

# Question-answering strategies and conceptual knowledge

MURRAY SINGER

*University of Manitoba, Winnipeg, Manitoba, Canada*

To test the proposal that question-answering strategy is independent of the type of searched representation, an experiment was designed to demonstrate the application of two strategies, direct retrieval and plausibility judgment, to the same representation. Subjects judged whether or not concept pairs belonged to the same category. The distractor ("no") items presented either similar (e.g., *wren bee*) or dissimilar (e.g., *wren pea*) concept pairs. Dissimilar *different* distractors should favor the plausibility strategy and result in relatively fast responses; whereas similar *different* distractors should require direct retrieval. The selected strategy was predicted to be extended to the *same* pairs, which were identical in the two conditions. Consistent with this analysis, "same" responses were faster in the dissimilar-*different* condition. Because all answers were based on conceptual knowledge, this outcome supports the proposal that answering strategy is independent of type of searched representation.

The study of question answering has led to the identification of different answering strategies. One can either *directly retrieve* a queried fact from memory, or *judge its plausibility* in the context of pertinent knowledge (Reder, 1982, 1987). However, it has recently been documented (Singer, in press) that answering strategy has sometimes been confounded with the type of representation to which the strategy is applied. In particular, some theorists have associated direct retrieval with the search of episodic representations, such as a discourse text base, and have associated plausibility judgment with the search of general knowledge structures, such as a script.

I argue, in contrast, that answering strategy is orthogonal to the type of searched representation; that is, any answering strategy may be applied to any representation (Singer, in press). Three types of evidence can be adduced to support this proposal. First, both intuition and previous research provide examples of the application of each strategy to both episodic and semantic representations. In this regard, if asked, "Who invented dynamite?" one has little alternative but to *retrieve* the requested information from *general knowledge* (Camp, Lachman, & Lachman, 1980). Second, empirical evidence confirms that both direct retrieval and plausibility judgment can be applied to both episodic representations (Reder, 1982, 1987) and general knowledge structures (Camp et al., 1980; Lorch, 1981). Third, "extrinsic" question factors, such as test delay, task instructions, and composition of

the test items, influence the selection of answering strategy (Reder, 1982, 1987; Reder & Anderson, 1980; Reder & Ross, 1983; Reder & Wible, 1984). As I have demonstrated (Singer, in press), these factors exert similar effects regardless of whether plausibility must be evaluated with reference to episodic or semantic representations.

The impact of test item composition on strategy selection is particularly relevant to the present study. This effect can be explained with reference to the materials of Reder and Anderson (1980). Suppose that one learns some facts about Laura's ski trip, such as that *Laura bought a pair of downhill skis at the Co-op*, and *Laura fell while skiing down the steepest stretch*. Then, if all distractor test questions are unrelated ones, such as *Did Laura think that the clowns were very funny?*, the strategy of plausibility judgment may be adopted: any fact that links Laura to skiing may readily be answered "yes." However, if the distractor items are related ones, such as the question *Did Laura groan at the price of good ski boots?* then it is necessary to *retrieve* the queried facts. Otherwise, one will incorrectly judge that the fact that *Laura groaned at the price of good ski boots* was in the learned set. Consistent with this analysis is the evidence that related distractors favor direct retrieval, whereas unrelated distractors support plausibility judgment (Lorch, 1981; Reder, 1987; Reder & Anderson, 1980; Yekovich & Walker, 1986).

In this paper, I describe an experiment designed to provide additional support for the proposal that different answering strategies may be applied to one type of representation. The experiment was focused on the representation of conceptual knowledge. The subjects were required to judge whether two concepts belonged to the same category or not (cf. Rips, Shoben, & Smith, 1973; Schaeffer & Wallace, 1970). For example, the answer for the concepts included in the phrase *geranium lettuce* is "different," because a geranium is a flower, whereas lettuce is a

---

This research was supported by Grant A9800 from the Natural Sciences and Engineering Research Council of Canada. The data of the experiment were presented at the annual meeting of the Canadian Psychological Association, Halifax, June 1989. I would like to thank Nancy Black, who conducted the experimental sessions. Requests for reprints should be addressed to Murray Singer, Department of Psychology, University of Manitoba, Winnipeg R3T 2N2, Canada.

vegetable. In the experiment, the concepts were judged with reference to three animal categories—mammal, bird, and insect—and three plant categories—flower, fruit, and vegetable.

Two lists were constructed. In one list, all *different* pairs presented two concepts from similar categories. For example, *cow grasshopper* is a *different* pair, but both concepts represent animals. In the second list, *different* pairs, such as *cow plum*, were constructed from dissimilar categories—namely, one animal and one plant. The two lists included the identical set of *same* pairs.

It is well documented that people need less time to correctly verify dissimilar *different* pairs as opposed to similar *different* pairs (Rips et al., 1973; Schaeffer & Wallace, 1970). The crucial hypothesis in the present experiment was that answering strategy would be influenced by composition of the test items. In particular, dissimilar *different* pairs would favor plausibility judgments and similar *different* pairs would support direct retrieval (see, e.g., Lorch, 1981; Reder & Anderson, 1980). The favored strategy would be applied to *same* pairs as well as *different* pairs. As a result, answer times for *same* test items on the dissimilar-*different* list would be faster than answer times for *same* test items on the similar-*different* list.

## METHOD

### Subjects

One hundred twenty male and female students in introductory psychology at the University of Manitoba participated in the experiment. All participants were native speakers of English who participated in partial fulfillment of a course requirement.

### Materials

The materials were two lists of word pairs. The words were drawn from three animal categories: mammal, bird, and insect; and three plant categories: flower, fruit, and vegetable. The words were selected from the category norms of Battig and Montague (1969). Eighteen words were chosen from each category, beginning with the word with the highest production frequency, and proceeding down the list. Terms mentioning their categories, such as *bluebird*, and those consisting of more than one word, such as *blue jay*, were excluded.

From these words, two lists of 72 word pairs were constructed. In one list, the *different* pairs came from similar categories (i.e., both animals or both plants), and in the other, the *different* pairs came from dissimilar categories. The *same* pairs were identical in the two lists. Six of the 18 words chosen from each category were randomly selected to form the *same* pairs. Each of these words was randomly paired with 2 of the 5 other *same* words from its category. This yielded six *same* pairs per category, for a total of 36 pairs across the six categories. Each half of each list included three *same* pairs from each category. (The complete lists are available from the author on request.)

All words in the *same* pairs appeared twice in the list. The alternative procedure of selecting additional words from each category to generate enough pairs would have had the undesired outcome of yielding items of very low category typicality (e.g., *emu* as an instance of bird). Because the main hypothesis focused on the comparison of identical *same* pairs in the two lists, rather than the comparison of *same* and *different* pairs, this procedure did not compromise the evaluation of the hypothesis.

Each *same* word appeared once in the first half of the list and once in the second. Furthermore, each *same* word functioned as the first member of one of its pairs and the second member of the other.

The *different* word pairs were constructed from the remaining 12 words from each category. In the similar-*different* list, each *different* word was randomly paired with a word from a different, but similar category: that is, animals with animals, and plants with plants. For example, a

mammal name could be paired with either a bird name or an insect name. This procedure linked 6 words from each category with a member of each of the other two similar categories. There were 36 such pairs in total.

In the dissimilar-*different* list, the 12 *different* words from each category were randomly paired in equal frequency with words from the three dissimilar categories. For example, four mammals were randomly paired with flower names, four with fruits, and four with vegetables. This procedure generated 36 *different* pairs. In both lists, all categories appeared equally often as the first member of *different* pairs.

To construct the lists, the 36 *same* pairs were first randomly assigned to a list position from 1 to 72. Each half of the list included three *same* pairs from each category. Each *same* pair occupied the identical position in the two lists. In each list, the *different* pairs were randomly assigned to the remaining 36 positions. Each possible *different* pairing of categories (e.g., *mammal-fruit* in the dissimilar-*different* list) appeared equally often in both halves of the list. No more than three *same* pairs or *different* pairs occurred consecutively.

Each list was preceded by a block of 24 practice trials constructed from eight additional words from each category. By necessity, the practice words either had lower typicality than the experimental words or violated one of the restrictions on word selection. The practice words were arranged to form two *same* and two *different* pairs representing each category. The *different* practice pairs reflected the list that they preceded, similar-*different* or dissimilar-*different*. The practice lists were constructed following the same principles as the experimental lists, except that they did not consist of two symmetric halves.

### Procedure

The experimental sessions were conducted with groups of 1 to 4 individuals. Each participant was tested at a station consisting of a monochrome video monitor positioned 22 cm from the participant, and a response panel. The experimental events were controlled by a micro-computer.

The similar-*different* list and the dissimilar-*different* list were viewed by 61 and 59 randomly assigned participants, respectively. The participants were told that they would see word pairs constructed from the categories mammal, bird, insect, flower, fruit, and vegetable. They were instructed to answer "same" if both words represented the same category, and "different" if they did not. The participants answered by pressing response buttons labeled "same" and "different." They were randomly assigned to use either their left index finger for "same" and their right finger for "different," or vice versa.

On each trial, a fixation point was displayed on the screen for .5 sec, followed by a word pair. The two words appeared in capital letters. They were arranged side by side, with the first letter of the right-hand word 22 character spaces from the first letter of the left-hand word. The participants had 4 sec in which to respond "same" or "different," after which the word pair disappeared. They were instructed to respond as quickly as they could without sacrificing accuracy. Trials on which no answer was registered were counted as errors. The responses and response times were recorded automatically by the computer.

## RESULTS

Table 1 shows the mean correct answer times and proportion of errors as a function of the type of *different* item and response. Analysis of variance (ANOVA) was applied

Table 1  
Mean Correct Answer Times (in Milliseconds) as a Function of  
Type of *Different* Item and Response

Type of <i>Different</i>	Response			
	"Same"		"Different"	
	<i>M</i>	<i>PE</i>	<i>M</i>	<i>PE</i>
Similar	1,756	.083	1,944	.089
Dissimilar	1,589	.043	1,662	.035

Note—PE = proportions of errors.

to these scores, alternately treating participants ( $F_1$ ) and items ( $F_2$ ) as the random variable. In the participants-random analysis, type of *different* item was a between-subjects variable and response was a within-subjects variable. In the items-random analysis, type of *different* item and response were both between-items variables. An alpha level of .05 was used throughout.

Answer times for "same" responses were 131 msec faster than those for "different" responses [ $F_1(1,118) = 85.5$ ,  $MS_e = 13,770$ ,  $F_2(1,140) = 11.6$ ,  $MS_e = 52,868$ ]. Items in the dissimilar-*different* list were answered 225 msec faster than those in the similar-*different* list [ $F_1(1,118) = 16.3$ ,  $MS_e = 185,555$ ,  $F_2(1,140) = 34.4$ ,  $MS_e = 52,868$ ]. The response  $\times$  type-of-*different*-item interaction was significant with subjects-random [ $F_1(1,118) = 20.8$ ,  $MS_e = 13,770$ ], but not with items-random [ $F_2(1,140) = 2.25$ ,  $MS_e = 52,868$ ,  $p = .14$ ]. This marginal interaction reflects the fact that the latency advantage of the dissimilar-*different* items was somewhat greater for the "different" response than for the "same" response.

Of greatest concern for the present purposes was the fact that "same" answer times were 167 msec faster in the dissimilar-*different* condition than in the similar-*different* condition. Tests of simple main effects revealed that this difference was significant [ $F_1(1,118) = 7.84$ ,  $MS_e = 92,320$ ,  $F_2(1,140) = 9.52$ ,  $MS_e = 52,868$ ].

ANOVAs were also applied to the error proportions, using the same designs as those for the answer-time analyses. Only the main effect of type of *different* item reached significance [ $F(1,118) = 18.1$ ,  $MS_e = 56.1$ ,  $F(1,140) = 12.6$ ,  $MS_e = 50.2$ ]. This reflected error rates of 8.6% and 4.4% for the similar-*different* and dissimilar-*different* conditions, respectively.

## DISCUSSION

The answer-time data supported the main hypothesis: namely, answer times were faster for the *same* pairs in the dissimilar-*different* condition than in the similar-*different* condition. The outcome is consistent with previous evidence that the nature of distractor test items affects answering strategy (Lorch, 1981; Reder, 1987). Following this interpretation, the present findings can be explained as follows. First, dissimilar-*different* items permit the comparison of concepts through the assessment of their semantic overlap, a plausibility judgment. In contrast, similar-*different* distractors require the retrieval and comparison of their category relations. Second, the evaluation of semantic overlap is faster than the retrieval of categorical information (Smith, Shoben, & Rips, 1974). Third, the strategy favored by the distractor type is applied to all test questions. As a result, answer time is faster for *same* items in the dissimilar-*different* condition than it is for *same* items in the similar-*different* condition.

This analysis can be articulated further with reference to the two-stage model of the comparison of concept pairs (Smith et al., 1974). At the first stage, the semantic features of the concepts are compared, yielding an index of concept similarity,  $x$ . If  $x$  falls below a low criterion,  $c_0$ , or above a high criterion,  $c_1$ , then the subject can register fast responses of "no" or "yes," respectively. If  $x$  falls between  $c_0$  and  $c_1$ , then, in a second stage of search, a more painstaking examination of the relevant category relations is made. The comparisons requiring the second stage of search take longer (Smith et al., 1974).

In the framework of this model, dissimilar-*different* distractors might permit  $c_0$  and  $c_1$  to be set close together, toward the middle of the distribution of similarity values. The similarity index of the test concept pairs would then frequently fall in the large tails of the distribution, lowering

the reliance on the second stage of search, and so producing fast judgment times. The presence of similar-*different* distractors, in contrast, would require that the criteria have relatively extreme values. Otherwise, the similarity index of similar-*different* pairs, such as *wren bee*, might exceed the high criterion,  $c_1$ . This would result in an incorrect "same" response. With extreme criteria, however, the similarity index of some *same* pairs, such as *cow mouse*, would also not exceed  $c_1$ . Therefore, the impact of extreme values for the criteria is that second-stage search would frequently be needed for both *same* and *different* pairs. The corresponding judgment times would be slow.

Although the present experiment was not designed to distinguish this analysis from other possibilities, this account illustrates how the selection of answering strategies might be explained in terms of well-defined mechanisms of memory search. Furthermore, this analysis should not be interpreted as indicating that the present results were due "merely" to a criterion shift. Rather, the proposed adjustment of the similarity criteria would directly influence one's relative reliance on semantic overlap as opposed to category evaluation, the comparisons respectively proposed to constitute plausibility judgment and direct retrieval in the present task (Lorch, 1981; Reder, 1987).

The present results bear some similarities to those of Lorch (1981). His participants verified property statements, such as *A rabbit has fur*. Lorch manipulated the composition of the distractor test items. His Experiment 1 included only related false items, such as *A sword HAS a weapon* (capitals added). Experiment 2 added some "unrelated" false fillers, such as *A garage has seeds*. Lorch predicted that the unrelated false fillers would favor the adoption of a plausibility strategy, according to which decisions would be based on the semantic overlap of the concept pair (see also Reder, 1987; Reder & Ross, 1983; Reder & Wible, 1984).

The *relatedness effect* measured by Lorch (1981) supported his prediction. This term refers to the observation that it takes longer to correctly verify highly related false statements, such as *A dolphin is a fish*, than it does to correctly verify weakly related ones, such as *An oak is a fish* (see Rips et al., 1973; Schaeffer & Wallace, 1970). Lorch (1981) detected a positive relatedness effect of 80 msec in the presence of unrelated false fillers (his Experiment 2), but a negative effect of -51 msec in their absence. He concluded that when all false items were related ones, the subjects had to retrieve the category relations linking the concepts, in order to answer correctly.

However, in some ways, Lorch's (1981) measurements did not reflect the effect of test composition on answer strategy. First, unlike the present experiment, the answer strategies that Lorch detected for false property statements were not extended to true property statements. For his true items, the relatedness effect was -110 msec under the condition of direct retrieval (unrelated fillers absent) and -74 msec for plausibility judgments (unrelated fillers present). In other words, answer times for true property statements of high relatedness were faster than those for true property statements of low relatedness, regardless of test composition. Second, as in the present experiment, Lorch's materials included statements that required an evaluation of category relations—for example, *A cobra is a snake*. However, in all of Lorch's conditions, the relatedness effect for category statements was negative. Therefore, his category materials did not reflect the manipulation of test composition. Finally, Lorch measured answer times for highly related true category statements of 1,021 and 1,018 msec in the absence and presence of unrelated false fillers, respectively. In the present experiment, the answer times for the *same* items, comparable to Lorch's highly related true category statements, varied with list composition: Answer time was 167 msec faster in the presence of dissimilar-*different* items than in the presence of similar-*different* items.

One reason that Lorch's data did not consistently reflect the use of different answering strategies may be that his unrelated fillers (Lorch, 1981, Experiment 2) were added to the related false items rather than substituted for them. The presence of both unrelated and related false items would create a weaker preference for the plausibility strategy, possibly resulting in strategy mixing (Reder, 1982).

In conclusion, in the present experiment, people's application of different answering strategies to their long-term conceptual knowledge was studied. Because no experiment-specific message representation was available, both strategies had to be applied to structures of general knowledge. Likewise,

the results of Reder (1982, 1987) reflect the application of both strategies to an episodic memory structure—namely, a text base. These findings support the proposal that question-answering strategy is orthogonal to the type of representation that is searched.

#### REFERENCES

- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monographs*, **80**(3, Pt. 2).
- CAMP, C. J., LACHMAN, J. L., & LACHMAN, R. (1980). Evidence for direct-access and inferential retrieval in question-answering. *Journal of Verbal Learning & Verbal Behavior*, **19**, 583-596.
- LORCH, R. F., JR. (1981). Effects of relation strength and semantic overlap on retrieval and comparison processes during sentence verification. *Journal of Verbal Learning & Verbal Behavior*, **20**, 593-611.
- REDER, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, **89**, 250-280.
- REDER, L. M. (1987). Strategy-selection in question answering. *Cognitive Psychology*, **19**, 90-134.
- REDER, L. M., & ANDERSON, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology*, **12**, 447-472.
- REDER, L. M., & ROSS, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 55-72.
- REDER, L. M., & WIBLE, C. (1984). Strategy use in question-answering: Memory strength and task constraints on fan effects. *Memory & Cognition*, **12**, 411-419.
- RIPS, L. J., SHOBE, E. J., & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, **12**, 1-20.
- SCHAEFFER, B., & WALLACE, R. (1970). The comparison of word meanings. *Journal of Experimental Psychology*, **86**, 144-152.
- SINGER, M. (in press). Independence of question-answering strategy and searched representation. *Memory & Cognition*.
- SMITH, E. E., SHOBE, E. J., & RIPS, L. J. (1974). Structure and process in semantic memory: A feature model of semantic decisions. *Psychological Review*, **81**, 214-241.
- YEKOVICH, F. R., & WALKER, C. H. (1986). Retrieval of scripted concepts. *Journal of Memory & Language*, **25**, 627-644.

(Manuscript received September 15, 1990.)