picture Encoding times ( $b$ ) for the different viewing conditions. The reason the  $b$ 's vary is certainly not clear, although there are many picture encoding strategies that might account for the variation. In the subsequent experiments in which more care was taken to counterbalance for certain kinds of picture encoding strategies, the  $b$  effect disappeared altogether (Clark & Chase, in preparation). So the  $b$  effect appears to be quite specific to the several viewing conditions of this experiment, whereas the  $a$  and  $c$  effects are quite independent of the viewing conditions.

The estimates for  $a$ ,  $b_1$  and  $c$  for Seymour's data are 53, 38 and 53, respectively. These estimates are considerably lower than the comparable estimates from Expt. I alone of 75, 86 and 86, respectively. It is not clear why these two sets of parameters should be so different.

The most important thing this model has accomplished has been to separate the two encoding difficulties from each other and from the verification difficulties. First of all, this model asserts – in agreement with the results of Expts. I and IV – that *above* takes 81 msec. less time to interpret than *below*, no matter what the display is and no matter what comparison has to be made. This assumption is completely in keeping with all the subsequent data of Clark & Chase (in preparation), which show (1) that this *above–below* difference is consistently about 80 to 100 msec. and (2) that this difference is independent of the display to be verified as well as other attributes of the task. Secondly, the model also asserts that the 'true'–'false' difference is also constant at 96 msec. and is independent of specific attributes of the displays. Again, this is quite consistent with all the results of Clark & Chase (in preparation). But most importantly, the model quite appropriately asserts that the three visibility conditions affect only one thing – the relative difficulty of encoding the ABOVE and BELOW positions of the circle. In other words, the interpretations of *above* and *below* are not affected by variations in the viewing conditions, nor is the verification process itself affected: the only part affected by covering up various parts of the display is the process of representing the location of the circle in memory, just as one would expect. In short, the present model has several advantages over Seymour's attentional-scanning hypothesis. First, it correctly accounts for the data in Expts. I and IV. Secondly, the model explicitly separates out the subprocesses of the task and correctly correlates changes in the experiment with the appropriate changes in the model. And thirdly, the present model has wide generality, for in its more general form it is able to account for many other locational and non-locational tasks for which the attentional-scanning hypothesis makes either incorrect predictions or no predictions at all.

At this point, one might still suspect that the present results can somehow be accommodated by a mental imagery model (e.g. DeSoto *et al.*, 1965; Huttenlocher, 1968). However, such a model fails to explain our later results (Clark & Chase, in preparation), and on quite independent grounds. The problem is with negative sentences such as *A isn't above B*.

First, it is impossible in general to form a single image, or even an image surrogate as in Seymour's proposal, of negative sentences (e.g. *A isn't above B*). The general property of such sentences is that they indicate what a picture is not, and only rarely does this uniquely define what a picture is. But since the imagery model requires that a single image be matched against the picture, the model cannot work in general for negative sentences. And since it cannot account for negatives, it is unlikely that it accounts for positive sentences either. Rather, we suppose that both positive and negative sentences are verified in the same way, and this assumption allows the imagery model to be rejected on *a priori* grounds.

The imagery model, however, can be ruled out on empirical grounds as well. To apply the imagery model to the Clark & Chase results, we would have to make two assumptions: (1) that the imagery model could provide a plausible account of why

an image of *A* above is easier to construct than an image of *B* below for positive sentences; and (2) that subjects use the same procedure to form a single image from negative sentences like *B isn't above A* and *A isn't below B*, a strategy that *was* possible in this particular experiment. The results were incompatible with the imagery model on two counts. First, it was found that *B isn't above A* took less time to verify than *A isn't below B*. By the imagery model, forming an image from bottom up would have to be easier than forming an image from top down for these negative sentences, whereas exactly the opposite would hold true for positive sentences (cf. Assumption 1). Hence the imagery assumption is not consistent with these results. Secondly, it was found that for positive sentences true responses are faster than false responses, but for negative sentences true responses are *slower* than false responses. To account for this result, the imagery model would have to assume that each mental image contains information about whether it was constructed from a positive or negative sentence. Images as normally conceived simply do not do this. In contrast, the linguistic model we have proposed correctly accounts for these results in every detail.

In summary, the present experiments have allowed us to make both a negative and a positive point. First, Expts. II, III and IV rule out an attentional-scanning process as a possible explanation of the *above*-ABOVE *below*-BELOW asymmetry in Expt. I. Secondly, these same experiments suggest a model that consists of at least four additive stages: a Sentence Encoding and a Picture Encoding stage, in which the word and picture are represented mentally as abstract symbols, a Comparison stage, in which the resulting two representations are compared symbolically, and a Response stage where the result of the Comparison – either *true* or *false* – is mapped on to a 'true' or 'false' button press. And it was shown that it is the semantic interpretation of the words and pictures that lies at the heart of the *above*-ABOVE *below*-BELOW asymmetry.

This work was supported by Public Health Service Research Grant MH-07722 from the National Institute of Mental Health. We would like to thank Calvin Nodine for suggesting Expt. II and Michelene Chi Chase for running subjects. We are indebted to Michael Posner for his critical comment on an earlier draft.

#### REFERENCES

- CLARK, H. H. (1969). Linguistic processes in deductive reasoning. *Psychol. Rev.* **76**, 387–404.
- DESOTO, C., LONDON, M. & HANDEL, S. (1965). Social reasoning and spatial paralogic. *J. Person. soc. Psychol.* **2**, 513–521.
- HUTTENLOCHER, J. (1968). Constructing spatial images: a strategy in reasoning. *Psychol. Rev.* **75**, 550–560.
- POSNER, M. I. (1969). Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (eds.), *Advances in Learning and Motivation*, vol. 3. New York: Academic Press.
- POSNER, M. I. & MITCHELL, R. F. (1967). Chronometric analysis of classification. *Psychol. Rev.* **74**, 392–409.
- SEYMOUR, P. H. K. (1969). Response latencies in judgements of spatial location. *Br. J. Psychol.* **60**, 31–39.
- STERNBERG, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta psychol.* **30**, 276–315.

(Manuscript received 19 June 1970)