COGNITIVE PSYCHOLOGY 3, 472–517 (1972)

On the Process of Comparing Sentences Against Pictures[1]

HERBERT H. CLARK
Stanford University

AND

WILLIAM G. CHASE
Carnegie–Mellon University

The present study outlines a theory of how people compare sentences against pictures. This theory was tested in four experiments in which Ss were timed as they judged whether a sentence (e.g., *Star isn't above plus*) was true or false of a picture (e.g., $\overset{+}{*}$). The latencies in these tasks were consistent with the thesis that: (1) sentences are represented in terms of elementary propositions; (2) pictures are encoded in the same interpretive format; (3) these two codes are compared in an algorithmic series of mental operations, each of which contributes additively to the response latency; and (4) sentence encoding, picture encoding, comparing, and responding are four serially ordered stages, and their component latencies are additive. From these results, it was also possible to rule out certain explanations based on visual imagery, conversion (e.g., converting *isn't above* into *is below*), reading time, normative word frequencies, and other factors. Finally, it was shown that this theory is consistent with previous studies on sentence comprehension, sentence verification, concept verification, and other related phenomena.

Although people find it exceedingly easy to compare information from linguistic and pictorial sources, e.g., to decide whether a sentence is true or false of a picture, little is known about how they do it. This lack of knowledge is particularly serious since the process of "sentence–picture comparison" is an integral part of many common psychological tasks, e.g., following instructions for perceptual judgments, concept formation, pattern identification, problem solving, sentence verification, formation of visual images. Also, we would argue, this process is fundamental to the study of psycholinguistics, since language is often used to describe perceptual experiences. In the present paper, we are concerned with sentence–picture comparison both broadly and narrowly. Our main aims are to outline a general theory of sentence–picture comparison and then to test it in a series of experiments on the verification of sentences against pictures. Thus, the plan of the paper is to consider, in turn: (1) certain general constraints on how such a theory should be constructed; (2) a specific model designed to account for how people verify sentences like *Star isn't above plus* against simple pictures; (3) confirmatory results from three experiments on sentence verification; (4) several plausible, but empirically defective alternative models for the same process; (5) corroborative evidence from previous experiments; and finally, (6) questions about the processes that are left unanswered by the present theory.

GENERAL CONSIDERATIONS

The main *a priori* requirement of any theory of sentence–picture comparison is that for a sentence and picture to be compared they must be represented, ultimately, in the same mental format. One cannot, for example, compare the printed word *orchestra* directly with a picture of an orchestra and judge them to "mean" the same thing, for there are no properties intrinsic to the graphemic and pictorial modes to indicate that the word and picture represent the same concept. What, then, could the common mental format be? We will assume that this format normally represents the interpretations, rather than the acoustic or pictorial properties, of the sentence and picture, and that these interpretations are coded in terms of propositions. Although these two assumptions will only be illustrated here, they have been more fully justified elsewhere (e.g., Chase & Clark, in press; Grasselli, 1969; Watt, 1966).

The first assumption is that the mental format codes the interpretations rather than the perceptual properties of sentences and pictures. This assumption seems essential because it is their interpretations that lie at the basis of every sentence–picture comparison. Consider the verification of the sentence *It is a young woman* against Leeper's (1935) famous ambiguous drawing which is interpreted on some occasions as a young woman and on other occasions as an old hag. The person "seeing" the young woman in the picture would judge the sentence to be true, but the person "seeing" the old hag would judge it to be false. Or consider the verification of *They are eating apples*, interpreted either as "They are apples for eating" or as "Apples are being eaten by them," against a picture of two delicious looking apples. The person interpreting the sentence in the first way would judge it to be true of the picture, but the person with the second interpretation would judge it to be false.

---

[1] This research was supported by United States Public Health Service Research Grants MH-07722 and MH-20021 from the National Institute of Mental Health. We especially thank Eve V. Clark, Gary Olson, Michael Posner, Edward Smith, William C. Watt, and Richard Young for their helpful comments on earlier drafts of this paper, and Micheline Chi Chase, Barbara Schneiderman, and Dianne Horgan for their assistance in the experiments.

In either case, the perceptual properties of the sentence and picture remain constant, yet the comparison process changes with their interpretations. On the other hand, many different sentences and pictures can be assigned the same interpretation, e.g., *That is water, Das ist Wasser, C'est de l'eau,* and three different pictures of water, yet all of the sentences would be judged true of all of the pictures. According to these illustrations, judgments of truth are independent of the particular perceptual properties of sentences and pictures and are determined completely by the interpretations given to them.

The second assumption is that interpretations are coded in terms of elementary propositions, that is, in terms of names of objects combined with names of their properties. This assumption is best illustrated by a recent empirical study of Nevada cattle brands by Watt (1966, 1967a, 1967b, 1969). What Watt did was (1) construct an "iconic" characterization of the Nevada cattleman's tacit knowledge of which Nevada cattle brands are well formed, (2) construct a linguistic characterization of the correct "readings" or "blazons" the cattlemen assign to each brand (e.g., "Flying W Hanging Tumbling T," "H E Combined Cross S," and "Reversed B D Combined Quartercircle"), and (3) show how these two characterizations are related. In this analysis, Watt found that the cattle brands could be characterized by a generative grammar complete with phrase structure and transformational rules (much like the current generative grammars of English), and so could the corresponding blazons. But in constructing these grammars, Watt found it necessary to make use of simple relational predicates (or functions) with nominals as arguments. In the blazons, the nominals were typically realized as nouns (e.g., "W," "H," and "Quartercircle"), and the predicates as verbs, adjectives, or prepositions (e.g., "Cross," "Combined," and "Tumbling"). In the brands, the nominals were typically realized as primitive geometrical figures (e.g., alphabetic letters and curved lines), and the predicates as spatial relations (e.g., the positioning of one letter above and touching another) or as geometrical alterations (e.g., the sprouting of wings on a letter or the rotation of a letter by 45°). Thus, the use of propositions was essential to the characterization of both the brands and the blazons.

Watt's analysis, however, is also important because it shows the relation between sentence interpretations and picture interpretations. Watt found that the grammar for the blazons, rather than being different from the grammar for the brands, is essentially the same grammar supplemented by a few late "translational" rules. That is, the blazons are simply "translations," or alternative surface realizations, of the brands; at a deeper level, each brand and its corresponding blazon are the same in terms of the rules that generate them. It is impossible to get from the brand to blazon, or vice versa, without making use of the common underlying representation. Because Watt's system characterizes the cattleman's tacit knowledge of brands and blazons, it is an example *par excellence* of a naturally occurring language–picture system. Thus, Watt's evidence is fully consistent with the assumption that people actually do construct interpretations for sentences and pictures and that these interpretations are in a common mental format consisting of propositions.

Although these general considerations suggest that sentences and pictures are interpreted in terms of elementary propositions, they reveal little about how people actually compare one interpretation with another. To examine this process, we now turn to a detailed investigation of a specific theory of sentence–picture comparison.

A THEORY OF SENTENCE–PICTURE COMPARISON

The theory of sentence–picture comparison to be described has been designed mainly to account for a limited type of sentence-verification task. In this task, the subject is shown a display containing a sentence like *Star isn't below line* and a picture of, say, a star above a line; he is typically required to read the sentence, look at the picture, and indicate as quickly as possible whether the sentence is true or false of the picture, all while he is timed. The sentences of this task always make use of *above* or *below* and describe the vertical location of two geometrical figures; the pictures invariably depict two geometrical objects one above the other. The theory presented is meant primarily to account for the latencies of the subject's judgments in this task.

Because the present theory deals in part with the verifications of negative sentences, portions of it can be traced to the extensive earlier work on negation by Wason (1961), Eifermann (1961), Wason and Jones (1963), Gough (1965, 1966), Slobin (1966), Jones (1966a,b, 1968), Wales and Grieve (1969), Greene (1971a,b), and Just and Carpenter (1971). Indeed, part of the comparison process to be proposed has been informally outlined by Wason, Gough, and Slobin. More recently, Clark (1970; in press) and Trabasso (1970; Trabasso, Rollins, & Shaughnessy, 1971) have independently formalized almost the identical general model for the comprehension of negation and have shown how it accounts for most of the previous results on negation. This general theory will be found embedded within the present theory.

The Sentence-First Model: Model A

The first model to be discussed, Model A, applies only when the *S* reads the sentence before he looks at the picture. In this model, the task is formally divided up into four relatively identifiable stages: at Stage 1, the S forms a mental representation of the sentence; at Stage 2, he forms a mental representation of the picture; at Stage 3, he compares the two representations; and at Stage 4, he produces a response. These four stages will be discussed in turn.

Stage 1: Mental Representation of Sentences

At Stage 1, the S must represent the semantic or ideational content of the sentence—its interpretation—in a propositional form. An assumption that meets this requirement is the deep structure assumption (cf. Clark, 1969, in press), which is that sentences are represented in terms of their underlying propositions; it is consistent with recent research (cf. Clark, in press) which shows that it is these propositions that are put to use in sentence verification and reasoning tasks. By this assumption, the sentence *A is above B* would be represented as a single proposition denoted by $(A\ above\ B)_{\mathrm{Sen}}$. The negative *A isn't above B* would consist of the two propositions (*A above B*) and (*it is false*), with the first proposition embedded within the second, as in *that A is above B is false*, or *it is false that A is above B*.[2] For convenience, this embedding will be denoted by $(false\ (A\ above\ B))_{\mathrm{Sen}}$. Similarly, *A is below B* and *A isn't below B* would be represented, respectively, as $(A\ below\ B)_{\mathrm{Sen}}$ and $(false\ (A\ below\ B))_{\mathrm{Sen}}$.

As they stand now, however, these representations fail to capture the fact that *above* and *below* are antonyms and not two arbitrarily related prepositions. A more adequate coding is suggested by the following two linguistic facts about English. First, even though (1) *A is above B* and (2) *B is below A* both refer to the same physical situation, (1) describes the position of A with respect to a "point of reference" at B, whereas (2) does just the reverse (Clark, 1971, in press). Seconds, 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 relations are *high, low, tall,* and *short.*[3] It is well known that *high* and *low* semantically presuppose the dimension of *height* (not *lowness*), and they mean roughly "of much height" and "of little height," respectively (Vendler, 1967; Clark, 1969; Givon, 1970). But height is always measured upward from a point of reference at the bottom, no matter whether much or little height is being measured; *height* means "distance upward," never "distance downward." The analogous statements hold for *tall, short,* and their underlying dimension of *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.

Considered together, these two facts suggest that *A is above B* is the normal or neutral description of an A above a B because the point of reference B is at the bottom of the described dimension in exact agreement with the presuppositions of English adjectives; similarly, *B is below A* is the abnormal or semantically marked description because its point of reference A does not have this property. The coding scheme chosen for *above* and *below* should reflect these semantic facts. The antonymy of *above* and *below* is taken care of in the representations [+Verticality [+Polar]] and [+Verticality [−Polar]], respectively, in which the feature +Verticality stands for all the verticality relations that *above* and *below* have in common and ±Polar indicates the polarity of the comparison. The markedness of −Polar, however, is more difficult to take care of. Perhaps −Polar demands an extra step in coding because it is the abnormal case; or perhaps it consists of two parts, one indicating the normal polarity and the second indicating the negative of it. In any case, *above* will be considered to have a simpler coding than *below*. With these considerations in mind, we will nevertheless continue to use the notation $(A\ above\ B)_{\mathrm{Sen}}$ and $(B\ below\ A)_{\mathrm{Sen}}$ for brevity and convenience.

These linguistic considerations lead to several predictions about the latency differences resulting from Stage 1. We assume that different representations are constructed from their respective surface structures with different speeds. First, the coding speed for *above* and *below* are presumably different, and since *above* appears to be less complex linguistically, we assume that it is coded faster than *below*. Second, the coding speeds for positive and negative representations are also presumably different, and since the positive is the less complex code, we assume that positives are encoded faster than negatives. For conveni-

---

[2] We assume here that positive and negative sentences are embedded in some sort of additional implicit performative (Ross, 1970), giving expressions like *I hereby say to you that it is false that A is above B.* This performative implicitly asserts the truth of the sentence it embeds. Such an assumption, for example, makes it unnecessary to embed the positive sentence in the proposition (*it is true*). If there were such a proposition, then it would have to be attached to negative sentences too since they are also true, and this would lead to the absurd result that sentences could have an indefinite number of such embedding strings—*it is true that it is true that it is false that A is above B,* etc.

[3] *Deep* and *shallow* do not belong to this set because they are more general in meaning, referring to "distance into an enclosed space from its surface" (Bierwisch, 1967; Clark, 1971).

ence, the *above–below* and positive–negative differences will be denoted by parameters a and b, respectively. Finally, we make the strong assumption that a and b are additive. Thus, whereas a sentence with *below* takes a longer than one without, and a sentence with a negative takes b longer than one without, a sentence with both *below* and a negative takes $a + b$ longer than one without either. These assumptions are all empirical claims that will be tested in the following experiments.

Stage 2: Mental Representations of Pictures

While the subject is holding the sentence representation in temporary memory, he must encode the picture in the same general format. Because of this requirement, we assume that he encodes the picture $\overset{\mathrm{A}}{\mathrm{B}}$ either as $(A\ \mathit{above}\ B)_{\mathrm{Pic}}$ or as $(B\ \mathit{below}\ A)_{\mathrm{Pic}}$, depending on whether he is encoding the position of A with respect to B, or vice versa. In anticipation of the results, Model A was formulated such that the subject will represent this picture as $(A\ \mathit{above}\ B)_{\mathrm{Pic}}$ whenever the sentence he has just read contains *above*, but as $(B\ \mathit{below}\ A)_{\mathrm{Pic}}$ whenever the sentence he has just read contains *below*. As for latency predictions, one might assume *a priori* that the *above* representation should be faster to construct than the *below* representation, just as in Stage 1. But since the two codes have been found empirically to take approximately equal amounts of time to construct (cf. Experiment II and Clark & Chase, in preparation), Model A makes no provision for the different encoding latencies of *above* and *below* at Stage 2.

Stage 3: Comparison of Sentence and Picture Representations

At Stage 3, the subject must compare the representations produced by Stages 1 and 2 to see if they "mean" the same or not. This comparison is assumed to be based on the "principle of congruence" (Clark, 1969), which states that two underlying representations can be compared only for identity and that there must be extra operations in case there is a mismatch. The simplest comparison process would, therefore, be to compare the sentence and picture representations for overall identity, and to say "true" if there is identity and "false" if there is not. This, however, will not do, because, for example, $(\mathit{false}\ (B\ \mathit{above}\ A))_{\mathrm{Sen}}$ is actually true of $(A\ \mathit{above}\ B)_{\mathrm{Pic}}$ even though they are not identical. For this reason, Stage 3 must be endowed with a series of comparison operations, each checking for the identity of subparts of the two representations, and each adding to the computation of the answer *true* or *false*. Although there are many *a priori* models that fulfill these requirements, the previous results on negation and the results about to be presented favor one model over the rest. To save time, we will present the one model plus its justification.

In Model A, the goal of the Stage 3 comparison operations is to keep track of a truth index, whose only values are *true* and *false*. As an indication of whether the sentence is true or false of the picture, this index is initially set at *true*, under the supposition that the sentence is true unless there is evidence to the contrary, and then successive operations change the *true* to *false*, and *false* back to *true* again, whenever a mismatch is found. The final value of the truth index is then passed on to Stage 4, which produces a response that corresponds to that value.

In its most succinct form, the comparison process is given by two ordered production rules:

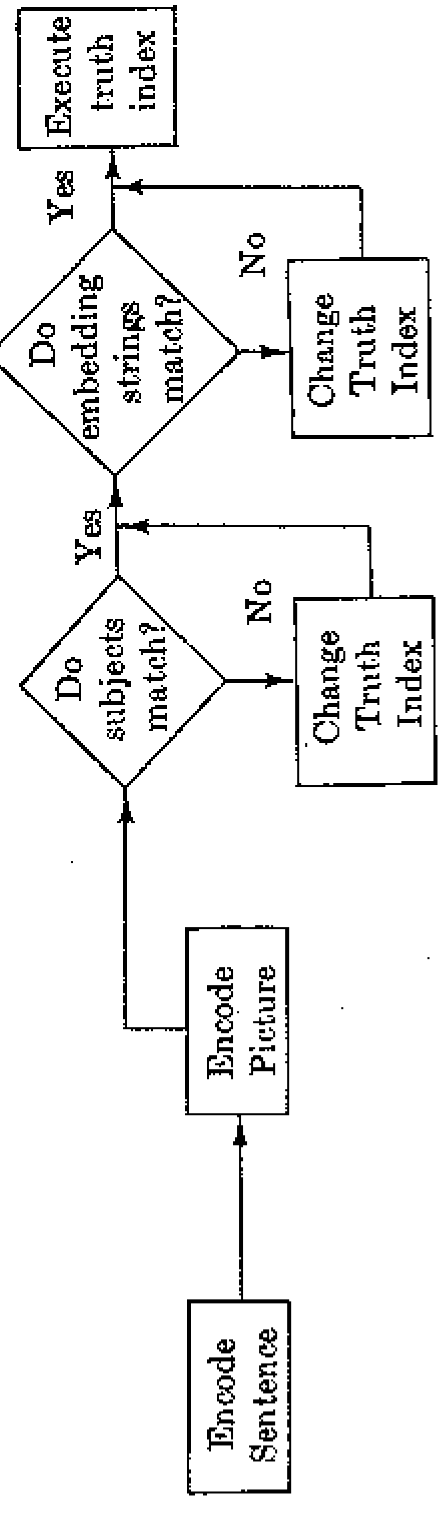
(1) If the embedded strings do not match, change the truth index.
(2) If the embedding strings do not match, change the truth index.

These are illustrated in the flow diagram of Table 1, which also lists the eight basic types of sentences, their representations at Stage 1, the representation of the picture of an A above a B at Stage 2, the changes their truth indices undergo as a result of the comparison operations, and the final values of their truth indices. The flow diagram shows the two production rules expanded into two *pairs* of operations. Operation 1 compares the first noun of the embedded strings of the sentence and picture representations. If they are identical, the process goes on to Operation 2; if not, it goes on to Operation 1a. Operation 1a changes the current value of the truth index into its opposite; then the process goes on to Operation 2. The mismatch of Operation 1 and the translation of Operation 1a are together assumed to consume a fixed amount of time c. Then Operation 2 tests for the identity of the embedding strings of the sentence and picture representations. If neither contains a negative, then the process goes on to Stage 4 directly; but if the sentence contains *false* and the picture does not, the process goes on to Operation 2a. Operation 2a changes the current value of the truth index; then the process goes on to Stage 4. The second mismatch and translation together are assumed to take a fixed amount of time d. It is further assumed that these operations are carried out one after the other so that the increments are additive.

It is helpful to work through a concrete example in Table 1—the true negative sentence *B isn't above A*. With the sentence represented as $(\mathit{false}\ (B\ \mathit{above}\ A))_{\mathrm{Sen}}$ and the picture as $(A\ \mathit{above}\ B)_{\mathrm{Pic}}$, Operation 1 would compare the *B* of the former against the *A* of the latter and find a mismatch, so Operation 1a would change the truth index from its pre-

TABLE 1

Model A and Its Consequences for Eight Types of Sentences



Sentence type			Sentence code	Picture code			
Positive	True	above	(A above B)	(A above B)			T
		below	(B below A)	(B below A)			T
	False	above	(B above A)	(A above B)	true → false		F
		below	(A below B)	(B below A)	true → false		F
Negative	True	above	(false (B above A))	(A above B)	true → false	false → true	T
		below	(false (A below B))	(B below A)	true → false	false → true	T
	False	above	(false (A above B))	(A above B)		true → false	F
		below	(false (B below A))	(B below A)		true → false	F

supposed value of *true* to *false*. Operation 2 would then compare the embedding strings and find a mismatch between $(\textit{false}\ (\ ))_{\mathrm{Sen}}$ and $((\ ))_{\mathrm{Pic}}$—the equivalent of $(\mathrm{true}\ (\ ))_{\mathrm{Pic}}$—, so Operation 2a would change the truth index, now at *false*, back to *true* again. Together, the two mismatch and translation operations would consume an amount of time given by $(c + d)$. The value of the truth index, *true*, would be passed on to Stage 4.

Stage 4: Response Production

The formal outcome of the Stage 3 comparison operations is only an index indicating whether the sentence is true or false of the picture. The duty of Stage 4 is to translate this outcome—*true* or *false*—into a response, which in the present experiments was always a push of an appropriate button. Although processes at this stage are far from trivial (Brooks, 1968; Fitts & Posner, 1967; Smith, 1968; Sternberg, 1969), we will assume that they are subsequent to, and separable from, Stages 1 through 3. Hence, processing times at Stage 4 are assumed to be additive to the components of the model, contributing only a constant time to the whole process. Although it is important to keep this stage separate, little more will be said about it.

Summary of the Latency Predictions

Model A is capable of predicting the time it takes subjects to verify sentences only because of the assumption that the times for the separate processes are additive in the sense of Sternberg (1969). Under this assumption, the claim of Model A is that the total verification latency for each condition shown in Table 2 consists of the addition of one or more of five parameters. First of all, Stage 1 adds a amount of time to sentences that contain *below* and b amount of time to those that are negative. Second, Stage 3 adds c amount of time to the sentences that require Operation 1a (because of a mismatch on their embedded strings), and d amount of time to those that require Operation 2a (because of a mismatch on their embedding strings). Finally, the time not accounted for by these parameters is thrown into a wastebasket parameter t_0. The constituency of the latencies for the eight conditions is shown in Table 2. It should be noted, however, that b and d are confounded, since both parameters occur only with negative sentences. They, therefore, need to be handled together as a single parameter $(b + d)$ in the present experiments, although they can be teased apart experimentally (cf. Chase, Young, Singer & Clark, 1971; Young & Chase, 1971; and below). In summary, Model A predicts that the verification latencies can be accounted for by four parameters: a, hereafter called Below Time; $(b + d)$, here-

TABLE 2

Breakdown of Latencies Predicted by Model A for Eight Types of Sentences

Sentence type			Sentence	Latency components	Latencies		Percent error
					Observed	Predicted	
Positive	True	above	A is above B.	t_0	1744	1763	6.7
		below	B is below A.	$t_0 + a$	1875	1856	7.9
	False	above	B is above A.	$t_0 + c$	1959	1950	8.8
		below	A is below B.	$t_0 + a + c$	2035	2043	7.5
Negative	True	above	B isn't above A.	$t_0 + b + c + d$	2624	2635	12.5
		below	A isn't below B.	$t_0 + a + b + c + d$	2739	2728	12.5
	False	above	A isn't above B.	$t_0 + b + d$	2470	2448	7.1
		below	B isn't below A.	$t_0 + a + b + d$	2520	2541	14.6

after called Negation Time, since it estimates how much longer negatives will take than affirmatives; *c*, hereafter called Falsification Time, since it estimates how long it takes to discover a mismatch of the embedded strings; and t_0, hereafter called the Base Time. Experiment I was designed to test these predictions.

EXPERIMENT I

Method

The *S*s in Experiment I were shown displays consisting of a sentence on the left and a picture on the right and were required to push one of two buttons as quickly as possible to indicate whether the sentence was true or false of the picture. The eight sentences were *star is above plus, star is below plus, start isn't above plus, star isn't below plus,* and the four sentences with *star* and *plus* interchanged. Each sentence was paired with one of two pictures: either an asterisk directly above a plus, $\overset{*}{+}$, or a plus above an asterisk, $\overset{+}{*}$. The pairings of sentences with pictures yielded 16 different displays. The sentence and picture of each display were typed in lower case elite type on 5 × 8 index cards and were viewed in a Polymetric tachistoscope from a distance of 18 in. The positive sentence subtended a visual angle of approximately 4.7°, and the negative, 5.4°, and the picture subtended a visual angle of 0.8° vertically. The picture was centered 10° to the right of the beginning of the sentence.

Each *S* was run on 11 blocks of 16 trials, each block consisting of a different random order of the 16 displays. The first block was discarded as practice. On each trial, the *S* was presented with one of the 16 displays and was required to read the sentence, then fixate the picture, and then push either a "true" or a "false" button as quickly as possible; the time interval from presentation to response was measured to the nearest 10 msec. Half the *S*s pushed "true" with their right thumb, and half pushed "true" with their left. The *S* 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 *S* pushed a third button situated midway between the "true" and "false" buttons, and one second later—time enough for the *S* to replace his thumbs in the proper positions—the next display appeared. The intertrial interval was about 10 sec, there were about 30 sec between blocks, and the experiment lasted about 30 min.

The 12 *S*s were Carnegie–Mellon University students fulfilling a course requirement for introductory psychology. They were instructed to carry out the task as quickly as possible while keeping their errors to a minimum.

the strongest statistical support for the present theory lies in the lack of any unpredicted interactions in this experiment.

Sternberg (1969) suggested that models of this type be evaluated by considering the magnitude of the RMSD relative to the smallest parameter of the model.[5] In the present case, the 16-msec RMSD is much smaller than any of the parameters, and, tested against a pooled error term,[6] the F for lack of fit was less than 1.

Finally, an examination of the residuals does reveal some possible sources of lack of fit. If one looks at the residuals closely, it can be seen that most of the lack of fit can be attributed to differences in a for positive vs negative sentences, and for true vs false sentences (although these interaction Fs do not approach significance). There are several possible causes, other than random error, for these residuals, and the best way to trace these causes is to look at the data of individual Ss.

Of the 12 Ss only three showed any significant lack of fit ($p < .05$, see footnote 6), and two other Ss showed nonsignificant but systematic residuals suggestive of lack of fit. Four of the five Ss' deviations could be described as tendencies on some trials to convert negative sentences to positives before the Comparison Stage.[7] This strategy, to be described later, leads to a model which is derivable from the general theory, yet is slightly different from Model A. It should be emphasized, however, that the bulk of the data is still describable by Model A, that even the small deviations are representative of similar additive models, and that even at the level of individual Ss, the theory handles the data well.

To sum up, we conclude that the model receives excellent support in the data. We make three points. First, the parameters a, c, and $(b + d)$ predicted by the model are reliable, even at the level of individual Ss, and, as we shall see later, they agree with estimates from other experiments. Second, the additivity of the model is confirmed by the lack of significant interactions among the parameters. The small nonadditive residuals appear to be due to mixtures of additive processes within a few Ss. Finally, the best way to evaluate the theory is by further experimentation, and that is what the rest of the paper is about.

Errors. The overall error rate (9.7%) was relatively low (see Table 2), and unlike the latencies, error rates in particular conditions are not very reliable. It is enough to emphasize one point about the errors. The error rates listed in Table 2 are correlated ($r = .75$) with the observed latencies. Sentences requiring more mental operations generally elicited more errors. Thus, there was a 1.8% difference in errors corresponding to Below Time a, a 4.0% difference corresponding to Negation Time $(b + d)$, and a 1.2% difference corresponding to Falsification Time c. This correlation rules out the possibility of a serious trade-off between latencies and error rates on any subset of conditions.

The Picture–First Model: Model B

Model A was specifically designed to account for what happens when the S reads the sentence before he looks at the picture to verify it. The generality of Model A would be strengthened considerably if Model A could also be shown to account for what happens when the S reads the sentence only after he looks at the picture. But Model A cannot apply to the picture-first task without modification, because it assumes that the coding of the picture is contingent on the coding of the sentence, an assumption that is extremely implausible when the S codes the picture first. Model B is a modified version of Model A suitable for the picture-first task.

In Model B, shown in Table 3, the coding of the picture (now Stage 1) occurs before the coding of the sentence (now Stage 2). Since the coding of the sentence at Stage 2 must be independent of the coding of the picture at Stage 2, it is simplest to assume that the picture is coded in the same way on every trial. For the linguistic reasons noted in conjunction with Model A, the picture is, therefore, assumed to be invari-

---

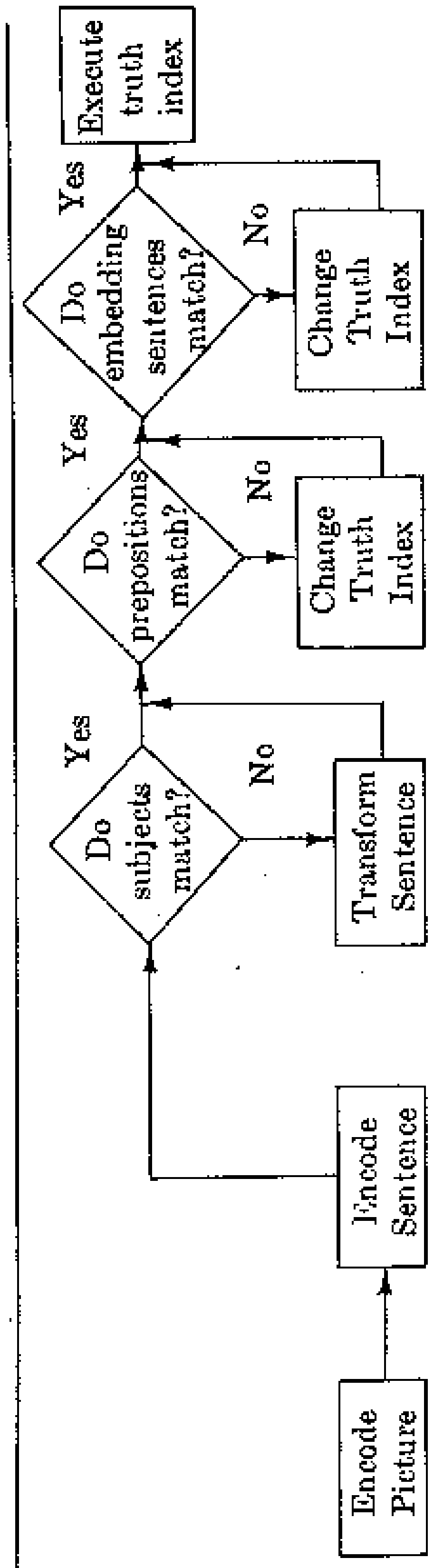
[5] This can be misleading because the magnitude of the RMSD depends upon how many observations are in the means. If the theory is true, the RMSD can be made arbitrarily small by taking many observations. For example, the RMSD's for individual Ss averaged 84 msec because there is 1/12 as much data per point, and the best S had an RMSD of 24 msec. On the other hand, with more motivated and practiced Ss, and twice as many trials, the RMSD for an individual S on Model A was less than 10 msec (Young & Chase, 1971).

[6] There are problems in trying to estimate the underlying "pure error." Just pooling the error terms will not do because there are real effects, unrelated to the model, which might inflate the error term and thus hide any lack of fit. What we did was eliminate all the $S \times$ Parameter interactions from the pooled-error term, and we got an estimate of 176 msec. Even so, this probably underestimates the underlying error because it is based on means of 10 observations. Another estimate is that from individual Ss across trials, removing all the Trial × Parameter interactions. This estimate, based on about 1000 *df*, was 446 msec (220 msec for the best S). Of course, the test for lack of fit would be smaller with this error term.

[7] The data from one S can be described as a "conversion" strategy where he converts the negative sentences to positives before the Comparison Stage by, for example, *A isn't below* → *B is below.* The data of three other Ss can be described as mixtures of Model A with a similar conversion strategy: *A isn't below* → *A is above.* The fifth S had a significant deviation from Model A due to extra long latencies for the true-negative sentences. These conversion strategies have been systematically studied by Young and Chase (1971), and they showed that these strategies are handled quite well by an additive model, similar to Model A, derived from the general theory.

TABLE 3

Model B and Its Consequences for Eight Types of Sentences



Sentence type			Picture code	Sentence code				
Positive	True	above	(A above B)	(A above B)				T
		below	(A above B)	(B below A)	→ (A above B)			T
	False	above	(A above B)	(B above A)	→ (A below B)	true → false		F
		below	(A above B)	(A below B)		true → false		F
Negative	True	above	(A above B)	(false (B above A))	→ (false (A below B))	true → false	false → true	T
		below	(A above B)	(false (A below B))		true → false	false → true	T
	False	above	(A above B)	(false (A above B))			true → false	F
		below	(A above B)	(false (B below A))	→ (false (A above B))		true → false	F

ably coded as $(A \textit{ above } B)_{\mathrm{Pic}}$, the normal, neutral representation for an A above a B. At Stage 2, Model B assumes that the four types of sentences are represented in exactly the same form as they were in Model A.

It is at Stage 3 that complications arise. In Model A, the comparison operations were given by two production rules:

(1) If the embedded strings do not match, change the truth index.
(2) If the embedding strings do not match, change the truth index.

In Model A, there was a simple test to see whether the embedded strings matched. Since Stages 1 and 2 guaranteed that the two prepositions would always be identical, the test needed only to compare the subject nouns of the two strings. In Model B, however, such a simple test is not possible since the two prepositions are not guaranteed to be identical. The test in Model B, for example, must show that $(A \textit{ above } B)_{\mathrm{Pic}}$ and $(B \textit{ below } A)_{\mathrm{Sen}}$ "match" (or are synonymous) even though the two strings are not identical. To accomplish this, apparently, Model B requires two production rules in place of (1). A pair of such rules is illustrated as (1′) and (1″) in the following revised model:

(1′) If the subjects of the embedded strings do not match, translate $(X \textit{ above } Y)_{\mathrm{Sen}}$ into $(Y \textit{ below } X)_{\mathrm{Sen}}$, or vice versa.
(1″) If the prepositions of the embedded strings do not match, change the truth index.
(2) If the embedding strings do not match, change the truth index.

An equally plausible sequence would be to compare the prepositions in (1′) and the subjects in (1″). In anticipation of the results of Experiment II, however, we will present the details of the given sequence only, since it predicts the mean latencies correctly, whereas alternative sequences do not.

Operations 1 and 1a of Model A are, therefore, replaced by Operations 1′, 1′a, 1″, and 1″a in Model B, as shown in Table 3. Operation 1′ first tests for the identity of the subjects of the two representations. If they match, the process goes on to Operation 1″; if not, it goes on to Operation 1′a. Operation 1′a must now manipulate the representation of the sentence so that it matches in at least one respect with the picture representation. Operation 1′a does this by transforming $(X \textit{ above } Y)_{\mathrm{Sen}}$ into $(Y \textit{ below } X)_{\mathrm{Sen}}$ or vice versa, whichever is needed; although only the former transformation is made use of in Experiment II, both will be required in Experiment III. This mismatch and translation operation is assumed to take a fixed amount of time e. Then Operation 1″ checks for the identity of the prepositions of the two embedded strings: if they match, the process goes on to Operation 2; and if they do not, the

TABLE 4
Breakdown of Latencies Predicted by Model B for Eight Types of Sentences

Sentence type			Sentence	Latency components
Positive	True	above	A is above B.	t_1
		below	B is below A.	$t_1 + a \quad + e$
	False	above	B is above A.	$t_1 \quad + e + f$
		below	A is below B.	$t_1 + a \quad + f$
Negative	True	above	B isn't above A.	$t_1 \quad + b + e + f + d$
		below	A isn't below B.	$t_1 + a + b \quad + f + d$
	False	above	A isn't above B.	$t_1 \quad + b \quad + d$
		below	B isn't below A.	$t_1 + a + b + e \quad + d$

process goes on to Operation 1″a. Operation 1″a changes the value of the truth index into its opposite and then passes the process on to Operation 2. This mismatch and translation operation is assumed to take a fixed amount of time *f*. Operations 2 and 2a are the same as in Model A. Operations 1′, 1′a, and 1″, therefore, jointly check the synonymy of the two embedded strings, and Operation 1″a changes the truth index only when the two strings are not synonymous.

Model B uses five parameters to predict the latencies in Experiment II: Below Time *a*, Negation Time $(b + d)$, Subject Mismatch Time *e*, Falsification Time *f*, and Base Time t_1. One could argue, however, that Operation 1″a in Model B is just the same as Operation 1a in Model A, and so *f* should actually be set equal to *c*. But a closer look at Operations 1a (of Model A) and 1″a (of Model B) shows that one is triggered by a mismatch of nouns, and the other by a mismatch of prepositions. Thus, there is no reason to think that *f* should be exactly equal to *c*. Nevertheless, in testing Models A and B jointly in Experiment II, it will be assumed that *f* and *c are* equal. As it turns out, this is approximately correct. The predictions of Model B in terms of these parameters are shown in Table 4.

EXPERIMENT II

Method

The displays and general procedure in Experiment II were the same as in Experiment I, except that now Ss were asked in half the conditions to view the picture first and then the sentence. For this purpose, a second deck of 16 displays was prepared with the picture on the left and the sentence on the right. In addition to an initial practice block, every S was run on 12 blocks (of 16 trials). The 12 blocks were divided into four sequences of three blocks each. The four sequences differed in what the S was instructed to fixate first: the sentence which was on the left, the sentence which was on the right, the picture which was on the left, or the picture which was on the right. The order of these four sequences was counterbalanced across the 24 *Ss*, all of whom were students at Carnegie–Mellon University who had not served in previous experiments.

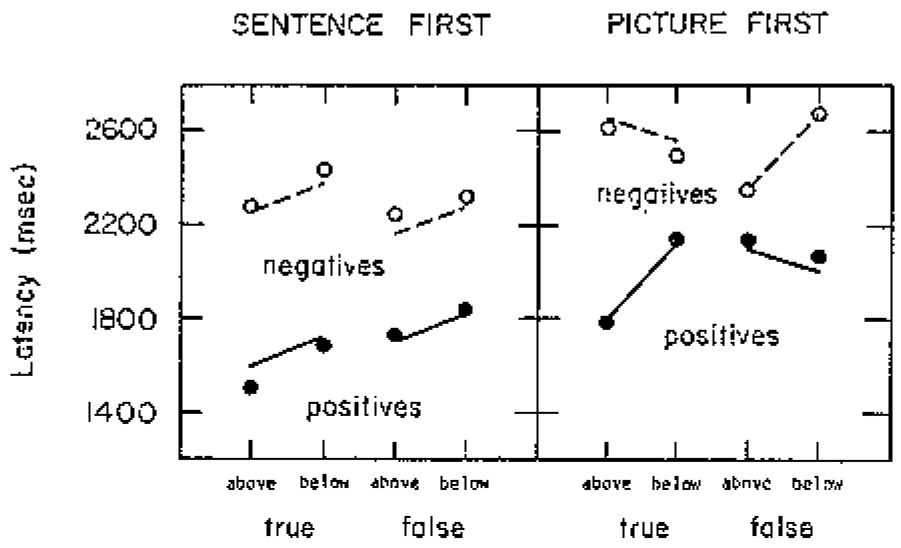


FIG. 2. Observed mean latencies (points) and predicted mean latencies (lines) for sentence-first and picture-first conditions of Experiment II. Positive sentences are represented by solid circles and solid lines (———), and negative sentences, by open circles and broken lines (- - - -).

Results

Latencies. The latency data from Experiment II strongly support the joint validity of Models A and B. The observed and predicted mean latencies for the eight main sentence-first and picture-first conditions are shown in Fig. 2 and Table 5. The patterns of latencies in the sentence-first and picture-first conditions obviously differ from each other. The sentence-first pattern is very similar to the pattern found in Experiment I, as it should be, and is rather well accounted for by Model A, whereas the picture-first pattern requires another model, specifically Model B. Since Models A and B have three parameters in common, they were treated jointly, deriving the least-squares estimates of the six parameters Below Time *a*, Negation Time $(b + d)$, Falsification Time *c*, Subject Match Time *e*, and Base Times t_0 and t_1. The overall estimates are shown in Table 5, with the estimates of the first three parameters shown separately for the sentence-first and picture-first conditions. The four parameters *a*, $(b + d)$, *c*, and *e* accounted for 97.5% of the variance among the 16 means (14 *df*). The RMSD of 47.5 msec with 10 *df* was rather large, and the 2.5% residual variance was highly significant,[3] $p < .001$.

---

[3] This was tested against a pooled error term representing a standard deviation of 272 msec.

TABLE 5

Observed and Predicted Latencies, Percentage of Error, and Parameter Estimates from Sentence-First and Picture-First Conditions of Experiment II

Condition			Sentence first			Picture first		
			Observed	Predicted	Percent error	Observed	Predicted	Percent error
Positive	True	above	1500	1603	6.2	1783	1798	4.7
		below	1681	1720	7.8	2139	2122	12.8
	False	above	1728	1701	8.6	2130	2103	6.8
		below	1838	1818	7.0	2077	2008	7.6
Negative	True	above	2269	2257	17.4	2614	2659	19.5
		below	2337	2374	14.3	2499	2564	14.6
	False	above	2246	2159	10.4	2354	2349	11.2
		below	2319	2276	13.3	2678	2678	16.7

Parameter	Sentence first	Picture first	Overall
a	106	128	117
$(b + d)$	608	504	556
c	104	91	98
e	—	212	—
t	1603	1793	—

An analysis of variance was performed on the means of the three or fewer correct responses for each display under each viewing condition for each S, a total of 1536 means. According to this analysis, the parameters a, $(b + d)$, c, and e were each significantly greater than zero ($p < .001$) with $F(1,23)$ of 71, 250, 39, and 56, respectively. More importantly, there were no interactions among these parameters except for a minor unpredicted interaction between $(b + d)$ and c; in this interaction, c was 135 msec longer for positive than negative sentences, $F(1,23) = 9$, $p < .01$, an effect either totally absent or insignificant on other experiments of this kind we have run. The significance of the residual variance can be traced to this minor interaction and to the fact that Negation Time $(b + d)$ was 104 msec larger in the sentence-first than in the picture-first condition, $F(1,23) = 8$, $p < .01$, another relatively small deviation from the model. Again, this deviation has not been replicated in later experiments we have run.

There were four other significant results in Experiment II. Sentences with *plus* as subject were verified 51 msec faster than those with *star* as subject, $F(1,23) = 16$, $p < .001$; this result is similar to one in Experiment I. In the picture-first condition, displays with the plus (+) above were verified 67 msec faster than displays with the star (*) above, $F(1,23) = 12$, $p < .005$. Also, the Base Time t_0 of the sentence-first condition was 189 msec less than the t_1 of the picture-first condition, $F(1,23) = 31$, $p < .001$. These three results are again important for the additivity of the model since these factors (all presumably encoding effects) did not interact with the effects of the comparison stage. Finally, the *above-below* difference was 47 msec larger when the sentence was on the right than on the left (93 vs 140 msec), $F(1,23) = 9$, $p < .01$; this last effect can only be assumed to be spurious.

As shown in Table 5, the overall error rate in Experiment II (11.2%) was slightly higher than in Experiment I. Here again, however, there was a high correlation ($r = .84$) between the error rates and latencies of the 16 main conditions. As in Experiment I, conditions requiring more mental operations generally elicited more errors. The differences in error rates corresponding to the separate parameters were as follows: a 1.2%, $(b + d)$ 7.0%, c 1.6%, e 4.4%, and $(t_1 - t_0)$ 1.1%, all of which were positive.

Discussion

Experiment II gives moderately strong confirmation to Models A and B together. The parameters of the models jointly accounted for most of the variance in the mean latencies, and there were only minor deviations from the model. Thus, the two models—which are actually two realizations of the same theory—present a relatively unified picture of how people compare sentences and pictures in two quite different situations.

The Ss questioned about Experiment II all noted that the picture-first conditions seemed much harder than the sentence-first conditions. Indeed, the picture-first conditions took 296 msec longer, on the average, than the sentence-first conditions. This difference is important because it can be predicted by Models A and B with the addition of several simple assumptions. Recall that Model B contains two pairs of operations ($1' - 1'a$ and $1'' - 1''a$) where Model A has only one ($1 - 1a$). Then consider: (1) the assumption made in Experiment II that $1''a$ and $1a$ take the same amount of time c (98 msec); (2) the much more dubious assumption that $1'$ and 1, both subject comparison operations, take the same (undeterminable) amount of time; and (3) the assumption that the Subject Mismatch Time e (212 msec) estimates the time consumed solely by $1'a$. From these assumptions it follows that the Base Time t_1 for the picture-first conditions should be greater than Base Time t_0 for the sentence-first conditions by the amount of time consumed by the extra operation ($1''$) required in the picture-first conditions. Thus, one could estimate the time consumed by Operation 1″ as 190 msec. It hardly needs to be said that the exact assumptions made here are rather dubious. Yet the analysis seems correct in spirit, and it shows that the difficulty of the picture-first conditions is consistent with the two extra comparison operations required in Model B.

EXPERIMENT III

In Experiment I, *Ss* reported informally that they attended to the top or bottom of the picture depending on whether the sentence they had just read contained *above* or *below*. It was partly on this evidence that coding the picture was assumed to be contingent on coding the sentence in the manner indicated in Model A. If the *S* looked at the top of the picture, he was assumed to have coded the picture as $(A \textit{ above } B)_{\mathrm{Pic}}$, but if he looked at the bottom, the code was instead $(B \textit{ below } A)_{\mathrm{Pic}}$. One way to test this assumption is simply to direct the *Ss* to attend to the top or bottom of the picture in a picture-first experiment. Model B, with the appropriate picture codes, should be capable of predicting the latencies in such a task.

Method

The displays of Experiment III were the same as in Experiment I. In this case, however, *Ss* were instructed to fixate the picture before the sentence in one of three ways: (1) attend to the figure as a whole; (2) attend only to the top of the figure; and (3) attend only to the bottom of the figure. Each *S* was run for five blocks of trials on the displays under each set of instructions, and the first block of each set was discarded as practice. The order of the three instructions was counterbalanced across the 24 *Ss*, all of whom were students at Carnegie-Mellon University who had not served in previous experiments. In all other respects, Experiment III was identical to Experiment I.

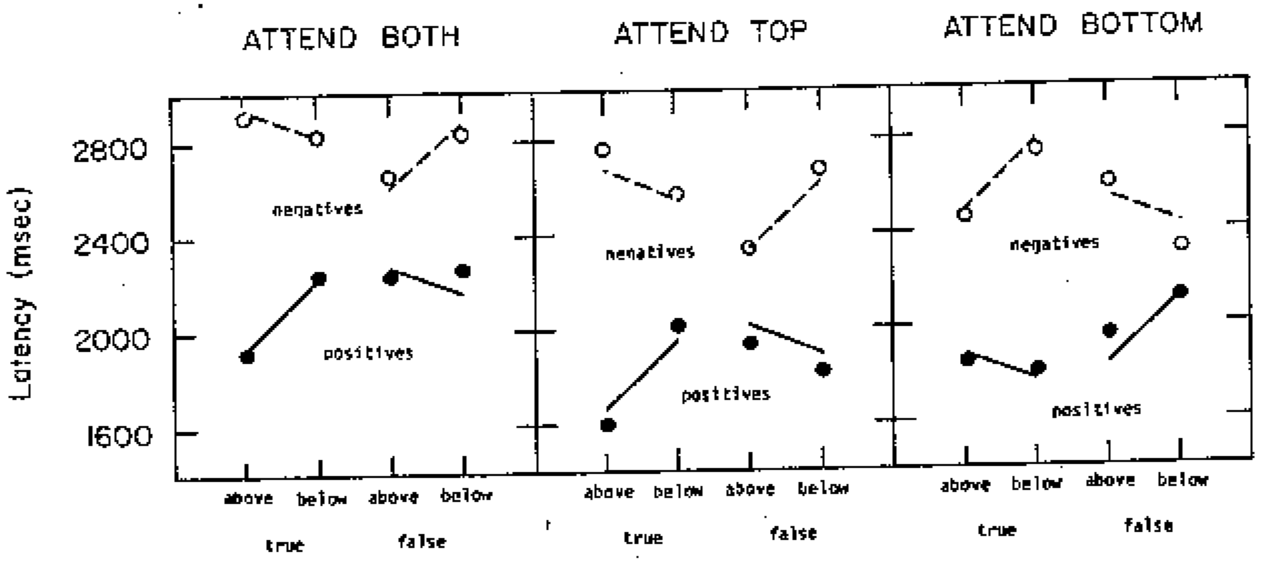


FIG. 3. Observed mean latencies (points) and predicted mean latencies (lines) for the "attend-both," "attend-top," and "attend-bottom" viewing conditions of Experiment III. Positive sentences are represented by solid circles and solid lines (———), and negative sentences, by open circles and broken lines (- - - -).

Results

Figure 3 and Table 6 show the observed and predicted latencies and percentage errors for the eight main sentence conditions under each of the three instructions. In order to predict these latencies with Model B, it is assumed that the picture is always encoded as $(A \textit{ above } B)_{\mathrm{Pic}}$ whenever either the whole or the top of the picture is attended to, but as $(B \textit{ below } A)_{\mathrm{Pic}}$ whenever the bottom is attended to. Under these assumptions, Model B fits the data quite closely, as shown in Figure 3 and Table 6. The four parameters a, $(b + d)$, e, and f accounted for 97.4% of the variance among the 24 means (22 *df*) shown in Figure 5 and Table 6. The RMSD was 60.7 msec with 18 *df* and the residual 2.6% of the variance did not approach significance.[9]

From the analysis of variance, the parameters a, $(b + d)$, e, and f were each significantly greater than zero (at least $p < .005$) with $F(1,23)$ of 12, 117, 51, and 23, respectively, and the parameters did not differ across conditions and did not interact with other factors in the experiment, just as Model B predicts. The Base Times (t_0, t_1), however, did differ across the three viewing conditions, $F(2,46) = 4.45$, $p < .025$; 99.6% of this variance was due to a higher Base Time for the attend-both condition than the other two conditions, $F(1,46) = 8.87$, $p < .005$. For this reason, two Base Times were used to fit the model—a t_1 of 1938 for the attend-both condition and a t_1 of 1675 for the other two conditions—leaving 22 *df* to fit the model.

There were two other significant effects in Experiment III. First, the pictures in which the plus (+) was attended to were verified 60 msec faster than those in which the star (*) was attended to, $F(1,46) = 10.3$, $p < .005$. And second, the sentences with *plus* as subject were verified 80 msec faster than those with *star* as subject, $F(1,23) = 12.6$, $p < .005$. Both of these differences are in complete agreement with Model B, since they reflect encoding times of the picture and sentence, respectively, and should be statistically independent of each other and the other parameters.

The errors in Experiment III had the same properties as they did in Experiments I and II. The error rates and the observed latencies in

---

[9] This was tested against a pooled error term representing a standard deviation of 343 msec.

TABLE 6

Observed and Predicted Latencies, Percentage of Error, and Parameter Estimates from Experiment III

Condition			Both			Top			Bottom		
			Observed	Predicted	Percent error	Observed	Predicted	Percent error	Observed	Predicted	Percent error
Positive	True	above	1911	1938	3.8	1596	1675	1.0	1841	1871	4.3
		below	2238	2218	11.5	1014	1955	12.0	1808	1758	4.8
	False	above	2231	2279	6.8	1927	2016	9.5	1946	1820	5.8
		below	2255	2167	8.8	1816	1903	6.8	2115	2100	6.3
Negative	True	above	2906	2939	14.0	2758	2676	14.0	2459	2480	19.8
		below	2823	2827	16.3	2567	2564	21.3	2726	2760	15.0
	False	above	2652	2598	5.8	2337	2335	4.3	2592	2531	15.8
		below	2827	2878	10.5	2662	2615	14.0	2311	2419	11.5

Parameter	Both	Top	Bottom	Overall
a	111	110	30	84
$(b + d)$	643	743	594	660
e	140	261	187	196
f	147	115	173	145
t	1938	1675	1675	—

Table 6 were again highly correlated ($r = .70$), and the differences in the percentage errors corresponding to the separate parameters were as follows: a 2.8%, $(b + d)$ 7.6%, e 2.0%, and f 3.8%, again all positive differences.

Discussion

The significance of Experiment III is that it ties together two assumptions critical to Models A and B. The first assumption was that Ss in Experiment I coded the pictures as $(A\ \textit{above}\ B)_{\text{Pic}}$ or $(B\ \textit{below}\ A)_{\text{Pic}}$, depending on whether the sentence had contained *above* or *below*, respectively. This assumption was plausible since Ss reported attending to the top or bottom figure depending on whether the sentence contained *above* or *below*, respectively. All that remained to be shown was that when Ss were *instructed* to "attend" to the top or bottom figure, their picture codes actually were $(A\ \textit{above}\ B)_{\text{Pic}}$ or $(B\ \textit{below}\ A)_{\text{Pic}}$, respectively. The results of Experiment III are completely consistent with this assumption. A secondary, but important result is that it took Experiment III Ss no longer to code the pictures under the "attend top" instruction than under the "attend bottom" instruction. This confirms the assumption made in Model A that $(A\ \textit{above}\ B)_{\text{Pic}}$ takes no longer to construct at Stage 2 than $(B\ \textit{below}\ A)_{\text{Pic}}$.

The second assumption was that when Ss are asked to view an A above a B without any other constraints, they will use the normal or neutral coding $(A\ \textit{above}\ B)_{\text{Pic}}$. This assumption was made in the picture-first condition of Experiment II and was required for Model B to fit the results. This assumption, too, was confirmed in Experiment III, where it was explicitly shown that viewing the picture as a whole and viewing the top of the picture both gave the same pattern of verification latencies as in the picture-first condition of Experiment II. Furthermore, the pattern of latencies changed in a predictable way when Ss were instructed to attend to the bottom figure in the picture. For some reason, Experiment III Ss took over 200 msec longer to code pictures in the attend-both condition than in the attend-top condition. Perhaps this is because Ss felt compelled in the attend-both condition to encode both figures—as $(A\ \textit{above}\ B)_{\text{Pic}}$—whereas in the attend-top condition they needed to encode only the attended-to figure—simply as $(A\ \textit{above})_{\text{Pic}}$. This difference in coding, of course, would have no consequences for the following comparison processes.

ALTERNATIVE EXPLANATIONS

Although Experiments I, II, and III have so far been used only as support for Models A and B and their interlocking assumptions, they can also be used to rule out certain alternative explanations for the present experiments. Some of these alternatives are full-fledged models, but others are meant to account for only restricted aspects of the data. In this section, we will present five such alternatives and examine the difficulties they encounter in the present data. Although these alternatives can be evaluated only with respect to the present data, this examination may reveal some of the more general shortcomings of these classes of explanations.

The Visual Imagery Models

In the spirit of recent suggestions by Paivio (1969), Seymour (1969), and others, the first model to be examined is one in which the Ss construct a single picture-like mental image (perhaps only schematic in detail) from the given sentence and then compare this image against the given picture for a match; if the image and picture match, then the sentence is true, but if they do not, then the sentence is false. Although Seymour's particular version of the imagery model has already been shown to be incorrect (Chase & Clark, 1971), this model can more generally be shown to fail both logically and empirically for the present type of task.

The model runs into logical difficulties because it is impossible, in general, to form a single visual image, or image surrogate, from negative sentences. The sentence *A isn't above B*, for example, is consistent with many different visual configurations—of A beside B, of A in front of B, of A in back of B, of A below B, and so on. Negative sentences indicate what a picture is not, and only rarely does this uniquely define what the picture is. The imagery model, however, quite properly assumes that each sentence is represented as a single image, not as a broad mélange of incompatible images, and, therefore, the model cannot accommodate negative sentences for which there is no unique image. Thus, for logical reasons alone, this imagery model is generally unable to account for negatives.

These logical objections, however, do not necessarily apply to the present restricted experiments, because here the Ss could have predicted exactly which picture was true for each negative sentence and could therefore have formed a single image for every sentence. Nevertheless, this imagery model fails to account for the present results because it is unable to explain why positive and negative sentences behave differently with respect to truth and falsity. Consider the four sentences meant to describe the picture $\genfrac{}{}{0pt}{}{\text{A}}{\text{B}}$. In the imagery model, the true sentences *A is above B* and *B isn't above A* would both be represented as the image $\genfrac{}{}{0pt}{}{\text{A}}{\text{B}}$. According to this model, the *S* cannot decide on the truth of each sentence until he has constructed the image and compared it against the picture. Therefore, the original form of the sentence—whether it was positive or negative—cannot enter into the comparison process: if true is faster than false for positive sentences, then true must also be faster than false for negative sentences. This prediction, however, is disconfirmed by the highly significant true–false by positive–negative interaction in Experiments I, II, and III (parameters *c* and *f* in our models). Put more generally, any model in which positive and negative sentences are represented in an indistinguishable form will necessarily fail to account for the present—and previous—positive–negative by true–false interactions.

This imagery model also fails to account for the differential behavior of *above* and *below* in picture-first tasks, and for similar reasons. Note that the true sentences *A is above B* and *B is below A* would both be represented as $\genfrac{}{}{0pt}{}{\text{A}}{\text{B}}$, and the false sentences *B is above A* and *A is below B*, as $\genfrac{}{}{0pt}{}{\text{B}}{\text{A}}$. Since these images do not retain the information that they originated from sentences containing *above* or *below*, there is no possibility that *above* and *below* sentences could behave one way when true and another way when false. But this is just what happened in Experiments II and III, with highly significant interactions between *above–below*, true–false, and negation (parameter *e* in our model). Thus, the imagery model is inconsistent with these facts, too.

The imagery model, therefore, is inconsistent with all the effects in the previous experiments (parameters *c*, *d*, *e*, and *f*) except those represented by parameters *a* and *b*, and even these effects cannot be predicted by the model without extra assumptions about the decoding of *above* versus *below* and positive versus negative sentences. That is, the imagery model cannot account for the presence of the mental operations underlying parameters *c*, *d*, *e*, and *f* in the present experiments. The deviations of the data from this imagery model are highly significant.

Although the imagery model chosen here is probably the one most appropriate to the assumptions psychologists normally make about imagery, it is obviously not the only possibility, since imagery is a particularly flexible concept. At the present, we can only specify several minimum requirements that an imagery model must meet if it is to account for the present data. First, the image must contain, or have associated with it, some indication of negativity. For instance, *A isn't above B* might be represented as $\genfrac{}{}{0pt}{}{\text{A}}{\text{B}}$ plus a (nonimagery) tag indicating that the state of affairs represented by this image is not the case. Second, the image must contain, or have associated with it, some indication of point of reference. This is required to distinguish the image for *A is above B* from that for *B is below A*. Unfortunately, both of these requirements are foreign to the basic notion that mental images are like pictures: pictures do not contain indications of whether they are the case or not, nor do they distinguish between comparison objects and points of reference. Thus, the main challenge to any imagery model is to demonstrate these properties and to distinguish it from the present Models A and B, which represent both negation and point of reference quite directly.

The Conversion Models

The next class of models to be considered, the "conversion" models, all assume that negative sentences can be represented directly in a positive form. In the most obvious such model, the S would (1) read *A isn't above B*, (2) represent it directly as (*A below B*)$_{\text{Sen}}$, having "converted" *isn't above* into *below*, and then (3) attempt to verify the positive representation against the picture. Like the imagery model, however, these models fail both logically and empirically because they represent these models in indistinguishable form. In general, as was pointed out previously, negative sentences (like *A isn't above B*) are not equivalent to any single positive representation (say, *A is below B* or *B is above A*), and, therefore, the conversion models cannot work in the general case. But even in the present experiments, because these models do not distinguish between positive and negative sentences at the comparison stage, they are incapable of predicting the different behavior of positive and negative sentences with respect to truth and falsity. Thus, the "conversion" models cannot account for the presence of mental operations causing parameters c and d in the three experiments.

All this evidence, however, does not say that Ss are incapable of using conversion strategies if they want to. Wason (1961), for one, noted that about half of his Ss reported converting sentences like *Nine isn't odd* into *Nine is even* before trying to decide whether the sentence was true or false. In a recent series of experiments, in fact, Young and Chase (1971) have demonstrated that conversions of this kind are open not only to introspection, but also to instruction. Their technique was to repeat the present Experiment I, but to instruct Ss deliberately to convert one type of sentence into another. The four instructions Young and Chase tried out were: (1) convert *isn't above* into *below* and *isn't below* into *above* wherever either conversion is applicable; (2) convert *X isn't above Y* into *Y is above X* and *X isn't below Y* into *Y is below X* wherever applicable; (3) convert *X is below Y* into *Y is above X* and *X isn't below Y* into *Y isn't above X* wherever applicable; and (4) convert *below* into *isn't above* and *isn't below* into *above* wherever applicable. Under the first instruction, for example, the S would read *Star isn't above plus*, think to himself *Star is below plus*, and then look over at the picture to verify the sentence. These four instructions were compared to a control "no conversion" instruction. Young and Chase found that the verification latencies under each conversion instruction were accounted for completely by Model A with only two additional assumptions: (1) each conversion takes an increment of time (k) to carry out; (2) the sentence finally represented at Stage 1 is the sentence produced by the conversion, and not the sentence as originally constituted. From the Young and Chase study, then, one can be even more confident that the Ss in Experiment I did not follow the strategy of converting negatives into positive codes (with the caveat mentioned in Experiment I), because the pattern of latencies in Experiment I was completely unlike that produced by any of the conversion strategies. Consistent with this conclusion, the Experiment I, II, or III Ss who were questioned, unlike Wason's Ss, did not report using conversion strategies. In sum, the Young and Chase study is an important piece of supporting evidence for Models A and B and their underlying assumptions, and it specifically rules out the conversion models as accounts of the present experiments. For more discussion of conversion models see Clark (1970; in press), Trabasso *et al.* (1971), and Trabasso (in press).

Reading Time

One commonly proposed component of comprehension time is reading time: one sentence appears to take longer than another to comprehend only because it takes longer to read. Logically, however, reading time can only account for latency differences that correlate with surface features of the sentences used. In the present experiments, therefore, it is capable of accounting for Below Time a and the b increment of Negation Time ($b + d$), but not for Falsification Time c or f, or Subject-match time e, since the latter two differences are independent of the displayed sentence. Indeed, we will argue that reading time cannot even account for Below Time or for anything but a minor part of the b increment of Negation Time.

A priori, it seems highly unlikely that *above* and *below* differ in reading time, since they both contain five letters and two syllables. This suspicion is confirmed in the data of Young and Chase (1971). Under one of their conversion instructions, Ss were required to convert *isn't above* into *is below* and *isn't below* into *is above*. Whereas *isn't above* was faster than *isn't below* under the "no conversion" control instructions, *isn't above* was *slower* than *isn't below* under this particular conversion instruction, and by approximately the same amount of time. In other words, constructions finally coded as "above" were always faster than those finally coded as "below," regardless of which word (*above* or *below*) actually appeared in the sentence. So reading time can be eliminated as an explanation of Below Time.

It is far more plausible, however, to suppose that reading time accounts for a significant part of Negation Time, since the negative actually adds an extra syllable to the sentence. To begin with, we will assume that Negation Time consists of three components—reading time r (the extra time consumed in reading *isn't* over *is*), encoding time b, and comparison time d. One approach to the investigation of r is to estimate r independently and see what proportion of Negation Time it accounts for. The crudest, but most direct method of doing this is to time Ss as they repeat a positive or negative sentence over and over again and to calculate the difference in repetition rates for positive and negative sentences. By this method, r is about 25 msec. By another subtractive method, Young and Chase (1971) were able to estimate r independently of b and d at about 90 msec. By either method, r is relatively small compared to a Negation Time of 600 msec.

A more satisfactory approach to the problem, however, is to demonstrate that b and d are real quantities and that r cannot completely replace either b or d. And the evidence appears to show that b and d are real quantities. First, Gough (1966) found in a sentence-verification task that even when 3 sec intervened between the end of the sentence read to the S and the presentation of the picture to be verified, negatives still took about 250 msec longer than affirmatives. Since 3 sec is presumably ample time to have completed the coding of the sentence, the 250 msec is a relatively pure estimate of d. Second, in another timed verification task, Trabasso *et al.* (1971) gave their Ss positive and negative sentences of the same length, e.g., *kov green* (meaning "is green") and *luz green* (meaning "isn't green"). By using a method that separated encoding time from comparison time, they found that negatives took about 250 msec extra in encoding time and about 200 msec extra in comparison time. Taken together, these results imply that the 600-msec Negation Time of the present experiments consists of three parts, and by the roughest of estimates, $(r + b + d)$ might be given as (100 msec + 250 msec + 250 msec). In a word, reading time accounts for none of Below Time and for only a small fraction of Negation Time, leaving Models A and B fundamentally intact.

Frequency of Occurrence

Closely akin to reading time is frequency of occurrence. One could assume that some words (or constructions) are more difficult to understand because they are less frequent in the language (cf. for example, Goldman-Eisler and Cohen, 1970). Since frequencies correspond only to superficial linguistic categories (like words or constructions), frequency is by nature incapable of accounting for anything but Below Time a and the b fraction of Negation Time; in this respect, frequency is just like reading time. Otherwise, frequency can be interpreted as referring to (1) the frequency of the words or constructions themselves, or (2) the frequency of the meanings of the words or constructions. Evidence from the present experiments goes counter to either interpretation.

Under its first interpretation, frequency predicts the faster comprehension of: (1) *above* over *below*, since *above* is more frequent than *below* by a ratio of 941 to 529 in the Lorge magazine count (Thorndike & Lorge, 1944); (2) *star* over *plus*, since *star* is more frequent than *plus* by a ratio of 527 to 70 in the Lorge count; and (3) affirmatives over negatives, since affirmatives occur more frequently than negatives by a factor of about 10 to 1 (Goldman-Eisler & Cohen, 1970). Although the results of Experiments I, II, and III are consistent with predictions (1) and (3), the results of Experiments II and III, in which the encoding of *star* and *plus* was separable from the encoding of the pictures * and +, showed that *star* was significantly *slower* than *plus*, exactly counter to prediction (2). Furthermore, Young and Chase demonstrated that even prediction (1) is not upheld for the first interpretation of frequency; as discussed previously, they showed that the superiority of *above* over *below* did not correspond to the printed words *above* and *below*, but to the interpretations "above" and "below" resulting from any conversion that was made. The first interpretation of the frequency explanation must, therefore, be incorrect.

The second interpretation of frequency also seems untenable because of the *plus-star* evidence just presented, although for a strong disconfirmation one would want to know the relative frequencies of the several meanings of *star* and *plus*. The second interpretation, however, runs into further difficulties in accounting for the experiment that follows.

EXPERIMENT IV

Method

The purpose of Experiment IV was to contrast the verification task we have been using with an almost identical forced-choice identification task in order to show that *above* is not always encoded more quickly than *below*. The four displays used for both tasks contained a "sentence" on the left—either the word *above* alone or the word *below* alone—and a picture on the right—a star (*) either above or below a plus (+); in other respects, the displays were the same as those in Experiment I. The

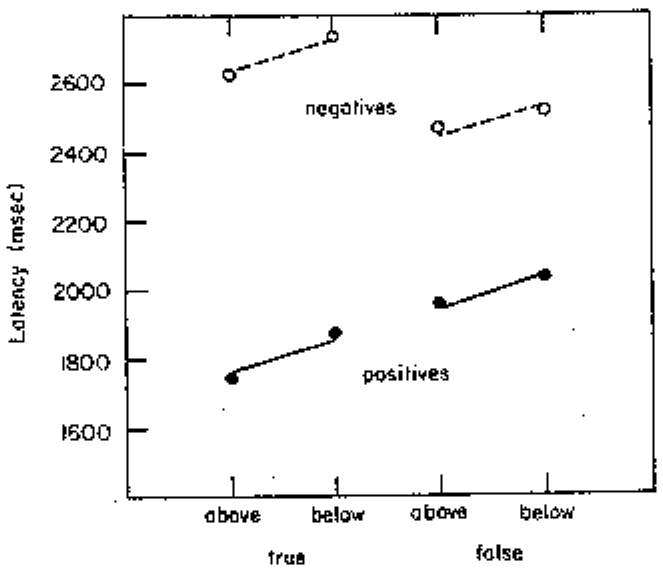


FIG. 1. Observed mean latencies (points) and predicted mean latencies (lines) for Experiment I. Positive sentences are represented by solid circles and solid lines (———), and negative sentences, by open circles and broken lines (- - - -).

Results

Latencies. The latency data of Experiment I strongly support Model A. Table 2 and Fig. 1 show the "observed" and "predicted" mean latencies for the eight main conditions. The "observed" latencies were calculated by averaging only the correct responses in 10 trials for each of the 16 conditions and for each *S*, and then averaging across the 12 *Ss* and across *plus* and *star* as subject of the sentence. The "predicted" latencies were calculated by estimating the four parameters of Model A by the method of least squares and by applying these values to the formulae in Table 2. The estimates were as follows: Below Time a was 93 msec; Negation Time $(b + d)$ was 685 msec; Falsification Time c was 187 msec; and Base Time t_0 was 1763 msec. The three parameters a, $(b + d)$, and c accounted for fully 99.8% of the variance among the eight means shown in Table 2 and Figure 1. The residual 0.2% of the variance represents a Root Mean Squared Deviation (RMSD) of 16 msec, based on 4 *df*.

Evaluation of the model. How do we evaluate the goodness of fit of the model? Two of the more traditional criteria—number of parameters and percentage of variance accounted for—are of little use here. For example, one might argue that the model accounts for so much variance only because it has a large number of parameters, and any *post-hoc* model with this many parameters would do as well. This would be true if there were no underlying theory motivating the derivation of the model, and if this were the only experiment we ever performed to test the model. But the theory states that in order to verify a sentence, both underlying propositions must be checked, so there must be two parameters corresponding to these mental operations, and similarly, there must be a parameter for the semantic difference between *above* and *below*. In short, the theory demands that in order to comprehend and verify the complex sentences under study, there must be at least three parameters corresponding to the underlying mental operations (in addition to the Base Time t_0). It is an empirical question, of course, whether the times are additive.

One might also argue that the model accounts for 99.8% of the variance among the eight critical means mainly by virtue of the positive–negative difference (91.3%), and any model that simply predicts a positive–negative difference will predict over 90% of the variance. This argument must be rejected because although parameters a and c account for only 1.7 and 6.8% of the variance, respectively, these are real parameters. This discrepancy in the variance is due to the fact that the amount of variance accounted for by a parameter is proportional to the square of the parameter. Another reason for the large variance attributable to the positive–negative difference is that it is the sum of two parameters $(b + d)$.

It is, therefore, desirable to concentrate on the magnitudes of the parameters because, first of all, the theory is concerned with the real-time characteristics of these mental operations, and second, the relative magnitudes contain the same information as the variances. We are concerned, therefore, with two things: (1) the empirical determination of the time it takes to perform these mental operations, (2) the determination of whether these times are reliable.

To test the reliability of the parameters, we evaluated the model with an analysis of variance. The parameters a, $(b + d)$, and c, tested against their variability across *Ss*, were found to be significantly greater than zero at at least $p < .005$ with $F(1,11)$ of 21, 95, and 16, respectively. Moreover, the four interactions among these three parameters, as Model A predicts,[4] did not approach significance. The only other significant factor was that sentences with *plus* as subject were verified 97 msec faster than those with *star* as subject $F(1,11) = 5$, $p < .05$. This effect is an empirical result unpredicted by the model, but it must occur at the Sentence Encoding Stage, presumably for good linguistic reasons. As such, the model must predict that this *plus-star* effect be statistically independent of, i.e., not interact with, the parameters of the Comparison Stage. In fact, this parameter did not interact with anything. Perhaps

---

[4] Strictly speaking, a can interact with part of Negation Time, b of $(b + d)$, since these are both sentence-encoding effects. This must be allowed since we have made no assumptions about the details of the Sentence Encoding Stage.

gets underway, the Ss are able to treat *above* and *below* simply as if they were arbitrary nonlinguistic "signals" as to where to look, and they therefore take no longer in using one signal than the other. Although this description is only suggestive, it can be firmly concluded that the *above–below* difference disappears in the forced-choice task because Ss can somehow forego the full encoding requirements of Stage 1 of the verification task.

No matter how the results of Experiment IV are to be explained, however, they cannot be accounted for easily by frequency. Since *above* and *below* "mean" the same thing under the two instructions, frequency would have to predict that *above* would be faster than *below* by the same amount under the two instructions, a prediction contrary to the results. More generally, frequency explanations are by nature difficult to apply to comprehension tasks, for they normally do not specify whether frequency should affect reading time, encoding time, comparison time, or response time, and what the mechanism would be behind such an effect. On the other hand, the present model for *above* and *below* and their representations does offer a plausible account for why *above* should be more frequent in the language than *below:* since *above* is less complex to encode, it should be the term normally used for verticality, and *below* should be used less often. The present theory, or one with similar properties, is therefore preferable to frequency explanations for three reasons: (1) the present theory can account for normative frequencies; (2) frequency predictions were disconfirmed by Experiments II and III for *star* and *plus;* and (3) frequency predictions were disconfirmed by Experiment IV for *above* and *below*.

The Picture-Negation Model

The final model to be examined makes a radically different, but quite plausible assumption about the Stage 3 comparison operations. Briefly, it assumes that an initial mismatch in the comparison process brings about a change in the picture code from positive to negative. For example, a comparison of $(B\ \mathit{above}\ A)_{\mathrm{Sen}}$ against $(A\ \mathit{above}\ B)_{\mathrm{Pic}}$ would cause the latter code to be transformed into $(\mathit{false}\ (B\ \mathit{above}\ A))_{\mathrm{Pic}}$, a code whose embedded string now matches the embedded string of the sentence code. A further comparison is made after this transformation. The two Stage 3 production rules for the picture-negation model that correspond to Model A would be as follows:

(1) If the embedded strings do not match, transform $(A\ \mathit{above}\ B)_{\mathrm{Pic}}$ into $(\mathit{false}\ (B\ \mathit{above}\ A))_{\mathrm{Pic}}$ or $(B\ \mathit{below}\ A)_{\mathrm{Pic}}$ into $(\mathit{false}\ (A\ \mathit{below}\ B))_{\mathrm{Pic}}$, whichever is appropriate.

(2) If the embedding strings do not match, change the truth index.

Thus, rule (1) of this model changes positive picture codes into negative ones on finding a mismatch, whereas rule (1) of Model A changes the truth index on finding such a mismatch. Rule (2) is the same in this model as in Model A. *A priori,* this model has much to recommend it: it produces the correct answer; it does so by only a single alteration of the truth index; and it makes use of two transformations that are logically valid. Nevertheless, this model and its variations are not supported by the present results.

The predictions this model would make for the latencies in Experiment I are shown in Table 8. These formulae were arrived at by the same assumptions as used for Model A: Below Time a is the extra Stage 1 time needed for encoding *below;* the increment b is the extra Stage 1 time needed for encoding negatives; the increment k is the extra time needed whenever production rule (1) is used; and the increment d is the extra time needed whenever production rule (2) is used. The major disparity between these formulae and those of Model A (in Table 2) is that d is now required for False Positive sentences and is no longer required for True Negatives. Estimating these four parameters from the data of Experiment I, we find that a is 93 msec, b is 685 msec, k is 187 msec, and most significantly of all, d is 0 msec. Whereas the picture-negation model predicts a positive difference d between true and false sentences, the results show no difference at all. This result is also found in Experiment III, in Clark and Chase (in preparation), in Young and Chase (1971), and in Experiment IX of Trabasso *et al.* (1971). For the picture-negation model to hold, d must be a nonzero interval, since the verification studies done so far have all shown that mismatches and conversions invariably consume more time than matches. A further deficiency is that this model does not jibe with Ss' informal reports of "changing their answer" once for False Positives and False Negatives and twice for True Negatives;

TABLE 8
Breakdown of Latencies Predicted by the Picture-Negation Model for Eight Types of Sentences

Sentence type			Sentence	Latency components
Positive	True	above	A is above B.	t_0
		below	B is below A.	$t_0 + a$
	False	above	B is above A.	$t_0 + k + d$
		below	A is below B.	$t_0 + a + k + d$
Negative	True	above	B isn't above A.	$t_0 + b + k$
		below	A isn't below B.	$t_0 + a + b + k$
	False	above	A isn't above B.	$t_0 + b + d$
		below	B isn't below A.	$t_0 + a + b + d$

the picture-negation model would not have them change their answer at all for True Negatives. To sum up, the present results lend no support whatever to the picture-negation model.

GENERALITY OF THE MODEL

Since the present theory would be difficult to maintain from the present experiments alone, it is important to demonstrate just how general individual aspects of the theory are. In the following sections, we will therefore discuss the generality of: (1) the coding of *above* and *below* at the encoding stage; (2) the two main comparison operations of the comparison stage; (3) the initial value of the truth index at the comparison stage; and (4) the semantic coding of the picture.

The Coding of Above and Below

In the present theory, it is claimed that *above* takes less time to encode at Stage 1 than *below*. The available evidence shows that this claim has at least some generality. First, the *above–below* difference itself is strikingly constant across experiments. From the data of Seymour (1969), Experiments I and IV of Chase and Clark (1971), Experiments I, II, III, and IV of the present study, and Experiments II, III, and IV of Clark and Chase (in preparation), Below Time *a* can be estimated, respectively, as 53, 75, 83, 93, 117, 84, 90, 136, 113, and 124 msec; these estimates have an average of 97 msec and a standard deviation of only 24 msec. Such stability in Below Time helps attest to the fact that it is an encoding difference separable from radical changes in the rest of the task. This phenomenon is also shown to be general by its consistency over other prepositional pairs that are linguistically similar to *above* and *below*. Clark (1971; in press) has argued that asymmetries should also be found in the encoding of the "front-back" prepositions, with those referring to "front" having the advantage, as well as in the encoding of the other "top–bottom" prepositions. Indeed, the available evidence reviewed in Clark (in press) demonstrates the expected asymmetries for the preposition pairs *on top of–under, in front of–in back of, ahead–behind, before–after,* as well as for the related adjective pairs *first–last,* and *early–late.* Thus, the *above–below* effect is not an isolated one, but appears to fit a broader pattern of encoding asymmetries in prepositions.

Two Operations at the Comparison Stage

As we pointed out above, Clark (1970; in press) and Trabasso (1970; in press; Trabasso *et al.*, 1971) have independently proposed almost identical models to account for the verification of negative sentences in general. In the present model, the comparison stage consists of two ordered production rules:

(1) If the embedded strings do not match, change the truth index.
(2) If the embedding strings do not match, change the truth index.

Trabasso's models contain two steps that are functionally equivalent to these two rules. In addition, both Clark and Trabasso have reviewed the extensive literature on negatives, demonstrating that their models account for almost all of the previous results with relatively precise quantitative predictions. And both have noted that the Ss in some of the previous experiments were able to bypass this two-rule model by "converting" negatives to positives and that the possibility of using these conversions depended on specific properties of the task and of the type of sentences used.

More importantly, despite their minor differences, the present model and Trabasso *et al.*'s are based on the same assumptions. First, the S seeks a match in the underlying codes of the sentence and picture. Second, whenever there is a mismatch, the S must carry out additional time-consuming operations and transformations, and this implies that mismatches require more time. And third, the two Stage 3 operations are sequential and the time increments they consume are therefore additive. Insofar as Clark (1970; in press) and Trabasso *et al.* (1971) have shown that these assumptions hold fairly well across a variety of studies on negation and other types of sentences, then they have also shown the generality of Stage 3 in the present Models A and B.

The Initial Value of the Truth Index

One assumption of both the present and Trabasso *et al.*'s models is that the initial value of the truth index is *true*. That is, the S considers the sentence to be true of the picture unless he finds evidence to the contrary. This assumption was made in order to account for the superiority of true affirmative sentences. These sentences match the picture directly, and if the truth index is assumed to be *true* initially, then no additional operation is required for the response to be readied, so it will be fast. But to show the generality of this assumption, we will present additional linguistic evidence that demonstrates that it is quite proper to assume that the truth index is initially *true*.

To make this demonstration clear we will first recast the present models in two non-essential respects. The first change to be made is to place the truth index within the sentence representation itself. In the present theory as it stands, there is a truth index attached to each sentence to indicate whether it is true or false. However, the theory could have been formulated with the truth index incorporated directly into the sentence, as in *It is* $true_{\text{index}}$ *that A is above B, It is* $true_{\text{index}}$ *that A isn't above B,* etc. whose respective representations would be ($true_{\text{index}}$ (*A above B*)), (*true*$_{\text{index}}$ (*false* (*A above B*))), etc. The advantage of this notation is that it makes explicit just how the truth index is related to the content of each sentence. The second change is to view the verification task as one in which the *S* is required to answer questions like *Is it true*$_{\text{index}}$ *that A is above B?* *Is it true*$_{\text{index}}$ *that A isn't above B?*, etc. For convenience, these questions can be denoted, respectively, as (*true*$_{\text{index}}$? (*A above B*)), (*true*$_{\text{index}}$? (*false* (*A above B*))), etc.

Under this recasting of the present theory, the claim that the initial value of the truth index is *true* instead of *false* is equivalent to the claim that Ss are answering the question *Is it true*$_{\text{index}}$ *that A is above B?* Instead of *Is it false*$_{\text{index}}$ *that A is above B?* Linguistic evidence shows why these questions should contain *true*$_{\text{index}}$ instead of *false*$_{\text{index}}$. Consider the neutral question *Is A above B?* This question can be accurately paraphrased as *Is it true that A is above B?* but not as *Is it false that A is above B?* By asking *Is it true?*, the questioner does not commit himself as to what he expects the answer to be, whereas by asking *Is it false?*, the questioner indicates that he expects A not to be above B and that he only wants confirmation of that negative expectation. (Related neutralization phenomena have been discussed by Clark (1969) and Fillenbaum (1968).) Thus, with the use of *true*$_{\text{index}}$ in the initial questions, the present theory entails no expectation on the *S*'s part about the truth or falsity of the sentence. But with the use of *false*$_{\text{index}}$ instead, the theory would have to have assumed that the *S* began with the expectation that the sentence was false and that he had only to conform his expectation. Clearly, the use of *true*$_{\text{index}}$ is more appropriate for the normal verification task. It is significant that English (and probably all other languages) is optimally designed to accept confirming rather than disconfirming evidence and that this property fits so conveniently into the present models for empirical reasons alone.

The Picture Representation

The present theory also assumes that the picture is coded at Stage 2 in a propositional format so that it can be compared against the sentence representation at Stage 3. The generality of this assumption is strengthened by evidence from the previous studies on the verification of positive and negative sentences. In each of these studies, the *S* was asked to verify a sentence against evidence external to the sentence. In some cases, the evidence consisted of pictures (Gough, 1965, 1966; Slobin, 1966; Trabasso *et al.*, 1971; the present experiments); in other cases, it consisted of previous knowledge, e.g., about the evenness and oddness of numbers (Wason, 1961; Eifermann, 1961; Wason & Jones, 1963) or about sums of digits (Wales & Grieve, 1969); in still other cases, the evidence consisted of other sentences (Greene, 1970a,b). The models that Clark (1970; in press) and Trabasso *et al.* (1971) have recently proposed to account for these studies assumed that all three types of evidence are encoded in the same format. If the models had not done this, then there would have had to have been three distinct models of negation, one for each type of evidence. Obviously, it is a more powerful generalization to construct one model instead of three. So there is evidence for the assumption that the *S* does indeed represent pictures, previous knowledge, and sentences in the same format. This agrees exactly with the assumption that seems also to be required in the present experiments to account for the various changes in the picture code induced by the different viewing instructions.

UNANSWERED QUESTIONS

The present theory obviously does not answer many questions about the process of comparing sentences against pictures. It is instructive to examine the most obvious of these questions and the reasons for them.

Sentence Representations

We have made no attempt in the present theory to account for how Ss actually construct semantic representations from the printed sentences they encounter—how Ss, for example, take in the printed sentence *Star isn't above plus* and from it construct (*false* (*star above plus*))$_{\text{sen}}$. This omission has been quite deliberate both because the rest of the theory does not seem to depend on how this process is carried out and because too little can be said about it at the present time. The problem here is one of research strategy. We know that Ss interpret sentences and that these interpretations are necessary for later comparisons. This makes it possible to study ~~the~~ structure of representations like (*false* (*star above plus*))$_{\text{sen}}$ without knowing precisely how they are formed. In contrast, it is difficult to study how Ss construct such representations from surface structure without knowing what the representations actually are. The present study and others like it, then, define the problem to be solved: how do Ss construct representations like (*false* (*star above plus*))$_{\text{sen}}$ from surface structures like *Star isn't above plus?* This question has not been ignored in recent years, witness a number of general reviews of the issue, e.g., by Bever (1970), Fodor and Garrett (1966), and Watt (1970), and much experimental work, e.g., by Clark and Begun (1968), Fodor and Garrett (1967), Fodor, Garrett, and Bever (1968), Foss (1969), Foss and Lynch (1969), Hakes and Cairns (1970), and others.

The present experiments, in fact, place several quite strong constraints on the process of constructing such representations. The results showed encoding differences between *above* and *below*, between *star* and *plus*, and between positive and negative sentences; furthermore, these differences were statistically independent of each other. Thus, the extra processing steps required for representing *below, star,* and negatives, however these steps are conceived, must be at least partly serial in nature, even though it is quite conceivable *a priori* that they could have been carried out completely in parallel.

Picture Representations

We have also not attempted in the present theory to spell out the processes by which the pictures are represented, and for exactly the same reasons as with the sentence representations. Probably even less is known about this process, and the present experiments add only a few facts. It was found, first, that without any other constraints, a picture of two vertically arranged objects is normally represented as $(A\ \textit{above}\ B)_{\text{Pic}}$; nevertheless, when the S attends specifically to the position of the upper or lower object, it is represented as $(A\ \textit{above}\ B)_{\text{Pic}}$ or $(B\ \textit{below}\ A)_{\text{Pic}}$, respectively. These facts will have to be accounted for in a theory of how Ss go about representing pictures.

A more serious problem for research in this area, however, is how to specify the level of abstraction at which a picture is represented. A plus (+), for example, can be represented at at least three levels of abstraction: (1) a horizontal and vertical line, of equal length, bisecting each other; (2) a plus; (3) a mathematical symbol denoting addition. Here level (1) is the "most primitive" code, level (2) more "interpreted," and level (3) the most "interpreted." The question is, which of these representations is used in the Stage 3 comparison process? Although the level (2) code has been assumed in the present theory, other representations are clearly possible and perhaps even necessary in other tasks.

In this regard, the extensive research of Posner and his colleagues, as reviewed in Posner (1969), details the circumstances under which level (1) codes and level (2) codes are made use of in judging matches of alphabetic characters. This research has demonstrated, for instance, that two letters can be judged to match more quickly if they are physically identical (i.e., AA) than if they have only the same name (e.g., Aa) but this happens only when the interval between the two letters is short or when the letter presented second is predictably of upper or lower case. This shows that although level (1) codes can be used most quickly, they are hard to keep around and are generally replaced by level (2) codes after a short interval unless the S knows that the level (1) information can be utilized directly. Tversky's (1970) work on the matching of faces can be interpreted in much the same way. In addition, Posner has shown that the level (1) code is not simply a visual iconic image. It is an "abstract idea"—to use Posner's term—since it has properties, for instance, that make it impervious to visual masking. Returning to the present results, we note that it took Ss longer to code the star (*) than the plus (+). This could have happened, then, because it takes longer with the star to construct the initial level (1) code, to derive the level (2) code from the level (1) code, or both.

Finally, there is the problem of how to extend this type of propositional analysis to more complex and detailed pictures. One could rightly complain that a star above a plus, for example, is not a "picture," but a "diagram." The implication of such a complaint is that diagrams, though not pictures, can be coded in a propositional form, so the present theory applies only to diagrams. This could, perhaps, be true. But it seems reasonable that in a picture-first task much of the detailed information of more complex pictures is simply not open to verification by sentence; perhaps only that information that has been coded, or can be deduced from the information coded, is open to such verification. In other words, the picture must conceptually be reduced to a diagram or series of interlocking diagrams so that its information can be coded in an "interpreted" propositional form before it can be verified. The details of such a proposal, however, clearly wait on future research.

Comparison Process

The comparison stages of Models A and B have much empirical backing. Yet it is possible to construct other equivalent, and perhaps even nonequivalent, models to account for the present results. Eventually, Models A and B must become part of an even more general theory which captures significant generalizations about an even larger variety of sentence–picture comparison tasks. To the extent that Models A and B do generalize to previous studies of negation, question-answering, concept identification, and other types of matching tasks, these models are very strong. Yet it is easy to demonstrate that the present models are both deficient and indeterminate in certain obvious respects. By way of illustration, we will examine two points that need further clarification.

As now specified, Models A and B will not give the correct answer for the sentence *Star is above circle* and the picture $\overset{*}{+}$. Consider Model A. It assumes that while Ss code the subject and preposition of each sentence in the present task, they do not code the object of the preposition (here *circle*), since it is always redundant. But in the above example, the object is not redundant, and Operation 1 must read something like "Compare the subject *and* object of the sentence and picture codes." This could well be shown to consist of two separate comparison operations with additive latencies. The present models would, therefore, have to be modified somewhat to handle *Star is above circle* as well as other sentences like *The star is above a mathematical symbol, The star isn't above anything, Nothing is below the plus, etc.* Since these modifications cannot be derived from the present results alone, they await future experimentation.

A second point needing clarification is how to alot time increments, say, to Operations 1 and 1a. The parameter *c* (Falsification Time) was identified as the increment of time required to find a mismatch and change the truth index. That is, *c* was attributed to the sum of (1) the extra time required by Operation 1 to find a mismatch over that required to find a match, plus (2) the time required by Operation 1a. Ideally, one would like to identify these two incremental parts of Falsification Time separately and demonstrate that the time increments required by Operation 1a and 2a—two formally identical operations—are the same. The present experiments do not provide an answer, nor is it obvious what experiments would. To sum up, the present models should be regarded simply as the essential first steps in the specification of a full theory of sentence–picture comparison.

CONCLUSIONS

In the present paper we have argued from a variety of evidence that people compare sentences against pictures through a series of discrete stages. People have to encode the sentence and picture they are to compare in a common abstract representation and then match the two representations against each other piece by piece. Significantly, the separate stages are carried out one after the other, and the times consumed by each stage accumulate in an additive fashion. What is striking about this theory is that many of its assumptions are necessary anyway to account for data in a variety of other cognitive tasks. The assumptions that sentences and pictures are coded in an abstract semantic format are consistent with previous studies in sentence verification, deductive reasoning, and even iconics and pattern recognition by computer. The comparison assumptions—particularly the principle of congruence and the assumption that the initial value of the truth index is *true*—also help to account for past studies in sentence verification, question answering, and deductive reasoning (cf. Clark, in press).

Considered more generally, the present theory can be thought of as a rapprochement between psycholinguistics and visual perception. One often hears psychologists speak of "verbal" and "perceptual" systems as if the two are quite separate and have little in common. The thrust of the present paper, in part, has been to demonstrate quite the reverse.

Underlying both language and perception, we have argued, is a common "interpretive" system that must be handled by one set of principles no matter whether the source of a particular interpretation is linguistic or perceptual. Thus, if the present theory proves to be correct, psychologists must come to regard certain thought processes not simply as "verbal" or "perceptual" in nature, but as "cognitive." These are the processes that deal with interpreted knowledge.

REFERENCES

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley, 1970.

Bierwisch, M. Some semantic universals of German adjectivals. *Foundations of Language,* 1967, 3, 1–36.

Brooks, L. R. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology,* 1968, 72, 349–368.

Chase, W. G., & Clark, H. H. Semantics in the perception of verticality. *British Journal of Psychology,* 1971, 62, 311–326.

Chase, W. G., & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, in press.

Chase, W. G., Young, R. M., Singer, M., & Clark, H. H. Practice effects in the comparison of sentences and pictures. Paper presented at Midwestern Psychological Association, Chicago, April 1971.

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, 76, 387–404.

Clark, H. H. How we understand negation. Paper presented at COBRE Workshop on Cognitive Organization and Psychological Processes. Huntington Beach, California, August 19, 1970.

Clark, H. H. Space, time, semantics, and the child. Paper presented at Conference on Developmental Psycholinguistics, State University of New York at Buffalo, August 1971.

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., & Begun, J. S. The use of syntax in understanding sentences. *British Journal of Psychology,* 1968, 59, 219–229.

Clark, H. H., & Chase, W. G. On generating and verifying descriptions of vertical location. In preparation.

Eiferman, R. R. Negation: A linguistic variable. *Acta Psychologica,* 1961, 18, 258–273.

Fillenbaum, S. Recall for answers to "conducive" questions. *Language and Speech,* 1968, 11, 46–53.

Fitts, P. M., & Posner, M. I. *Human performance.* Belmont, CA.: Brooks/Cole, 1967.

Fodor, J., & Garrett, M. Some reflections on competence and performance. In J. Lyons and R. Wales (Eds.), *Psycholinguistics papers.* Edinburgh: Edinburgh University Press, 1966.

Fodor, J. A., & Garrett, M. Some syntactic determinants of sentential complexity. *Perception and Psychophysics,* 1967, 2, 289–296.

FODOR, J. A., GARRETT, M., & BEVER, T. G. Some syntactic determinants of sentential complexity: II. Verb structure. *Perception and Psychophysics*, 1968, 3, 453–461.

FOSS, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 457–462.

FOSS, D. J., & LYNCH, R. H., JR. Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception and Psychophysics*, 1969, 5, 145–148.

GIVON, T. Notes on the semantic structure of English adjectives. *Language*, 1970, 46, 816–837.

GOLDMAN-EISLER, F., & COHEN, M. Is N, P, and PN difficulty a valid criterion of transformational operations? *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 161–166.

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.

GRASSELLI, A. (Ed.). *Automatic interpretation and classification of images*. New York: Academic Press, 1969.

GREENE, J. M. Syntactic form and semantic function. *Quarterly Journal of Experimental Psychology*, 1970, 22, 14–27. (a)

GREENE, J. M. The semantic function of negatives and passives. *British Journal of Psychology*, 1970, 61, 17–22. (b)

HAKES, D. T., & CAIRNS, H. W. Sentence comprehension and relative pronouns. *Perception and Psychophysics*, 1970, 8, 5–8.

JONES, S. The effect of a negative qualifier in an instruction. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 495–501. (a)

JONES, S. Decoding a deceptive instruction. *British Journal of Psychology*, 1966, 57, 405–411. (b)

JONES, S. Instructions, self-instructions and performance. *Quarterly Journal of Experimental Psychology*, 1968, 20, 74–78.

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

LEEPER, R. W. A study of a neglected portion of the field of learning: The development of sensory organization. *Journal of Genetic Psychology*, 1935, 46, 41–75.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, 76, 241–263.

POSNER, M. I. Abstraction and the process of recognition. In G. Bower and J. Spence (Eds.) *The psychology of learning and motivation*. New York: Academic Press, 1969. Vol. 3.

ROSS, J. R. On declarative sentences. In R. A. Jacobs and P. S. Rosenbaum (Eds.) *Readings in English transformational grammar*. Waltham, MA.: Ginn, 1970.

SEYMOUR, P. H. K. Response latencies in judgments of spatial location. *British Journal of Psychology*, 1969, 60, 31–39.

SLOBIN, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 219–227.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, 69, 77–110.

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.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

TRABASSO, T. Reasoning and the processing of negative information. Invited address to Division 3 of the American Psychological Association, Miami, September 1970.

TRABASSO, T. Mental operations in language comprehension. In J. B. Carroll and R. O. Freedle (Eds.) *Language comprehension and the acquision of knowledge*. Washington: V. H. Winston & Sons, in press.

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

TVERSKY, B. Pictorial and verbal encoding in a short-term memory task. *Perception and Psychophysics*, 1969, 6, 225–233.

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.

WATT, W. C. *Morphology of the Nevada cattlebrands and their blazons*, Part One. National Bureau of Standards Report 9050, Washington, D. C. 1966.

WATT, W. C. *Morphology of the Nevada cattlebrands and their blazons*, Part Two. Department of Computer Science, Carnegie–Mellon University, 1967. (a)

WATT, W. C. Structural properties of the Nevada cattlebrands. *Computer Science Research Review*, Carnegie–Mellon University, 1967, 20–27. (b)

WATT, W. C. Some remarks on iconic grammar. Paper presented at American Anthropology Association Meetings, New Orleans, November, 1969.

YOUNG, R., & CHASE, W. G. Additive stages in the comparison of sentences and pictures. Paper presented at Midwestern Psychological Association meetings, Chicago, April, 1971.

(Accepted January 21, 1972)