ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

TECHNICAL REPORTS

Technical Report No. 6

TWO FACES OF THE CONCEPTUAL PEG HYPOTHESIS

Richard C. Anderson, Ernest T. Goetz,
James W. Pichert, and Henry M. Halff

January 1976

LABORATORY for COGNITIVE STUDIES in EDUCATION

University of Illinois at Urbana-Champaign

Technical Report No. 6

TWO FACES OF THE CONCEPTUAL PEG HYPOTHESIS

Richard C. Anderson, Ernest T. Goetz,
James W. Pichert, and Henry M. Halff

January 1976

University of Illinois
Laboratory for Cognitive Studies in Education
236 Education Building
Urbana, Illinois 61801

The research described herein was supported by the National Institute of Education under Grant No. HEW NIE-G-74-0007.

Abstract

The present study investigated why it is that the more concrete the subject noun phrase of a sentence, the more likely the predicate is to be recalled when the subject noun phrase is the cue. The findings were that concretization dramatically influences both the probability of recognition of the subject noun phrase and the probability of recall of the predicate given recognition. These results were taken to mean that a concrete phrase makes a good conceptual peg because it is likely to be given a specific, stable encoding and because it tends to redintegrate the whole sentence. Regression analysis showed that the concreteness effect could not be attributed to an influence on comprehensibility. A model of sentence memory is offered which can account for the results.

Two Faces of the Conceptual Peg Hypothesis

It is well known that concrete language makes word pairs and sentences more memorable. The most widely accepted explanation for this fact is that concrete words readily form integrated, holistic units, often conceived to take the form of mental images (cf. Paivio, 1971). Several lines of evidence support the hypothesis that it is the holistic character of the mental representations to which concrete words and concretizing operations give rise that accounts for their facilitative effect. For example, Begg (1972) compared cued recall and free recall of noun phrases. He found that when both the adjective and noun were concrete, such as in rusty engine, cued recall of either phrase constituent using the other word as the cue was much better than free recall. For phrases composed of abstract adjectives and nouns, such as absolute truth, cued recall was no better than free recall.

When subjects do not spontaneously construct unified representations--or conditions are contrived to work against such organization--imagery instructions, pictures, drawings, and concrete words lose their potency. Experiments comparing separate pictures with pictures integrating the stimulus and response elements into a single scene invariably show a distinct advantage for the composite pictures (Epstein, Rock, & Zuckerman, 1960; Davidson, 1964; Reese, 1965). Bower (1972) found that subjects asked to bring to mind a scene of two objects interacting in some way recalled substantially more than subjects told "to imagine the two objects one at a time in their imaginal space, like two pictures being seen on opposite walls of a room." Finally, Rohwer (1967) found that placing a verb or a preposition

between two concrete nouns (the rock breaks the bottle, the rock behind the bottle) facilitated learning whereas connecting them with a conjunction (the rock and the bottle) did not. In summary, there is a quite convincing case that it is the quality of being easily organized into unified mental representations which accounts for the advantage of concrete elements.

A very different interpretation of the concreteness effect can be derived from the encoding variability hypothesis of Martin (1968) and the complementary concept of encoding specificity (cf. Tulving & Thompson, 1973). Martin has developed a two-stage model of cued recall performance. According to this model, when the cue is presented it must make contact with the representation of the cue stored during study. Next the response must be accessed from this stored representation. The first stage will succeed only when the cue is given the same encoding at the test as it was during study. If the two encodings fail to match, recall can not occur. It stands to reason that a concrete, denotatively specific term will permit fewer encodings than an abstract term. Consequently, concreteness could increase the probability that the encodings of the cue at test and study will match rather than the likelihood of an integrated representation.

It would appear, then, that there are at least two versions of the "conceptual peg" hypothesis (Paivio, 1969, 1972). The one which Paivio favors is that the concrete cue is more likely to evoke the whole idea, that is, that it has greater "redintegrative power" (Horowitz & Prytulak, 1969). But a plausible alternative is that concrete stimuli are more recognizable than abstract stimuli because there is a higher probability that they will have been given specific, stable encodings.

To date, the only direct experimental tests of these two explanations for the effects of concreteness have been completed by Wicker and his associates (Wicker, 1970; Wicker & Evertson, 1972), who have investigated picture-word differences using a paired associate task. Drawings or concrete nouns representing the same common objects were, respectively, the concrete and less concrete stimuli. A combined recognition and recall test included new drawings and words to serve as distractors. For each item, the subject indicated whether it was old or new and, if judged old, attempted to give the response element. These experiments have consistently shown better recognition of pictures than words, suggesting less encoding variability of the more concrete stimulus. However, no differences in recall conditional upon recognition have appeared, a fact inconsistent with the redintegration hypothesis. Hence, the results appear to differentially support the encoding specificity interpretation of the concreteness effect.

There are at least two reasons for not accepting Wicker's results at face value, however. First, like many other experiments of this general type (cf. Martin, 1967), it can be argued that recognition was confounded with capacity to recall. When a subject judged a cue to be old, he was immediately thereafter expected to produce a response. This task demand may have caused the criterion for saying "old" to shift depending upon whether a response was available. If this happened, of course, the data would give the appearance of locating the concreteness effect in the recognition phase even though it was really due to redintegration. In the experiment reported here, the recognition test preceded and was completely separate from the recall test. While response availability may still have

affected the old/new judgement, at least gross confounding was avoided; our procedures did not invite a recognition strategy based on whether or not the rest of the configuration could be recalled.

There is a second reason for not counting Wicker's data too heavily against the redintegration hypothesis. His pairs were completely arbitrary, and arbitrary concantenations do not necessarily form units (Horowitz & Prytulak, 1969). Redintegration depends upon there being holistic images (Begg, 1972) or, perhaps, unified propositions (Kintsch, 1974) so that there is something to be redintegrated.

The present experiment employed meaningful sentence pairs of the kind used by Anderson (1974). One sentence in each pair contained a concretely-modified subject noun phrase, the other a redundantly-modified subject noun phrase. Below are some examples:

The traditional customs fascinated the tourists.

The tribal marriage customs fascinated the tourists.

The parking regulations annoyed the salesman.

The official regulations annoyed the salesman.

The set of official regulations, for instance, is not much smaller than the set of all regulations, whereas parking regulations are a distinctly smaller subset. In other words, the concrete phrases were more denotatively specific than the redundant ones. Anderson (Experiment II) found that subjects were about one-and-one-half times as likely to recall the predicates of sentences which began with concrete than redundant subject noun phrases,

given the subject noun phrases as cues. The purpose of the experiment described herein was to determine whether the advantage of the concrete phrases can be attributed to their superior recognizability or their capacity to redintegrate whole ideas.

Method

Subjects. The subjects were 77 male and female undergraduates enrolled in introductory educational psychology, who participated to fulfill a course requirement. A total of 118 additional subjects from the same population took part in studies to norm the materials.

Materials. Forty pairs of sentences of the type already illustrated were constructed. Both members of any pair were identical save for the concrete or redundant adjective preceding the subject noun. The subject nouns were abstract or general terms. (A preliminary experiment showed that concrete modifiers have little impact when the subject noun is a specific, concrete term.) The predicates contained familiar, concrete words. The sentences were written to be readily understandable no matter which adjective was used.

Every sentence was rated on a scale from 1 to 7 where 1 meant very hard to understand and 7 meant very easy to understand. Each of two groups of 20 subjects saw 20 concrete and 20 redundant sentences. There were two counterbalanced lists such that sentences which were concrete for one group were redundant for the other. The reliability of the comprehensibility ratings, estimated using an analysis of variance procedure, was .91.

A total of 78 subjects rated the subject noun phrases for similarity to a set of distractors. Subjects received a mimeographed booklet containing 10 concrete and 10 redundant noun phrases. Also provided was a sheet listing 10 concrete and 10 redundant distractor phrases. Which set of phrases was the distractor set and which the target set was counterbalanced. For each phrase in the booklet the subject was to select the distractor item which was most similar in meaning, and then rate the similarity of this distractor and the target phrase on a seven-point scale where 1 meant very different in meaning and 7 meant very similar in meaning. The reliability of the similarity ratings was .79.

Design. The variable of interest was type of modifier, concrete or redundant. Sentence set and list were two additional variables required for counterbalancing. The forty sentence pairs were divided into two sets in a manner the experimenters judged would minimize intraset intrusions. Each set was partitioned into complementary lists containing one sentence from each pair, of which 10 were concrete and 10 redundant. Finally, the lists were divided into two blocks. Block order was counterbalanced. The order of items within blocks during study and the two tests was random, but on each occasion the block order was the same. Blocking was one of the procedures employed to minimize recall from short-term, nonsemantic memory. In this design, sentence set and list were between-subjects factors; modifier type and block position were within-subjects factors.

Procedure. The sentences were presented to groups of subjects in mimeographed booklets, one sentence per page. There were eight different random orders of items within blocks in the study booklet. The subjects

were given 10 seconds to study each sentence, paced by tape recorded "beeps." A filler task, to further reduce recall from short-term memory, the Surface Development Test (French, Ekstrom, & Price, 1963), was administered for five minutes after the study trial. Next came a successive recognition task. The 20 target and 20 distractor items were intermixed in a mimeographed booklet, one phrase per page, in a total of eight different random orders within block. The distractor items for any one group were the target items for subjects who had received another set and list. In addition to circling "old" or "new" for each item, subjects rated confidence on a scale from 1 (just guessing) to 5 (completely certain I'm right).

The final task was an unannounced recall test, which consisted of a booklet with one subject noun phrase per page. There were eight different random orders of phrases within blocks. The booklet contained only phrases from the study list, which was a fact subjects were told. The instructions emphasized that while a subject should try to recall each sentence verbatim, "if you do not remember the exact words of the sentence, but do remember the meaning, write down a sentence as close to the original as possible." Both the recognition and recall tests were subject paced.

Results

Recognition. Table 1 presents several measures of recognition performance. Minimum quasi F ratios calculated for hits, false alarms, and corrected recognition all showed significant ($\alpha = .01$ for these and all other tests reported in this paper) effects for type of modifier. For instance, for the corrected recognition measure, $\min F'(1,81) = 34.2$.

There were neither main effects or interactions involving set, list, or block position. The old/new data were pooled to obtain the d' and log β values which appear in Table 1. A more elaborate signal detection analysis is presented later.

Insert Table 1 about here

Recall. The mean proportions recalled at four levels of scoring appear in Table 2. Level I entailed a verbatim reproduction except for abbreviations and contractions. Level II scoring was also verbatim, but permitted changes of articles, auxiliaries, number, tense, and meaning-preserving changes of prepositions and word order. Level III allowed substitution of synonyms, hyponyms, close superordinates, and close cohyponyms for the substantive words of the original. At Level IV the reproduction was unmistakably derived from the presented sentence, but the meaning was distorted in some way. The score at any level included all sentences which were counted correct at preceding levels. In previous research using this scoring scheme, raters have proved very reliable (Anderson & Ortony, 1975).

Type of modifier had a significant effect on recall no matter what the level of scoring. At Level III, which is gist or substance recall, $\min F'(1,79) = 31.2$. There were no main or interaction effects involving set, list, or block position.

Insert Table 2 about here

Conditional recall. In Table 3 are mean proportions recalled given recognition. Once again, type of modifier was significant at every level

of scoring. For conditional substance recall, min $F'(1,81) = 25.8$. As in every other set of analyses, there were no effects due to set, list, or block position.

Insert Table 3 about here

Similarity and comprehensibility. The concrete and redundant phrases were rated equally discriminable from the distractors, $t(39) = .66$. However, the sentences containing concrete adjectives were judged significantly more comprehensible than the parallel sentences containing redundant adjectives, $t(39) = 5.35$.

The correlations of similarity rating, comprehensibility rating, and modifier concreteness with corrected recognition were -.16, .09, and .53, respectively. The multiple correlation was just .56, so it is apparent that most of the variance was due to concreteness.

Similarity, comprehensibility, and concreteness correlated .03, .33, and .48 with conditional substance (Level III) recall. The multiple correlation was .50. Only modifier concreteness had a significant regression coefficient. Approaching the matter in another way, there was still a significant correlation between concreteness and conditional recall when comprehensibility was partialed out.

Discussion

The present experiment provides evidence consistent with both the recognizability and the redintegration interpretations of the concreteness effect. Subjects were much better at recognizing the concrete noun phrases

than the redundant noun phrases. That this difference was not due to the failure to provide comparable distractor sets is suggested by the similarity rating data. As Wicker (1970) has previously argued a compelling explanation of the superior recognizability of concrete items is that they give rise to specific encodings. A concrete noun phrase refers to a narrower range of things than a redundant noun phrase. There is less latitude for variable encoding on different occasions.

The new finding in this research was the substantially higher level of recall given recognition for the sentences containing concrete adjectives compared to the ones containing redundant adjectives. A question might be raised about the conditional recall proportions since subjects had a looser criterion for calling redundant phrases "old." This fact probably led to an underestimate of the amount recalled given recognition, because the number of redundant phrases identified as old was larger than the number truly recognized, or at least there was a bias in that direction relative to the concrete phrases. Below we consider a model which discounts the effects of bias. In the mean time, note that the most conservative possible assumption is that for every false alarm there was a lucky hit. When conditional proportions are calculated dividing by hits minus false alarms instead of simply hits, there is still a substantial, significant difference favoring the concrete items, $F(1,73) = 13.2$. Thus, criterion bias cannot explain the difference in recall given recognition.

Next we will present a model which permits a more detailed analysis of the effects of modifier type. It is similar to the one suggested for a different paradigm by Bernbach and Kupchak (1972), and it borrows from

signal detection theory (Swets, Tanner, & Birdsall, 1961). For ease of exposition, just the case in which the subject noun phrase is the cue and the predicate is the response, the arrangement in this experiment, will be discussed.

Assume that a certain proportion, s, of the presented sentences enter a stored state, S, while the rest remain in an unstored state, N. It is important to note that State N effectively includes sentences whose subject noun phrases are encoded differently at time of initial presentation and time of test, as well as sentences which are not stored at all. The predicates of sentences in State N are never recalled, and the subject noun phrases from these sentences are indistinguishable from distractors. On the other hand, a certain proportion, r, of the predicates from sentences in State S are recalled, and the subject noun phrases from sentences in State S can be distinguished from distractors to some extent as being "old." The efficiency of this recognition process together with the probabilities s and r, constitute the parameters of the model, all of which may be estimated from the data at hand.

Consider the process by which people distinguish the subject noun phrases of sentences in State S from the distractors. Since all sentences whose predicates are recalled must be in State S, the ROC (relative operating characteristic) for recalled items alone appropriately represents the process. To construct this ROC, we derived a score, C, for each judgement, reflecting the extent to which the subject believed the subject noun phrase was old. For items judged "old," C was simply the confidence rating; for items judged "new," C was the negative of the confidence rating. Then,

for each possible value, x, of C, we obtained the proportion of items recalled at Scoring Level III for which C exceeded x, and the proportion of distractors for which C exceeded x. Figure 1 shows the former values plotted as a function of the latter on normal-normal coordinates.

Insert Figure 1 about here

According to the classical theory of signal detectability, subjects base their judgements on the value of a random variable, X, representing evidence that the item is old. If X is normally distributed for the distractors and items in State N, and normally distributed with the same variance for items in State S, and if C is a nondecreasing function of X, then the ROC plotted on normal-normal paper should be linear with a slope of 1. Furthermore, the efficiency of the recognition process can be represented by a single parameter:

$$d^* = \frac{E(X|S) - (E(X|N)}{\sigma_{X|N}}$$

Using the overall proportions of "old" responses to recalled and new items (the points indicated with circles in Figure 1) the estimates of d^* for concrete and redundant phrases are 3.62 and 2.39, respectively. The ROCs in Figure 1 are the ones predicted on the basis of these estimates. Statistical considerations aside, the fit of the model would appear to be acceptable, at least in the middle confidence range.

The other two parameters of interest are s, the storage rate, and r, the retrieval rate for stored sentences. The former parameter is reflected

in the difference between the ROC for recalled (i.e. stored) items and that for all presented items (both stored and not stored). Since d' = $[E(X|old) - E(X|N)]/\sigma_{X|N}$, it is not difficult to show that $d' = sd^*$. Dividing the estimates of d' given in Table 1 by the estimates of d^* given above, the estimate of s is .72 for concrete items and .62 for redundant items. The overall probability of recall is sr, the probability that an item will be both stored and recalled. Estimates of r can therefore be obtained by dividing the observed recall probabilities (given in Table 2, Column III) by the estimates of s. The resulting estimates of r are .55 for concrete items and .29 for redundant items.

Statistical evaluation of the effects of modifier type on the parameters of the model is quite difficult. Sample sizes were too small to allow for parameter estimates on a subject-by-subject basis. Instead, we used a statistical procedure described in Mosteller and Tukey (1968) known as the "jacknife." The jacknife involves developing a pseudoestimate for each case which reflects that case's contribution to the estimate derived from the entire set of n cases. The pseudoestimate for the ith case is, $y_{*i} = ny_{all} - (n - 1)y_{(i)}$, where y_{all} is the parameter estimate based on the full set of data, and $y_{(i)}$ is the estimate based on the data set with the ith case removed. (To get some intuitive feel for the behavior of these pseudoestimates, note that the pseudoestimates are the raw scores themselves in the case where y is a sample mean.)

There were a few occasions where hit rates were 1 or false alarm rates were 0. In these cases we substituted values of $k/(k + 1)$ and $1/(k + 1)$, respectively, k being the number of cases involved in the proportion. Such

a procedure was necessary to obtain finite normal deviates upon which to base estimates of d* and d'. The average pseudoestimates of d*, s, and r for concrete items were 3.52, .75, and .52, respectively. The corresponding averages for redundant items were 2.51, .51, and .32. Analyses of variance showed that modifier type had a significant effect on all three variables. Thus, assuming that the jacknife is appropriate here, the differences in parameter estimates were not due to chance.

The calculations based on the model of sentence memory further support the conclusions advanced earlier. The fact that s was lower for redundant than concrete items is consistent with the notion that redundant modifiers are more likely to be encoded variably from occasion to occasion, though it could also mean that sentences containing redundant modifiers are less likely to be stored at all. Especially telling is the d* difference, which indicates that there is poorer recognition of redundant subject noun phrases even when just those items which were recalled are considered. The one reasonable interpretation of this fact is that redundant phrases are likely to receive different interpretations at different times in different contexts. The higher mean value of r for concrete items is exactly what would be expected from the redintegration hypothesis.

Recently, Thorndyke (1975) has reported a study which he interprets as disconfirming the holistic integration hypothesis. He manipulated verb imageability in simple subject-verb-object sentences and tested recall using the subject, verb, or object as a cue. The finding was that verb imageability had an effect only when the verb was part of the to-be-recalled target, not when the verb was the cue. He argued that these results are

inconsistent with the integration hypothesis because if high imagery verbs lead to more integrated representations, then they should be better cues. He further argued that high imageability simply leads to better retention of the verb itself, discrete from any effect on the rest of the sentence. These results are, of course, exactly contrary to the striking concretization effect observed in this study and the previous one (Anderson, 1974). Thorndyke's study is also inconsistent with studies showing larger stimulus than response effects in noun pair learning (cf. Paivio & Yarmey, 1966). The contrast between Thorndyke's findings and those of other investigators may be due to a difference in how imageability affects nouns and verbs in sentences (cf. Yuille & Holyoak, 1974), perhaps because of the apparently lesser role of verbs in sentence storage (cf. Bobrow, 1970). Furthermore, Thios (1974) has reported results which are the precise opposite of Thorndyke's. Thios found that concrete verbs were better cues in a cued recall task than more general verbs, but did not eventuate in better recall when the subjects or objects were the cues, provided that synonyms substituted for the verbs were scored as correct. Obviously the effects of concretizing verbs is a matter which awaits future resolution.

Johnson, Bransford, Nyberg, and Cleary (1972) and Pezdek and Royer (1974) have argued that some, if not all, of the effects on concreteness can be explained in terms of the greater comprehensibility of concrete language. In the present case, though the concrete sentences were rated as more comprehensible than the redundant sentences, the comprehensibility ratings accounted for little variance in either recognition or conditional recall. Thus, we must conclude that concreteness has effects apart from

any it has on comprehensibility. This conclusion should not be pushed too far, however, for it is obvious that concreteness will generally contribute to comprehensibility. A sentence filled with words whose referents are unclear is by definition vague and, therefore, less than fully comprehensible. Pezdek and Royer (1974) showed that abstract sentences, such as The foreign faith aroused an enduring interest, will be given a semantic interpretation in an appropriate context. They are surely right that the context was helpful because it aided comprehension. However, the reason comprehension became more probable should not be overlooked. In the case of the sentence about the foreign faith, the facilitative context involved a Hindu girl talking about her religion to a class of fascinated American students. The context improved comprehension because it provided for a concrete instantiation of words whose referents were otherwise indeterminate (cf. Anderson & McGaw, 1973).

We speak of "concrete words," but this is potentially confusing shorthand. It is things and events which are concrete, and it is the specificity of mental representations which relates to comprehension. So-called "concrete words" usually have unambiguous referents and they usually permit the construction of instantiated mental representations, but not always. There are sentences composed of concrete words, such as The notes were sour because the seams split, which are incomprehensible to most people, unless they are provided with an illuminating context, in this case one involving bagpipes (Bransford & McCarrell, 1974). Again, the obstacle to comprehension is the difficulty in discovering what specifically the words are about.

In conclusion, the present research suggests that there are two distinct factors which make concrete stimuli efficacious conceptual pegs. First, concrete cues are identified with greater accuracy than abstract cues, most probably because they tend to be encoded the same way on each occasion. Second, concrete cues have greater power to reinstate the whole idea than less denotatively specific cues.

References

Anderson, R. C. Concretization and sentence learning. Journal of Educational Psychology, 1974, 66, 179-183.

Anderson, R. C., & McGaw, B. On the representation of the meanings of general terms. Journal of Experimental Psychology, 1973, 101, 301-306.

Anderson, R. C., & Ortony, A. On putting apples into bottles: A problem of polysemy. Cognitive Psychology, 1975, 7, 167-180.

Begg, I. Recall of meaningful phrases. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 431-439.

Bernbach, H. A., & Kupchak, P. G. Recognition and recall in short-term memory. Journal of Mathematical Psychology, 1972, 9, 237-242.

Bobrow, S. A. Memory for words in sentences. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 363-372.

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), Cognition in learning and memory. New York: Wiley, 1972.

Bransford, J. D., & McCarrell, N. S. A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W. B. Weimer and D. S. Palermo (Eds.), Cognition and the symbolic processes. Potomac, Maryland: Erlbaum, 1974.

Davidson, R. E. Mediation and ability in paired-associate learning. Journal of Educational Psychology, 1964, 55, 352-356.

Epstein, W., Rock, I., & Zuckerman, C. B. Meaning and familiarity in associative learning. Psychological Monographs, 1960, 74 (4), Whole No. 491.

French, J. W., Ekstrom, R. B., & Price, L. A. Kit of reference tests for cognitive factors. Princeton, N. J.: Educational Testing Service, 1963.

Horowitz, L. M., & Prytulak, L. S. Redintegrative memory. Psychological Review, 1969, 76, 519-561.

Johnson, M. K., Bransford, J. D., Nyberg, S. E., & Cleary, J. J. Comprehension factors in interpreting memory for abstract and concrete sentences. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 451-454.

Kintsch, W. The representation of meaning in memory. New York: John Wiley & Sons, 1974.

Martin, E. Stimulus recognition in paired associate learning. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 272-276.

Martin, E. Stimulus meaningfulness and paired associate transfer. Psychological Review, 1968, 75, 421-441.

Mosteller, F., & Tukey, J. W. Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology (Vol. 2, 2nd ed.). Reading, Mass.: Addison-Wesley, 1968.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart, and Winston, 1971.

Paivio, A. A theoretical analysis of the role of imagery in learning and memory. In P. W. Sheehan (Ed.), The function and nature of imagery. New York: Academic Press, 1972.

Paivio, A., & Yarmey, A. D. Pictures versus words as stimuli and responses in paired-associate learning. Psychonomic Science, 1966, 5, 235-236.

Pezdek, K., & Royer, J. M. The role of comprehension in learning concrete and abstract sentences. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 551-558.

Reese, H. W. Imagery in paired-associate learning in children. Journal of Experimental Child Psychology, 1965, 2, 290-296.

Rohwer, W. D., Jr. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. Final Report, 1967. Project 5-0605, Contract No. OE-6-10-273, U. S. Office of Education.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

Thios, S. J. Memory for general and specific sentences. Memory and Cognition, 1975, 3, 75-97.

Thorndyke, P. W. Conceptual complexity and imagery in comprehension and memory. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 359-368.

Tulving, E., & Thompson, D. M. Encoding specificity and retrieval processes in episodic memory. Psychological Review, 1973, 80, 352-373.

Wicker, F. W. On the locus of picture-word differences in paired associate learning. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 52-57.

Wicker, F. W., & Evertson, C. M. Prerecall and postrecall imagery ratings with pictoral and verbal stimuli in paired-associate learning. Journal of Educational Psychology, 1972, 92, 75-82.

Yuille, J. C., & Holyoak, K. Verb imagery and noun phrase concreteness in the recognition and recall of sentences. Canadian Journal of Psychology/Review of Canadian Psychology, 1974, 28, 359-370.

Table 1

Mean Performance on Recognition Measures

Type of modifier	Hits	False alarms	Corrected recognition	d'	log β
Concrete	.86	.08	.79	2.59	-.18
Redundant	.75	.21	.54	1.48	-.04

Table 2

Mean Proportions Recalled

	Level of scoring[a]			
Type of modifier	I	II	III	IV
Concrete	.17	.19	.40	.53
Redundant	.08	.10	.19	.27

[a]See text for description of scoring procedures.

Table 3

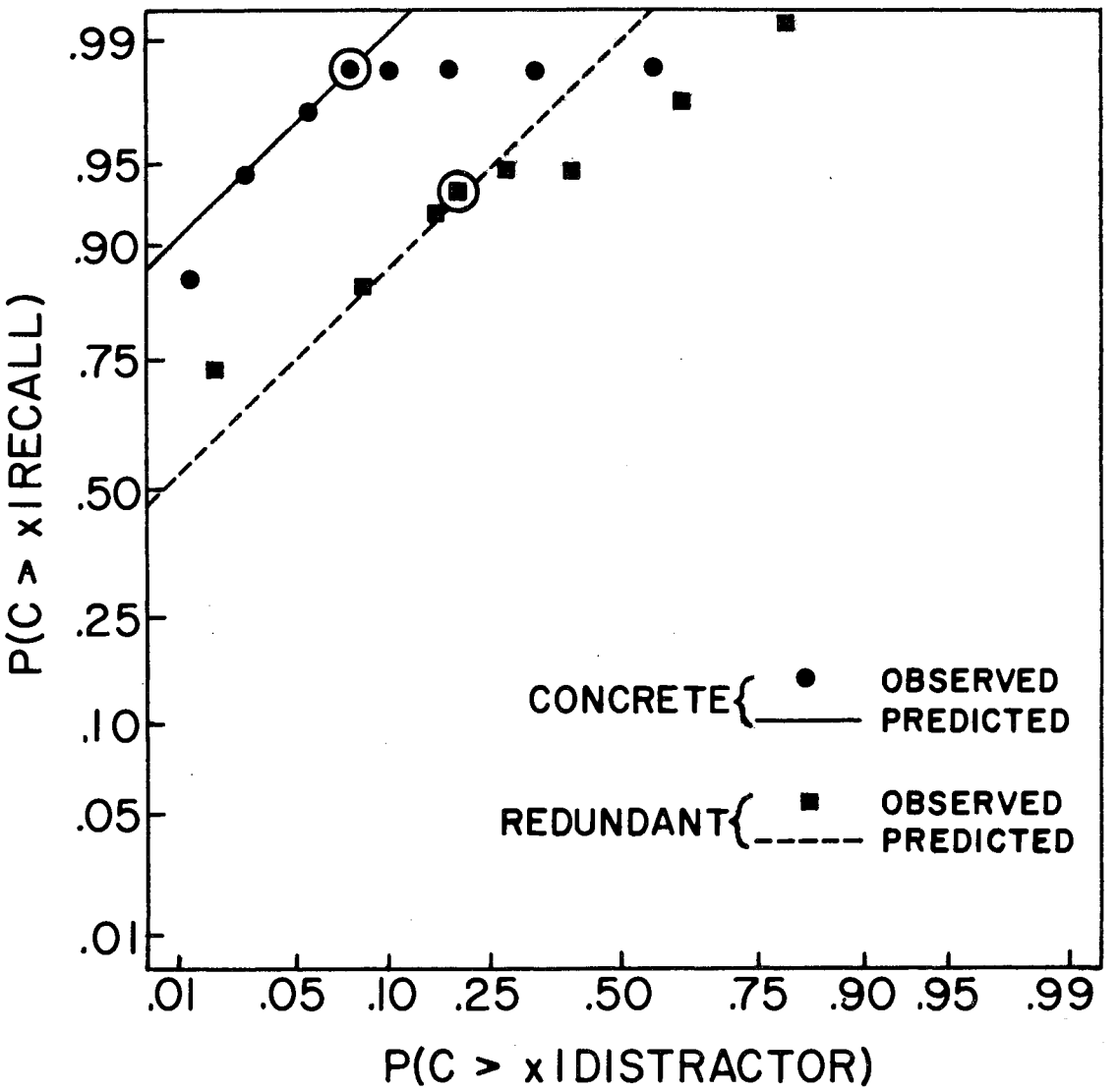
Mean Proportions Recalled Given Recognition

	Level of scoring[a]			
Type of modifier	I	II	III	IV
Concrete	.18	.21	.44	.59
Redundant	.09	.11	.22	.32

[a]See text for description of scoring procedures.

Figure Caption

Figure 1. ROCs for recalled items plotted on normal-normal paper as a function of modifier type. (Points indicated by circles are for $\underline{x} = -1$, i.e. for "old" responses.)


.99
.95
.90
.75
.50
.25
.10
.05
.01
P(C > x | RECALL)
.01
.05
.10
.25
.50
.75
.90
.95
.99
P(C > x | DISTRACTOR)
CONCRETE
OBSERVED
PREDICTED
REDUNDANT
OBSERVED
PREDICTED