The Role of Imagery in the Evaluation of Sentences: Imagery or Semantic Factors?[1]

KEITH J. HOLYOAK

Stanford University

Jorgensen and Kintsch (1973) found that sentences rated as easy to image were verified more quickly than sentences rated as hard to image. In the present study subjects rated the semantic relatedness of subject and predicate words in the sentences used by Jorgensen and Kintsch, the comprehensibility of the complete sentences, and also the difficulty of defining the verbs. Subjects found low-imagery sentences to be relatively difficult to understand, and verbs from low-imagery sentences were rated as more difficult to define. Subject and predicate words were more closely related in high-imagery true sentences and in low-imagery false sentences. It was concluded that the results of Jorgensen and Kintsch do not distinguish the effects of rated imagery on RT from the effects of semantic relatedness and semantic complexity.

In a recent study of semantic memory, Jorgensen and Kintsch (1973) first obtained ratings of the ease with which subjects could image simple sentences of the form Noun-has-Noun (for example, *Truck has oil*) or Noun-Verb-Noun (for example, *Robin eats worm*). They then measured the reaction time (RT) of other subjects in verifying these sentences. Although 108 sentences were included in the experiment, the results reported were based on a subset of 39 sentences drawn from the extremes of the distribution of imagery ratings. The major finding was that, for both true and false sentences, high-imagery sentences were verified more quickly than were low-imagery ones. This effect was significant across subjects, but apparently was not tested against item variability as recommended by Clark (1973). Explicit instructions to some subjects to use imagery to verify the sentences had no significant effect on their RT. On this basis it was concluded that image utilization is a natural and effective strategy for subjects in verification tasks.

However, there are reasons to doubt this conclusion. Jorgensen and Kintsch noted that no previous evidence suggests that subjects are able to form explicit images in the time (1 to 2 seconds) needed to respond in the verification task; they also discuss possible confoundings between rated imagery and semantic variables. The semantic representation of the relationships expressed in low-imagery sentences may possibly be more complex, making such sentences more difficult to encode. This was especially possible for the N-V-N sentences; while only concrete nouns were used in all sentences, the verbs in low- and high-imagery sentences were apparently not equated in any comparable way. It is therefore possible that the verbs in the low-imagery N-V-N sentences expressed more complex meanings. Semantic relatedness, or the degree of overlap between the meanings of subject and predicate words, also was not controlled in the sentences used in the Jorgensen and Kintsch study. Jorgensen and Kintsch suggest that subject and predicate words were possibly more closely related in high-imagery true sentences, but dismiss the

163

7

possibility that such a confounding could have been present for their false items. In fact, however, while high semantic-relatedness facilitates verification of true sentences (see Smith, Shoben, & Rips, 1974), it has generally been found to increase time to reject false statements (Kintsch, Crothers, & Berman, 1970; Meyer, 1970; Rips, Shoben, & Smith, 1973; Wilkins, 1971). In order to demonstrate that the Jorgensen and Kintsch results may have been mainly due to a confounding of rated imagery and semantic relatedness, it would therefore be necessary to show that for their false items subject and predicate words were more closely related in the low-imagery sentences.

In order to assess whether such confoundings were present, two measures were obtained for each of the 39 most extreme low- and high-imagery sentences used by Jorgensen and Kintsch.[2] For all items subjects rated the semantic relatedness of subject and predicate words, as well as the comprehensibility of the complete sentence. In addition, subjects rated the difficulty of defining each of the verbs from the 25 N-V-N sentences.

METHOD

A group of 20 subjects performed two rating tasks. In the first, each subject received a sheet of paper with 39 subject and predicate word-pairs taken from the Jorgensen and Kintsch sentences, listed in a column on the left-hand side of the page. The words of each pair appeared side-by-side in the order S-P; the order of the 39 pairs was random. A 1–7 scale appeared beside each pair. For each pair, subjects were instructed to circle a number "indicating how closely you feel the two words are associated in meaning," with 7 indicating high relatedness.

Following this task a second questionnaire was administered. This was similar to the first

[2] I am indebted to C. Jorgensen for providing me with the sentences used by Jorgensen and Kintsch.

except that on this sheet the 39 complete sentences were listed on the left, in a different random order. Instructions included the following: "Many of these sentences are unlikely to ever be true, but don't let this confuse you. Rate each sentence on how easily you are able to grasp its meaning, even if its meaning is improbable. All of these sentences are very simple, so if you feel the slightest hesitation in grasping a sentence's meaning, mark it down quite heavily." A rating of 7 indicated that the sentence was extremely easy to understand.

Another group of 28 subjects performed a different rating task. These subjects were given a sheet which listed in random order the 22 different verbs taken from N-V-N sentences. Subjects were instructed to rate each verb according to "how difficult it is to give a complete definition of the verb's meaning." A rating of 7 indicated the verb was extremely difficult to define. All verbs were written in the infinitive form. Four verbs appeared ambiguous; these were presented along with an N-V-N example sentence (not that used by Jorgensen and Kintsch) which illustrated the meaning subjects were to rate (for example, to drive was presented with Man drives car). Subjects were instructed that if they found any other verb ambiguous they were to "just rate the most obvious meaning." Prior to performing this task, subjects were given 13 verbs not used by Jorgensen and Kintsch, and asked both to write complete definitions of each verb and to rate the difficulty of doing so. Subjects were not asked to actually define the Jorgensen and Kintsch verbs, but were told to rate them in the same way as they had rated the previous ones.

In order to replicate the initial results, the relatedness questionnaire was also completed by a different group of 20 subjects. All subjects who completed the rating scales (68 in total) were Stanford students who participated either for pay, in order to satisfy a course requirement, or as volunteers.

RESULTS

Means for each of the three rating tasks, as a function of imagery level and truth value, are presented in Table 1. An analysis of variance, treating subjects as a random effect, was performed on each set of results. Since the set of sentences used by Jorgensen and Kintsch constitutes the complete population of items with which the hypotheses to be tested are concerned, items was treated as a fixed rather than as a random effect (Clark, 1973). The two groups of subjects who completed the relatedness ratings produced essentially the same results. Data from these two groups was therefore combined into a single analysis, which examined the effects of rated imagery of the sentence (high vs. low), verb-type (N-has-N vs. N-V-N sentences), and truth value (true vs. false). As would be expected, subject and predicate words drawn from true sentences were rated as more closely related than were those drawn from false sentences, $F(1, 39) = 683, p < .001$. The overall difference between imagery conditions was also significant, but was completely qualified by an interaction with truth value, $F(1, 39) = 136$, $p < .001$. The data in Table 1 show that for true sentences the subject and predicate words of high-imagery sentences were rated as more closely related, $t(39) = 8.83$, $p < .001$; but among the false items, word-pairs derived from low-imagery sentences were actually more closely related than those taken from high-imagery sentences, $t(39) = 3.30, p < .01$. A similar interaction was found between verb-type and truth value. Subject and predicate words from true N-has-N sentences tended to be rated as more related to one another in meaning than were those drawn from true N-V-N sentences, although this difference was not significant; the reverse ordering occurred for false sentences, $t(39) = 4.52, p < .001$.

The analysis of comprehensibility ratings yielded only two significant effects. High-imagery sentences were rated as more comprehensible than low-imagery sentences, $F(1, 19) = 47.8$, $p < .001$, and true sentences were understood more readily than false sentences, $F(1, 19) = 47.1, p < .001$. Ratings of the verbs used in N-V-N sentences produced similar results. The verbs from low-imagery sentences were rated as more difficult to define than were those from high-imagery sentences, $F(1, 27) = 36.8$, $p < .001$; this difference was significantly larger for verbs taken from false sentences, $p < .05$.

DISCUSSION

The greater comprehensibility reported for high-imagery sentences is consistent with an alternative explanation of the Jorgensen and Kintsch results which attributes them to confounding of imagibility of the sentences with their degree of semantic complexity. It might be argued that sentences which do not readily elicit an image are rated as less comprehensible for that very reason, that is, ease of imaging determines ease of comprehension. However, the semantic complexity hypothesis

TABLE 1

RATINGS (MAX = 7) OF RELATEDNESS OF S AND P WORDS, COMPREHENSIBILITY, AND DIFFICULTY OF DEFINING VERB, FOR LOW- AND HIGH-IMAGERY SENTENCES USED BY JORGENSEN AND KINTSCH (1973)

	Degree of relatedness		Ease of comprehension		Difficulty of defining verb[a]	
	Low—I	High—I	Low—I	High—I	Low—I	High—I
True	4.39	5.54	5.70	6.65	3.61	3.03
False	2.11	1.68	3.26	3.93	4.13	3.02

[a] Note: Based on N–V–N sentences only.

is strengthened by the findings concerning the verbs in the N-V-N sentences (which constitute 25 of the 39 sentences being examined). The verbs used in low-imagery sentences were more difficult to define than were those used in high-imagery sentences. This suggests that the differences in verification latency which Jorgensen and Kintsch obtained among N-V-N sentences may have been partly due to systematic differences in verb complexity, rather than to differences in imagibility of the overall sentence.

The differential relatedness of subject and predicate words in sentences of different imagery levels also highlights a serious artifact in the Jorgensen and Kintsch study. The greater relatedness of the nouns in their true high-imagery sentences, as well as in their false low-imagery sentences, would be expected for this reason alone to produce faster verification of high-imagery sentences in both true and false conditions (Smith et al., 1974). Consequently, the Jorgensen and Kintsch data offer no clear support for the hypothesis that imaginal processes play a role in sentence verification.

A possible explanation based on semantic relatedness can also be offered for an incidental finding in the Jorgensen and Kintsch study. N-has-N sentences were verified more quickly than N-V-N sentences, an effect which the authors were unable to explain. But the present results indicate that verb-type interacts with truth value much as rated imagery does—subject and predicate words tended to be more closely related in their true N-has-N sentences, and in their false N-V-N sentences. Therefore, the differences obtained in RT as a function of verb-type, as well as rated imagery, might be due to these confoundings with semantic relatedness.

The present results clearly indicate that the effects of rated imagery on RT cannot be distinguished from the effects of semantic variables on the basis of reported results obtained using these particular items. Many important questions remain concerning the possible role of imagery in sentence verification. But research on this topic can advance only if linguistic variables such as semantic relatedness, which are known to strongly influence RT, are carefully controlled when imagibility is varied.

REFERENCES

CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335–359.

JORGENSEN, C. C., & KINTSCH, W. The role of imagery in the evaluation of sentences. Cognitive Psychology, 1973, 4, 110–116.

KINTSCH, W., CROTHERS, E. J., & BERMAN, L. N. The effects of some semantic and syntactic properties of simple sentences upon the latency of judgements of semantic acceptability. In Studies in Mathematical Learning Theory and Psycholinguistics, University of Colorado, 1970.

MEYER, D. E. On the representation and retrieval of stored semantic information. Cognitive Psychology, 1970, 1, 242–300.

RIPS, L. J., SHOBEN, E. J., & SMITH, E. E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1–20.

SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. Structure and processes in semantic memory: A featural model for semantic decisions. Psychological Review, 1974, in press.

WILKINS, A. T. Conjoint frequency, category size, and categorization time. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 382–385.