JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 11, 265–277 (1972)

Difficulties People Have in Answering the Question "Where Is It?"[1]

HERBERT H. CLARK

Department of Psychology, Stanford University, Stanford, California 94305

It was proposed that people instructed to insert an object into an array must implicitly answer the question "Where should it be?" and then place the object there, and that this question is easier to answer for A than B given the instructions *A is higher/lower than B* and *A isn't as high/low as B* In Experiment I, *S*s read instructions like *Blue is lower than pink* and then placed a blue (or pink) object into an array; in Experiment II, *S*s answered questions like *If John is lower than Mary, then where is Mary?* Both proposals were supported by the parallel response latencies in the two experiments. Finally, it was shown that Experiment III, a variation on Experiment I, provides critical counterevidence to Huttenlocher's Imagery Theory of reasoning.

Recent studies have shown that children have trouble following certain instructions that require them to place objects into arrays (Bem, 1970; Huttenlocher & Strauss, 1968: Huttenlocher, Eisenberg, & Strauss, 1968). In several of these studies, for example, the experimenter showed the child a ladderlike structure with a block on one of its rungs, handed the child a second block, and instructed him, "Make it so that the blue block is on top of the red block " This instruction was easy for the child as long as he had to place the blue block into the array with respect to the red block; but it was difficult whenever the red block was movable and the blue block was already fixed in the array. These and other similar phenomena, it is proposed, can be accounted for by a "Question-answering" theory of following instructions, an hypothesis which is a natural extension of a recent theory of linguistic comprehension (Clark, 1969a). The first major goal of the present paper is to specify and test this proposal. But the paper has a second goal as well. Huttenlocher and her associates (Huttenlocher, 1968; Huttenlocher, Higgins, Milligan, & Kauffman, 1970; Huttenlocher & Higgins, 1971) have recently used these same phenomena as the basis for an "Imagery" theory of deductive reasoning, a theory designed to account for the difficulties people have in solving three-term series problems like *If Abel is better than Baker, and Charlie is worse than Baker, then who is best*? This theory, however, is incompatible in certain of its predictions with the Question-answering Theory and its progenitor. Thus, the second major goal of this paper is to test these predictions.

THE QUESTION-ANSWERING THEORY OF FOLLOWING INSTRUCTIONS

The Question-answering Theory assumes that following instructions is a process with at least two relatively distinct stages. Basically, the child following an instruction must (1) implicitly ask himself what he is supposed to do and (2) then do it. The question he asks himself at the first stage depends mainly on the extralinguistic constraints of the task. When told, "Make it so that the blue block is on top of the red block," the child holding a red block must implicitly ask himself something like "Where should I put the red block?" or "Where should the red block be?" He asks this

[1] This research was supported in part by Public Health Service Research Grant MH-20021 from the National Institute of Mental Health I wish to thank P. Carpenter, E. V. Clark, M. Just, C. E. Offir, G. Olson, R. Revlis, and E. E. Smith for commenting on earlier drafts of this paper and Suzy Weiner for conducting the experiments.

question of the red block because the task requires him to place the red block, not the blue block. At the next stage, then, he places the block in accordance with his answer. At the first stage, under this view, the child must solve an implicit quasilinguistic problem, here *If the blue block is on top of the red block, then where is the red block?* The difficulties in this type of problem are already predictable from a recently proposed theory of answering questions (Clark, 1969a). Fundamental to the theory is the "principle of congruence," which states that whenever the premise (here, *the blue block is on top of the red block*) and the question (here, *Where is the red block?*) are congruent in their underlying representations, then the question is easy to answer; otherwise, the question is difficult to answer. Thus, the Question–answering Theory claims that the difficulties in following simple instructions occur mainly at the question–answering stage, and that the difficulties at this stage are predictable from the Principle of Congruence.

For the application of the theory to locative instructions, consider the sentences *A is above B* and *B is below A.* They differ in that the former describes the location of A with respect to the "reference point" B, whereas the latter locates B with respect to the "reference point" A. Since the question *Where is A?* asks for the location of A, it is appropriately answered by *A is above B*, which describes the location of A, but not by *B is below A*, which instead describes the location of B. More formally, *Where is A?* is equivalent approximately to *A is at what place?* and *A is above B*, to *A is at a place up from B?* (Katz & Postal, 1964; Leech, 1970). So *Where is A?* and *A is above B* are congruent in their underlying representations with respect to their point of reference; *Where is A?* and *B is below A* are not. Other locative sentences have the same properties. By the Principle of Congruence, therefore, the child should find it easier to place the block mentioned in the subject of a locative sentence, since the implicit question he must answer at the critical stage in the process is congruent with that sentence. This is exactly what Huttenlocher and Strauss (1968) found and so their data are consistent with the Question-answering Theory (see also Clark, in press).

In the present paper, the Question-answering Theory will be tested more directly with the use of constructions like those in (1)–(4):

(1) A is higher than B.
(2) A is lower than B.
(3) A isn't as high as B.
(4) A isn't as low as B.

Like the locatives, each of these sentences describes the location of the subject term (A) with respect to the "reference point" specified by the predicate term (B). Thus, *Where is A?* is properly answered by all four sentences, and *Where is B?* by none of them. This shows that *Where is A?* is congruent in these crucial respects with sentences (1)–(4), whereas *Where is B?* is not. According to the Question-answering Theory, therefore, since *S*s should find it easier to answer the implicit question *Where is A?* than *Where is B?* for sentences (1)–(4), they should consequently find it easier to place A than B into an array for all four of these instructions. This prediction was tested in Experiment I. In addition, this theory implies that a question-answering task with the explicit questions *Where is A?* and *Where is B?* should produce results very similar to a placement task with its implicit questions. This prediction was tested in Experiment II.

Experiment I

Method The *S*s in Experiment I were presented with displays containing a sentence on the left (e g., *Blue is higher than pink*) and a pink or blue line on the right While timed, they were to indicate as quickly as possible where the missing second line should go.

The 32 displays used were constructed by pairing each of 16 sentences with each of two lines. The 16 sentences were: *Blue is higher than* (*is lower than, isn't as high as, isn't as low as*) *pink*, *Blue is better than* (*is worse than, isn't as good as, isn't as bad as*) *pink*, plus the same eight sentences with *blue* and *pink* interchanged. (The *good–bad* problems were included for reasons that will become clear later) The line on the right was either pink or blue Each display was centered in a 3.5 × 5-in. window and was viewed at a distance of 32 in. in a Gerbrands three-field tachistoscope. Each sentence was typed in elite type, and the 5/8-in. long line was centered an average of 1 in. to the right of the end of the sentence. Each S was given a practice block of 16 trials, which consisted of a random subset of the 32 displays, followed by five blocks of 32 trials, each block consisting of an individually randomized ordering of the 32 displays.

The S was instructed to read the sentence, then look over at the line, decide which line was missing, and "draw in" the missing line by pressing an upper or lower response button depending on whether the line was to go above or below the line already there On each trial, the S pressed a "start" button, and 500 msec later the blank "stare" field was replaced by the display. The S responded by pressing the upper or lower button on a hand-held box with the thumb assigned to that button The "start" button lay between the two response buttons within easy reach of either thumb. The S was timed in hundredths of a second from the instant the display appeared to the instant he pressed a response button (at which time the display went off) It was stressed that the S should try to make as few errors as possible yet respond as quickly as possible. For the sentences containing *good* and *bad*, the S was told to imagine placing good and bad objects in vertical arrays with the best object on top and the worst object on the bottom

The Ss, 12 Stanford University undergraduates, were paid $1 75 for their services The left hand was assigned to the upper response button for half the Ss and to the lower response button for the other half Both groups of Ss reversed this assignment halfway through the experiment (half the Ss between Blocks 2 and 3 and half between Blocks 3 and 4). The experiment required a single 45-min session for each S

Results. The results of most interest are the eight mean response latencies shown in Table 1. They were calculated as follows. First, the two displays that differed only in the color mentioned were considered one "class" of display. For instance, the display containing *Blue is higher than pink* and a pink line and the display containing *Pink is higher than blue* and a blue line were classed together as *A is higher than B* where the S was to "place A " Second, the mean latency was calculated for the ten or fewer correct responses for each of the 16 classes of displays for each S. Finally, the resulting 192 means, later submitted to an analysis of variance, were averaged across Ss and across *high–low* and *good–bad* to give the latencies listed in Table 1.

The latencies in Table 1 form a strikingly simple pattern. First, it was easier to "place A" than to "place B" by an average of 215 msec for all four types of sentences, $F(1, 11) = 21.45$, $p < .001$, and there was no interaction between this difference and the type of sentence involved. Second, instructions containing the "unmarked" adjectives *high* or *good* were easier than those containing the "marked" adjectives *low* or *bad* (Clark, 1969a) by an average of 229 msec, $F(1, 11) = 19.01$, $p < .001$. And finally, the comparative instructions were easier than the negative equative instructions by an average of 843 msec, $F(1, 11) = 184.64$, $p < .001$. Indeed, the means in Table 1 are quite accurately accounted for by a simple additive model requiring only four parameters (each estimated from all the data by the method of least-squares): $a = 215$ is the time taken to "place B" over that taken to "place

TABLE 1

LATENCIES (IN MSEC) TO PLACE A OR B FOR EACH SENTENCE IN EXPERIMENT I[a]

Sentence	Place A (I)	Place B (II)	Difference (II–I)
1 A is higher/better than B	2120 (1)	2375 (1)	+255
2 A is lower/worse than B	2338 (0)	2508 (2)	+170
3 A isn't as high/good as B	2951 (1)	3123 (7)	+172
4 A isn't as low/bad as B	3189 (3)	3452 (10)	+263

[a] Note. Numbers in parentheses are percent errors.

A"; $b = 229$ is the time taken by marked adjectives over that taken by unmarked adjectives; $c = 843$ is the time taken by negative equatives over comparatives; and $t_0 = 2113$ is the "base" time taken up by operations not included in a, b, and c. Thus, each condition listed in Table 1 should require t_0 plus a, b, or c (or some combination of a, b, and c), depending on whether the condition requires the "placing of B," contains a marked adjective, or is a negative equative instruction (or involves some combination of these). The latencies predicted by this model have a root mean-squared deviation (RMSD) of 35 msec from the actual mean latencies (Sternberg, 1969), and this deviation is not significant. The model, with its four parameters, accounts for 99.4% of the variance among the actual means, with their eight degrees of freedom. Thus, this is quite an accurate model.

The data in Table 1 were collapsed across *high–low* and *good–bad* because this difference in adjectives did not interact significantly with anything else in the experiment, although its interaction with markedness did approach significance, $F(1, 11) = 4.16$: *high* was 141 msec faster than *low*, whereas *good* was 318 msec faster than *bad*. The only other significant effect was that the adjective pair *high–low* was easier overall than the pair *good–bad* by an average of 123 msec, $F(1, 11) = 7.15, p < .025$.

The *S*s made errors on an average of only 3.1% of the trials (range .6–7.5%). The error rates for each condition, as shown in Table 1, never exceeded 10% and, despite their small magnitude, showed much the same pattern as the latencies. That is, there were fewer errors on the Place A conditions than the Place B conditions, fewer on unmarked adjectives than on marked adjectives, and fewer on comparative constructions than on negative equative constructions, all in agreement with the pattern of latencies.

Experiment II

Method. In Experiment II, the *S*s were presented individual premise-question pairs (e.g., *If John isn't as low as Mary, then where is Mary?*) and were required to indicate their answer as quickly as possible There were 16 displays, each consisting of one of eight premises—*John is higher than* (*is lower than, isn't as high as, isn't as low as*) *Mary* plus the four premises with *John* and *Mary* interchanged—and one of two questions—*Where is John?* and *Where is Mary?* Each *S* was given one practice and ten experimental blocks of 16 trials, each block consisting of an individually randomized ordering of the 16 displays. In all other respects, Experiment II was the same as Experiment I. The *S*s indicated their response by pressing either the upper or lower response button depending on whether the questioned person was above or below. The 12 *S*s, again paid Stanford University students, were counterbalanced for hand assignment, and they reversed this hand assignment between Blocks 5 and 6

Results. The analysis of Experiment II proceeded in the same way as in Experiment I. Means were calculated for the 10 or fewer correct responses for each display of each *S*, and the resulting 192 means were averaged appropriately for Table 2. The mean latencies

TABLE 2

LATENCIES (IN MSEC) TO ANSWER WHERE IS A? AND WHERE IS B? FOR EACH SENTENCE IN EXPERIMENT II[a]

Sentence	Where is A? (I)	Where is B? (II)	Difference (II–I)
1 A is higher than B	1914 (0)	2068 (2)	+154
2 A is lower than B	1970 (3)	2186 (1)	+216
3 A isn't as high as B	2376 (1)	2560 (5)	+184
4 A isn't as low as B	2492 (5)	2620 (12)	+128

[a] Note Numbers in parentheses are percent errors.

in Table 2 bear a striking resemblence to those of Experiment I. First, it was easier to answer *Where is A?* than *Where is B?* by an average of 170 msec, regardless of the sentence type, $F(1, 11) = 24.52$, $p < .001$, and there was no interaction between this difference and sentence type. Second, instructions containing the unmarked adjective *high* were easier than those containing the marked adjective *low* by an average of 86 msec, although this difference was not quite significant, $F(1, 11) = 4.11$, $p < .10$. And third, comparative instructions were easier than negative equative instructions by an average of 479 msec, $F(1, 11) = 78.17$, $p < .001$. These means can be accounted for accurately by the same additive model as in Experiment I. Here the estimates of the parameters t_0, a, b, c are 1906, 170, 86, and 479 msec, respectively. The RMSD of this model from the actual latencies is only 17.7 msec, and this is not significant. The four degrees of freedom of the model accounts for 99.5% of the variance in the eight means of Table 2. There were no other significant effects in the latencies.

The *S*s in Experiment II made errors on an average of only 3.8% of the trials (range .6–5%). Again, the error rates on each condition, shown in Table 2, were parallel to the latencies, with fewer errors on *Where is A?* than on *Where is B?*, fewer on unmarked adjectives than marked adjectives, and fewer on comparative than negative equative sentences.

Discussion

Experiments I and II support both the Question-answering Theory and the Principle of Congruence on which the theory is based.

The Principle of Congruence. This principle predicts that a question asked of a sentence will be easier to answer if it is congruent with that sentence than if it is incongruent. In Experiment II, this predicts that *Where is A?* should be easier to answer than *Where is B?* for sentences (1)–(4) in Table 2. The results confirmed this prediction in detail. To see the generality of congruence, one should contrast these findings with earlier findings on the so-called two-term series problems (e.g., *If John isn't as good as Bill, then who is best?*) reported by Clark (1969a). In that case, a sentence and question were congruent whenever the adjective underlying the question was the same as the adjective underlying the sentence. That is, the Principle of Congruence predicts that *Who is best?* (with its underlying adjective *good*) should be easier for sentences like (1 *A is better than B* and (3) *A isn't as good as B*, where the underlying adjective is *good*, but that *Who is worst*? (with its underlying adjective *bad*) should be easier for sentences like (2) *A is worse than B* and (4) *A isn't as bad as B*, where the underlying adjective is *bad*. The results of Clark (1969a) were consistent with these predictions. The contrast between these two sets of results can be summarized this way: With the questions *Who is best?* and *Who is worst?*, it is easier to answer questions about A than B for sentences like (1) and (2), but easier to answer questions about B than A for sentences like (3) and (4); on the other hand, with the questions *Where is A?* and *Where is B?* it is easier to answer questions about A than B for all four sentences, (1)–(4).

The findings on *where*-questions and *best/worst*-questions together rule out plausible alternative explanations of either result alone. For the *where*-questions alone, one might propose simply that information was more easily retrieved from sentences (1)–(4) about the subject term (A) than about the predicate term (B) But this theory fails to account for the results on *best/worst*-questions. Or, for the *best/worst*-questions alone, one might propose that information was more easily retrieved from sentences (1)–(4) about the term more extreme on the underlying adjectival scale, i.e., the "better" term for the sentences containing *good*, and the "worse" term for the sentences containing *bad*. But this theory fails to account for the results on *where*-questions. Variations on these two themes are inadequate for the same reason. The Principle of Congruence works, therefore, because it considers both the sentence and the question asked of it. Theories that disregard the sentence–question interaction, like the two just examined, necessarily fail to account for these two sets of findings.

Elsewhere (Clark, in press; Clark & Chase, in press), the Principle of Congruence has also been shown to be consistent with the question–answering experiments of Smith and McMahon (1970) and Wright (1969) and with most of the previous verification experiments on negation, for example, by Wason (1961; Wason & Jones, 1963), Gough (1965; 1966), Wales and Grieve (1969), Clark and Chase (in press), Just and Carpenter (1971), and Trabasso, Rollins, and Shaughnessey (1971). It is gratifying that the present data are consistent with this apparently well-established principle.

The Question-answering Theory Experiments I and II together also offer considerable support for the Question-answering theory as a theory of following instructions In this theory, it is assumed that the critical stage of the placement task is one in which *S*s implicitly answer a *where*-question. Therefore, the pattern of latencies in this task should be parallel to the pattern of latencies in a task in which *S*s explicitly answer the comparable *where*-questions. This prediction is supported in some detail. The main evidence is that the advantage of Place A over Place B in Experiment I is parallel to the advantage of *Where is A?* over *Where is B?* in Experiment II. This prediction is upheld even more fully by the fact that both sets of latencies could be fitted by the same additive model. Furthermore, the three parameters, a, b, and c—those specifying the Place A–Place B, *high–low*, and positive–negative differences, respectively—are quite similar for the two experiments, estimated at 215, 141 and 843 msec for Experiment I and at 170, 86, and 479 msec for Experiment II. However, the positive–negative parameter c, for some reason, was reliably different for the two experiments, $t(22) = 2.10$, $p < .05$.

A Test of the Imagery Theory

The Imagery Theory, as proposed in Huttenlocher (1968), views the *S* as placing the three terms of a three-term series problem into a mental array one by one. For the premise combination *Abel is better than Baker, Charlie is worse than Baker*, for instance, the *S* first places Abel and Baker of the first premise into the array, then he places the "new" term of the second premise (*Charlie*) into the array with respect to the "old" term of the second premise (*Baker*). The essence of the theory is that it makes predictions from actual placement tasks to reasoning tasks. It treats the "new" term in the second premise exactly as if it were the movable block—the one in the child's hand—in the placement task. Huttenlocher's argument goes as follows. Assumption 1: In reasoning tasks, the *S* has to place either A or B of the second premise *A is better* (*or worse*) *than B* into a mental array that already contains two items. Assumption 2: In actual placement tasks with such sentences, *S*s find it easier to place A than B into a physical array that already contains two objects. Conclusion: In reasoning tasks, *S*s should therefore find problems easier whenever the "new" term to be placed is A rather than B. In support of the Imagery Theory, this conclusion is consistent with most of the previous data on three-term series problems containing comparative premises (Hunter, 1957; DeSoto, London, & Handel, 1965; Huttenlocher, 1968; Clark, 1969a, b).

In a further test of the Imagery Theory, however, Clark (1969a) applied the same logic to problems containing negative equative premises (e.g., *If Charlie isn't as good as Baker, and Able isn't as bad as Baker, then who is best?*). In that test, the argument based on the Imagery Theory went as follows: Assumption 1′: In reasoning tasks, the *S* has to place either A or B of the second premise *A isn't as good* (*or bad*) *as B* into a mental array that already contains two items. Assumption 2′: In actual placement tasks with such sentences, *S*s find it easier to place A than B into an array that already contains two items. Conclusion: In reasoning tasks, *S*s should therefore find problems easier when the "new" term to be placed is A rather than B. This prediction, however, was disconfirmed in all of the negative equative problems examined by Clark (1969a); that is, these reasoning problems were found to be consistently harder, not easier, when the "new" term was A rather than B. By this argument, then, the Imagery Theory is incorrect.

Unfortunately, these two tests of the Imagery Theory—one confirmatory and the other disconfirmatory—were not complete, since they depended for their validity on Assumptions 2 and 2′ and neither of these assumptions had been tested by Huttenlocher (1968) or Clark (1969a). Obviously, both of these assumptions are plausible, since Experiment I showed that for both comparative and negative equative sentences *S*s find it easier to place A than B, at least into arrays *that contain only one object*. Nevertheless, to confirm Assumptions 2 and 2′ specifically, one must demonstrate the same advantage for arrays that already contain *two* objects. Experiment III was designed for just this purpose. It consisted of a placement task much like Experiment I except that there were always two objects already in the array instead of only one. If Experiment III is found to support Assumption 2′, then the previous disconfirmation of the Imagery Theory by Clark (1969a) will be complete. In addition, however, Experiment III was designed so that its results could be compared problem by problem against the most precise data available on three-term series problems (Clark, 1969b). If these comparisons are found to go counter to the predictions of the Imagery Theory, then one has to conclude that the Imagery theory is incorrect.

EXPERIMENT III

Methods. Experiment III was essentially the same as Experiment I except that instead of one line on the right side of each display there were two, one above the other. For one group of *S*s, there were 32 displays each constructed from one of eight sentences—*Blue is higher than* (*is lower than, isn't as high as, isn't as low as*) *pink*, plus the same four sentences with *blue* and *pink* interchanged—and from one of four different pairs of colored lines—black on top and blue on bottom, blue on top and black on bottom, black on top and pink on bottom, and pink on top and black on bottom. The other group of *S*s was given the same 32 displays with *good* and *bad* substituted everywhere for *high* and *low*, respectively. Each display was constructed just as in Experiment I, but here the two lines on the right were drawn one typewriter line above and below the line of the sentence.

The *S*s were told to indicate whether the missing blue or pink line went above or below *both* of the lines on the right by pressing the top or bottom button on their response panel On half the displays, however, neither response was appropriate, since the *S*s could not know whether the missing line went above or below the extra black line. One such indeterminate display consisted of *Blue is higher than pink* opposite a black line above a pink line; here, the blue line to be placed could go *either* above both lines *or* between the two lines and still be accurately described by the sentence. The *S*s were instructed to answer these indeterminate problems with a vocal "can't tell," and the latency to the beginning of this response was measured by means of a throat microphone and a voice-activated switch. In all other respects, the procedure was the same as in Experiment I. The two groups of *S*s each consisted of 12 paid Stanford University undergraduates

Results. Of most interest are the latencies shown in Table 3. Those in the top half of the table are for the "determinate" displays (for which *S*s should have pressed one of the two response buttons), and those in the bottom half are for the "indeterminate" displays (for which *S*s should have answered "can't tell"). The "determinate" displays were classified as if they did not contain the black line. The "indeterminate" displays were classified separately in exactly the same way. Note, therefore, that the "indeterminate" display consisting, for example, of *Blue is higher than pink* opposite a black line above a pink line was classified as Place A for *A is higher than B*. This classification is appropriate since the *S*'s task was to try to place the missing blue line (A) with respect to the pink line (B), even though he could not do so and had to answer "can't tell." Finally, the mean latency of the 10 or fewer correct responses was calculated for each of the 16 classes of display shown in Table 3 for each *S* separately, and these 384 means, later submitted to an analysis of variance, were averaged across *S*s to produce the latencies in Table 3.

The most important result of this experiment was that *S*s found it easier to "place A" than to "place B" for all eight sentences in Table 3. "Place A" had a 362 msec advantage over "place B" overall, $F(1, 22) = 48.42$, $p < .001$, although this advantage was a significant 199 msec larger for the comparative "determinate" and negative equative "indeterminate" problems than for the rest, $F(1, 22) = 7.91$, $p < .025$. In individual analyses of variance, "place A" was found to be favored over "place B" on each of the eight sentences separately, with $F(1, 22) = 30.51$, 73.72, 2.46, 5.02, 10.72, 7.21, 6.69, and 13.39, respectively; these F's are all significant at least $p < .05$, except for sentence (3) *A isn't as high/good as B* in the "determinate" displays. These eight individual analyses also indicated that the advantage of "place A" over "place B" for each sentence was independent of the adjective pair considered (*good–bad* or *high–low*), and so the data were combined over the two adjective pairs in Table 3.

The other results were: (a) Positive sentences were 903 msec faster overall than negative sentences, $F(1, 22) = 82$, $p < .001$. (b) Unmarked adjectives were an average of 173 msec faster than marked adjectives, $F(1, 22) = 22.88$, $p < .001$, and this difference did not interact with adjective pair. (c) The advantage of unmarked over marked adjectives was greater for negative sentences (315 msec) than for positive sentences (31 msec), $F(1, 22) = 15.75$, $p < .001$. And finally, (d) the "determinate" problems were resolved 303 msec faster overall than the "indeterminate" problems, $F(1, 22) = 47$, $p < .001$. Again, the overall error rate was a low 3.6% (range .1–8.8%) and the pattern of errors coincided fairly closely with the pattern of latencies (see Table 3).

The simple additive model used in Experiments I and III can also be fitted to the mean latencies in Table 3, but because of the interactions just reported, the fit is not nearly so good. The model could be applied to the top and bottom halves of the table separately, or with the addition of parameter d—the difference between the "determinate" and "indeterminate" problems—all 16 latencies can be fitted at once. In the latter case, with t_0, a, b, c, and d estimated as 2226, 362, 173, 903, and 303 msec, respectively, the model, with

TABLE 3

LATENCIES (IN MSEC) TO PLACE A OR B FOR EACH SENTENCE IN EXPERIMENT III[a]

Type of problem	Sentence	Place A (I)	Place B (II)	Difference (II–I)
Determinate	1 A is higher/better than B	2239 (1)	2678 (4)	+439
	2 A is lower/worse than B	2230 (1)	2777 (6)	+547
	3 A isn't as high/good as B	3153 (3)	3359 (4)	+206
	4 A isn't as low/bad as B	3401 (4)	3724 (10)	+323
Indeterminate	1 A is higher/better than B	2664 (2)	2937 (2)	+273
	2 A is lower/worse than B	2694 (1)	2940 (2)	+246
	3 A isn't as high/good as B	3348 (2)	3702 (2)	+354
	4 A isn't as low/bad as B	3597 (2)	4101 (10)	+504

[a] Note. Numbers in parentheses are percent errors.

5 degrees of freedom, accounts for 96.9 % of the variance among the 16 means, with a RMSD of 93 msec. But as pointed out above, there are significant deviations of the data from this model.

Discussion

The results of Experiment III are relevant to three issues: the validity of the Imagery Theory, the interpretation of certain related experiments by Huttenlocher *et al.* (1970), and the validity of the Question-answering Theory of following instructions.

The Imagery Theory. It was noted above that the disconfirmation of the Imagery Theory in Clark (1969a) depended on Assumption 2′, namely, that it is easier to "place A" than to "place B" for negative equative sentences in a task, like Experiment III, with two fixed objects. Indeed, Huttenlocher (1968) in referring to a placement task much like Experiment III said, "This task may be regarded as a concrete form of the three-term series problem (p. 553)." Experiment III confirmed Assumption 2′, both in the "determinate" and the "indeterminate" problems. Therefore, it follows that the original disconfirmation of the Imagery Theory in Clark (1969a) is complete.

Experiment III, however, was specifically designed so that it could be compared directly with a previous experiment on three-term series problems by Clark (1969b). Each of the 64 problems in Clark (1969b) can be classified as to whether it describes a "determinate" or "indeterminate" ordering of the three terms, as to whether its second premise is sentence type (1), (2), (3), or (4), and as to whether, according to the Imagery Theory, the last item to be placed into the *S*'s mental array is the subject term ("place A") or the predicate term ("place B") of the second premise. Table 4 presents the difficulty scores for the 64 problems classified in this way. Each of these scores is an average across four problems and is based on the proportion of the 100 *S*s unable to solve each problem correctly within 10 sec (see Clark, 1969b, for details).

According to the Imagery Theory, difficulties in the placement task (Table 3) should be directly reflected in difficulties in the reasoning task (Table 4). Let us define an "A-advantage" as the case when Place A is easier than Place B and so there is a plus in the final column of Table 3 or 4. A "B-advantage" is just the reverse. The Imagery Theory, then, predicts that whenever there is an A-

TABLE 4

PERCENT ERRORS FROM CLARK (1969b) ON THREE-TERM SERIES PROBLEMS CLASSIFIED ACCORDING TO THE IMAGERY THEORY[a]

Type of problem	Second premise	Place A (I)	Place B (II)	Difference (II–I)
Determinate	1 A is better than B	12	25	+13**
	2 A is worse than B	15	42	+27**
	3 A isn't as good as B	34	26	− 8*
	4 A isn't as bad as B	38	26	−12*
Indeterminate	1 A is better than B	25	23	− 2
	2 A is worse than B	25	26	+ 1
	3 A isn't as good as B	32	38	+ 6
	4 A isn't as bad as B	40	34	− 6

[a] Note. Each unsigned percentage is based on 400 data points.
* $p < .02$.
** $p < .001$.

advantage in Table 3, there should also be an A-advantage in Table 4. Obviously, however, the pattern of all pluses in Table 3 is not duplicated in Table 4, where there is a mixture of pluses and minuses. But let us examine these differences in detail.

In the "determinate" problems, the placement task (Table 3) showed an A-advantage for sentences (1)–(4), and so the reasoning task should show an A-advantage on all four sentences too. That this prediction is correct for sentences (1) and (2) is not surprising, for the Imagery Theory was originally developed to account for previous data on these problems. But the prediction of an A-advantage fails entirely for sentences (3) and (4). This failure can be seen in other data as well. The pattern of plus and minus A-advantage scores in Table 4 (viz., +13, +27, −8, and −12% errors on the four sentences) has been replicated by Clark (1969a) on *good–bad* reasoning problems (+340, +670, −220, and −450 msec), by Huttenlocher *et al.* (1970) on *good–bad* problems (+220, +910, −80, and −720 msec), and by Clark (unpublished data) on *high–low* problems (+165, +576, −900, −1585 msec). For some reason, Huttenlocher *et al.* failed to confirm this pattern on *high–low* problems, finding instead essentially no latency differences at all. But this null pattern is also inconsistent with the prediction of the Imagery Theory given the results of Table 3. In the "indeterminate" problems, the placement task (Table 3) again showed a strong A-advantage for all four sentences, so the reasoning task should turn up with the same pattern. While this prediction is weakly confirmed for sentence (3), it is disconfirmed by the null differences on sentences (1) and (2) and by the moderate B-advantage on sentence (4). That is, the Imagery Theory predicts a strong overall A-advantage on the "indeterminate" problems in Table 4, yet the table shows a strikingly null A-advantage (actually −0.06%) based on 3200 data points.

In all, therefore, the Imagery Theory is incorrect on at least five out of eight specific predictions (and on five out of six new predictions) in going from Table 3 to Table 4. As Huttenlocher *et al.* (1970) pointed out: "To demonstrate the falsity of (the Imagery Theory), one would have to show that the relative difficulty of the different forms of 'negative equative' syllogisms differs from the relative difficulty of arranging real items according to corresponding 'negative equative' descriptions (p. 335)." The comparison of Tables 3 and 4 constitutes just such a demonstration, and so by this criterion, the Imagery Theory is incorrect (see also Clark, 1971).

The Huttenlocher experiments. Experiment III also raises several questions about how one should interpret two "arrangement" tasks that Huttenlocher *et al.* (1970) have reported recently. In these tasks, the *S* was required to listen to the first premise of a three-term series problem, place two felt figures onto a felt board in accordance with this premise, then listen to the second premise and question, and, while timed, place the third felt figure into the array. The premises in the task of interest contained either *good* or *bad* in comparative or negative equative constructions. Superficially, this "arrangement" task resembles the placement task of Experiment III very closely, for in both cases the *S* was timed as he placed a third object into a physical array. One should expect, therefore, that the Huttenlocher *et al.* "arrangement" task should produce the same results as the placement task But it did not. Whereas the usual A-advantage was found in the "arrangement" problems with comparative premises, there was a B-advantage in the "arrangement" problems with negative equative premises. The latter result is exactly contrary to Experiment III.

The reason for this discrepancy could well be that the *S*s of Huttenlocher *et al.* viewed the "arrangement" task primarily as a reasoning task and only incidentally as a placement task. Recall that these *S*s were always required to listen to a complete three-term series problem (i.e., two premises and question) before placing the third felt figure. Thus, it is possible that they solved the problems in their heads, decided what the arrangement of figures should be from their reasoning, and placed the third figure accordingly.[2] By this argument, the *S*s were reasoning on this task, and the results should have coincided approximately with those of other reasoning tasks, as they did. Unlike the *S*s in the Huttenlocher *et al.* "arrangement" task, however, the *S*s in the present Experiment III did not receive a first premise or a question and could therefore not have been reasoning in this sense. This difference between the "arrangement" and placement tasks is critical. It suggests that the "arrangement" task is not a true test of Assumption 2′, as Huttenlocher *et al.* have claimed, since its results could have been contaminated by *S*s who reason before they arrange the third figure But because Experiment III does not contain this confounding, it is a proper test of Assumption 2′. It shows that the Imagery Theory is incorrect.

In arguing for the correctness of the Imagery Theory, Huttenlocher *et al.* (1970) had to explain why it was that B was easier to place into a mental array than A whenever the second premise was a negative equative in a "determinate" three-term series problem (see Table 4). In their account, Huttenlocher *et al.* proposed that *S*s "re-organize" such premises in particular ways to make them easier to comprehend. One proposal was that *A isn't as bad as B* is usually reorganized implicitly as *B is worse than A* and *A isn't as good as B*, often as *B is better than A*. Since B becomes the subject of a comparative sentence in such reorganizations, B should therefore be easier to place than A, and this accounts for the effect Huttenlocher *et al.* wanted to explain. But according to the Imagery Theory, placement and reasoning tasks are almost identical in their processing; therefore, the same reorganizations should appear in the present placement experiments. The reorganizations predict a partial or full B-advantage for sentences (3) and (4) in Experiments I and III. Instead, there was a significant A-advantage for these conditions in both experiments, and the prediction is disconfirmed. Even if the proposed reorganizations occurred only part of the time, then at the very least the A-advantage should be less for (3) and (4) than for (1) and (2) in Experiments I and III. The respective differences, however, are −5 msec. and +30 msec in favor of this prediction, and neither of these small differences approaches significance.

To account for results on their *high–low* "arrangement" task, Huttenlocher *et al.* also proposed that *A isn't as high as B* and *A isn't as low as B* are reorganized as *A is lower than B* and *A is higher than B*, respectively. After these reorganizations, there should be an A-advantage, since A has become the subject of a comparative sentence. Although this prediction by itself is consistent with Experiments I, II, and III, it implies that *high–low* premises should show a pattern of latencies different from *good–bad* premises. This implication receives no confirmation in either Experiment I or Experiment III, where it is possible to compare the two types of premises directly.

The proposal that *S*s reorganize premises, however, gets into difficulty even when the reasoning data in Table 4 are considered by themselves. In the Huttenlocher *et al.* (1970) proposal, simple comparatives like *A is better* (or *worse*) *than B* would never be reorganized. Thus, there should be an A-advantage on every reasoning problem in which the second premise is a simple comparative. But as Table 4 indicates, although the expected A-advantage appears for sentences (1) and (2) in the "determinate" problems, it does not appear for sentences (1) and (2) in the "indeterminate" problems, as this proposal predicts it should. Similarly, under the Huttenlocher *et al.* proposal, negative equatives would be reorganized so that there should always be a B-advantage on problems with such premises. Although the expected B-advantage

[2] The details of such a process are discussed more fully in H. H. Clark, "More about linguistic processes in deductive reasoning," in preparation.

appears for sentences (3) and (4) in the "determinate" problems, it appears for sentence (4) but not (3) in the "indeterminate" problems. In other words, the proposed reorganizations account only for the "determinate" problems; the proposal is incompatible with the data on the "indeterminate" problems. This difficulty might be overcome by assuming that *S*s reorganize sentences (1) and (2), for instance, only when the problems are "indeterminate." But such selective reorganizing would not work either, because the *S* could not know whether a problem was "indeterminate" or not until *after* he had solved it.

To sum up, it has been argued that the Huttenlocher *et al.* "arrangement" tasks appear to be confounded with reasoning, and so they are unreliable tests of the Imagery Theory. Nevertheless, even if one did accept the "arrangement" tasks at face value, then one would have to assume, with Huttenlocher *et al.*, that *S*s "reorganize" negative equative premises in placement and reasoning tasks. These reorganizations, however, receive no support whatever from Experiments I, II, and III. In addition, the assumed reorganizations are incompatible with the reasoning data in Table 4 even when considered alone. All this provides still more evidence that the Imagery theory is incorrect.

The Question-answering Theory. The same rationale that accounts for Experiments I and II also accounts fairly well for Experiment III. According to this theory, the question-answering stage comes first, and it should result in an A-advantage for sentences (1)–(4) regardless of whether the problem is "determinate" or "indeterminate." The data agree with this prediction. The placement stage comes second. Here, for "determinate" problems, the *S* is able to place the missing line immediately; however, for "indeterminate" problems, he has to decide that the position of the missing line is not necessarily either above or below the black third line and so he must answer "can't tell." This description at least suggests that the placement stage might take longer for "indeterminate" problems, where placement is blocked, than for "determinate" problems, where placement is immediate. The latencies are consistent with this suggestion showing about a 300-msec advantage for the "determinate" problems, although this advantage is confounded with mode of response. Thus, the effects of the question-answering stage, which predicts an A-advantage for all sentences, and the placement stage, which suggests an advantage for the "determinate" problems, are approximately separate and additive.

There were, however, significant deviations from this simple form of the model. The major deviation was that the A-advantage for sentences (1) and (2) was about 230 msec longer than for sentences (3) and (4) on the "determinate" problems, but about 170 msec smaller on the "indeterminate" problems. This suggests that the question-answering and placement stages are not quite separate and additive, but rather information from the first stage somehow affects the second. How could this occur? As proposed in Clark (1969a; 1970), the instructions containing *good*—sentences (1) and (3) in Table 3—presuppose the relative goodness of A and B, and those containing *bad*—(2) and (4)—presuppose the relative badness of A and B. One possibility is that *S*s are reluctant to place the missing line, or attempt to place the missing line, at the good or bad extreme —the top or bottom of the array of lines—when the presuppositions of the instruction are not congruent with that placement. This tendency, based on a modified Principle of Congruence, would predict an inflated A-advantage for sentences (1) and (2) on "determinate" displays and for sentences (3) and (4) on "indeterminate" displays, and a depressed A-advantage for the remaining four sentences, all at the placement stage of the process. This prediction is consistent with the data.

In summary, the present paper has demonstrated that the Question-answering Theory is able to account for the major processes of simple placement tasks as in Experiments I and III, and that the Imagery Theory is incorrect and unable to account for deductive reasoning in three-term series problems. Significantly, the Question-answering Theory is a natural extension of a proposal by Clark (1969a) that *can* account for the main aspects of deductive reasoning in three-term series problems. Together, therefore, the previous proposal and its offspring, the Question-answering Theory, are consistent with the data in both placement and reasoning tasks.

References

BEM, S. L The role of comprehension in children's problem solving *Developmental Psychology*, 1970, **2**, 351–358.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 384–404. (a)

CLARK, H. H. The influence of language in solving three-term series problems. *Journal of Experimental Psychology*, 1969, **82**, 205–215. (b)

CLARK, H. H. Comprehending comparatives. In G. Flores d'Arcais & W. J. M Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North Holland Press, 1970.

CLARK, H. H More about "Adjectives, comparatives, and syllogisms": A reply to Huttenlocher and Higgins. *Psychological Review*, 1971, **78**, 505–514.

CLARK, H. H Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics, Vol 12: Linguistics and adjacent arts and sciences* The Hague: Mouton, in press.

CLARK, H. H , & CHASE, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology*, in press

DESOTO, C., LONDON, M , & HANDEL, S Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 1965 **2**, 513-521

GOUGH, P. B. Grammatical transformations and speed of understanding *Journal of Verbal Learning and Verbal Behavior*, 1965, **5**, 107–111.

GOUGH, P B The verification of sentences· The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 492–496

HUNTER, I. M. L. The solving of three-term series problems. *British Journal of Psychology*, 1957, **48**, 286–298.

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, **75**, 550–560.

HUTTENLOCHER, J., EISENBERG, K., & STRAUSS, S. Comprehension: Relation between perceived actor and logical subject. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 300–304.

HUTTENLOCHER, J., & HIGGINS, E. T. Adjectives, comparatives, and syllogisms *Psychological Review*, 1971, **78**, 487–504.

HUTTENLOCHER, J., HIGGINS, E. T., MILLIGAN. C , & KAUFFMAN, B. The mystery of the "negative equative" construction. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 334–341.

HUTTENLOCHER, J , & STRAUSS, S. Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 527–530.

JUST, M., & CARPENTER, P. Comprehension of negation with quantification *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 244–253.

KATZ, J. J., & POSTAL, P. M. *An integrated theory of linguistic descriptions*. Cambridge, Mass.: M.I.T. Press, 1964.

LEECH, G. N. *Towards a semantic description of English*. Bloomington, Ind.: Indiana University Press, 1970.

SMITH, K. H., & MCMAHON, L. E. Understanding order information in sentences: Some recent work at Bell Laboratories. In G. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North Holland, 1970.

STERNBERG, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica*, 1969, **30**, 276–315.

TRABASSO, T., ROLLINS, H., & SHAUGHNESSY, E. Storage and verification stages in processing concepts. *Cognitive Psychology*, 1971, **2**, 239–289.

WALES, R. J., & GRIEVE, R. What is so difficult about negation? *Perception and Psychophysics*, 1969, **6**, 327–332.

WASON, P. C. Response to affirmative and negative binary statements. *British Journal of Psychology*, 1961, **52**, 133–142.

WASON, P. C., & JONES, S. Negatives: Denotation and connotation. *British Journal of Psychology*, 1963, **54**, 299–307.

WRIGHT, P. Transformations and the understanding of sentences. *Language and Speech*, 1969, **12**, 156–166.

(Received September 30, 1971)