

Reliability of retrieving information from knowledge structures in memory: Scripts

FRANCIS S. BELLEZZA
Ohio University, Athens, Ohio

Research has shown that the reliability of retrieving information from knowledge structures in memory is surprisingly low. For many knowledge structures only about half the information retrieved in one session is the same information retrieved in a second session 1 week later. A number of theorists have argued that knowledge structures known as scripts are particularly important and well-organized knowledge structures. The purpose of the two experiments reported here was to determine the reliability of retrieving script information from memory. In Experiment 1, 10 scripts were tested. It was found that the reliability of retrieving information from scripts had a mean value of .50, which is no larger a value than values found for many other knowledge structures. On the other hand, any script information recalled in both sessions was almost always recalled in the same relative order. The mean output-output correlation was .92. The results of Experiment 2 demonstrated that the strict ordering of script actions by the subjects was not the result of ordering instructions.

It is very important to cognitive scientists to determine the manner in which knowledge is represented in memory. This is because the way knowledge is organized influences the way it is used (Charniak & McDermott, 1985, chap. 1; Gardner, 1985). Investigators have proposed that *knowledge structures* in memory represent information about various real-world objects and domains that have been variously identified as natural categories, mental images, memory schemas, prototypes, and so forth (Alba & Hasher, 1983; Anderson, 1978; Mandler, 1984; Minsky, 1975; Norman, 1982; Paivio, 1986; Rumelhart, 1980). Various methods have been used to determine the content and organization of these knowledge structures (Wyer & Srull, 1984), but little research has been done on how reliably or completely information within these knowledge structures can be accessed and retrieved. Bellezza (1984a) proposed a procedure by which the reliability of retrieving verbal information from these knowledge structures can be estimated. By having subjects try to retrieve the same information from memory in two different sessions using open-ended questions, overlap scores can be computed indicating how much of the retrieved information was the same in each session.

For a wide variety of knowledge structures, only about half the information recalled in one session is recalled under identical conditions 1 week later. Bellezza (1987) reviewed the results of some of these studies. Low reliabilities have been obtained even though the retrieval cues used can be considered sufficient. That is, subjects were instructed to retrieve information from knowledge structures identified using their commonly used labels that ap-

peared to specify these structures without ambiguity (Bellezza, 1987).

One type of knowledge structure that has not been tested as to how reliably information can be retrieved from it is the script (Bellezza & Bower, 1982; Bower, Black, & Turner, 1979; Schank & Abelson, 1977). Scripts are used to comprehend spoken and written language, understand other people's behavior, and guide one's own behavior (Schank & Abelson, 1977). Because scripts are used to guide overt behavior, it may be that the reliability of retrieving information from scripts is better than that from other types of knowledge structures.

EXPERIMENT 1

Method

Subjects. Thirty-three undergraduates enrolled in introductory psychology courses at Ohio University participated for extra course credit.

Procedure. The procedure used was that proposed by Bellezza (1984a). All subjects were tested as a group in two sessions separated by 1 week. The subjects were given an example of a script including 20 typical activities involved in going to a disco on Saturday night. The subjects were then given 4 min to write down the activities that are involved for each of 10 common scripts. The instructions differed in two ways from the instructions given in previous studies investigating the reliability of retrieval of information from memory. First, the subjects were told to write down at least 20 activities for each script. Second, the subjects were told to write down the activities in the same order as they would perform them. The 10 scripts used were those used by Bower et al. (1979). In each session the scripts were tested in a different random order.

Results

Amount of information retrieved. Table 1 shows the mean number of activities recalled for each of the 10 scripts. An analysis of variance (ANOVA) showed that session was a significant effect [$F(1,32) = 10.22, p < .005, MSe = 25.53$]. More items were generated during Session 2 (18.6) than during Session 1 (17.4). Type of

I would like to thank Mindy Morrison for her help in collecting and scoring the data. Requests for reprints may be sent to Francis S. Bellezza, Department of Psychology, Ohio University, Athens, OH 45701.

Table 1
Characteristics of the Script Information Retrieved from Memory

Script	Number of Items Retrieved	Within-Subject Overlap	Output-Output Correlations
Making Coffee	16.1	.52	.88
Visiting the Doctor	18.7	.51	.96
Going to a Football Game	16.1	.46	.91
Going Shopping at the Grocery Store	17.2	.45	.93
Going to a Class Lecture	16.5	.39	.89
Getting Up in the Morning	19.0	.58	.95
Going to a Movie	19.3	.52	.97
Going to a Party	17.8	.51	.88
Going to Eat at a Restaurant	19.8	.54	.92
Going Swimming	19.3	.57	.96

script was also a significant effect [$F(9,288) = 13.25$, $p < .001$, $MSe = 10.05$]. A Tukey HSD test (Kirk, 1982) showed that any two script means differing by more than 2.4 activities were significantly different at the .05 level. The largest number of activities were retrieved for the script "Going to Eat at a Restaurant" (19.8) and the fewest for the script "Making Coffee" (16.1). The script \times session interaction was also significant [$F(9,288) = 5.83$, $p < .001$, $MSe = 4.53$]. The number of activities for the script "Making Coffee" decreased from 16.9 in Session 1 to 15.3 in Session 2. The number of activities for the script "Going Shopping at the Grocery Store" increased from 15.7 in Session 1 to 18.7 in Session 2. There were no apparent reasons for these differences.

Reliability of retrieval. The proportion of activities from each script that were the same in each session is shown in Table 1. The mean within-subject overlap score across scripts was .50, which is a value similar to that found for many other knowledge structures (Bellezza, 1987). An ANOVA showed type of script to be a significant factor affecting the overlap scores [$F(9,288) = 5.94$, $p < .001$, $MSe = .019$]. A Tukey HSD test indicated that any two overlap means differing by .11 or greater were significantly different. Hence, information in the scripts "Going Swimming" and "Getting Up in the Morning" was more reliably retrieved than information in the scripts "Going to a Football Game," "Going Shopping at the Grocery Store," and "Going to a Classroom Lecture." It is not clear why the reliability of retrieval for these various scripts should differ.

It appears from Table 1 that as the number of script actions increased, the overlap score increased. To determine if this effect was reliable, a correlation was computed for each subject between the number of activities generated for the 10 scripts and their overlap scores. An ANOVA on the Fisher z -transformations of these correlations showed the mean value of .27 to be significantly greater than zero [$F(1,32) = 9.03$, $p < .001$, $MSe = .29$]. This result is interesting because Bellezza (1984a), in investigating the reliability of recalling category instances from common categories, found the corresponding correlation to be $-.39$, which was significantly less than zero. Bellezza suggested that retrieval of instances

for a category occurred in a somewhat haphazard manner, and that the larger the category, then the less likely the same instances would be retrieved in each session. The positive correlation found here for retrieving script actions obviously demands a different explanation. It is possible that the more familiar a subject is with a particular script, then the more information he/she will retrieve from that script and also the more organized, and hence reliable, that retrieval will be.

Organization of recall. One of the characteristics of script actions is that they occur in a particular order (Schank & Abelson, 1977). In a restaurant script it would seem peculiar to have someone eat dessert, then eat a salad, and then look at the menu. The information in scripts seems to be retrieved in the order in which the script actions are performed (Barsalou & Sewell, 1985).

The consistency of order of recall was tested in the present experiment by computing an output-output correlation between script actions that were recalled by each subject for each script in both Session 1 and Session 2. To do this, the actions common to both recalls were ranked by order of output for each session separately. These ranks were correlated using Spearman's rank-correlation coefficient. These output-output correlations are shown in Table 1. The mean rank correlation was .92, which was significantly greater than zero [$F(1,32) = 8,152$, $p < .001$, $MSe = .035$]. Also, there were no significant differences among the output-output correlations of the 10 scripts. All the correlations were high in value.

Discussion

The primary purpose of this experiment was to determine the reliability of retrieving script information from memory as measured by content overlap scores (Bellezza, 1984a). The mean value of .50 for all scripts is not very high and not very different from the reliability of retrieving other types of information from knowledge structures in memory (Bellezza, 1987).

The most interesting result of this experiment was the strict ordering of script actions from one session to the next. The overlap scores indicate that the same script activities have only a 50% chance of being repeated, but the output-output correlations indicate that whenever a script action is retrieved it tends strongly to be in the same position relative to other retrieved script actions. However, the possibility remains that in the present experiment the instructions to recall the script actions in order contributed to the very high output-output correlations obtained. The instructions adapted from Bower et al. (1979) stressed recalling the

script activities in the order in which they would naturally occur and also suggested recalling about 20 activities for each script. In Experiment 2 the instructions to recall script actions in their order of occurrence was eliminated.

EXPERIMENT 2

In Experiment 2 no instructions were given to subjects to order the retrieved information in any way, and no suggestion was made as to how many items should be retrieved for each script. Another consideration is that when retrieving information about many different scripts, subjects may develop a pattern of writing down the information from the scripts not in the order it is recalled but in the order that it should be performed. To prevent any such pattern, the knowledge structures tested in Experiment 2 involved only one script (the restaurant script).

Method

Subjects. Forty-six undergraduates enrolled in introductory psychology courses at Ohio University participated for extra course credit.

Materials. Eight knowledge structures of various types were tested in the same order in two sessions that were 1 week apart. The knowledge structures were tested in the following order: one's good qualities (a self category; Bellezza, 1987), one's physical appearance (a self category; Bellezza, 1987), objects that could fall on one's head (an ad hoc category of Barsalou, 1983), Ronald Reagan's qualities (a famous person; Bellezza, 1984b), going to eat at a restaurant (a script), the qualities of one's best friend (a friend, Bellezza, 1984b), instances of fruits (a common category, Bellezza, 1984a), and one's goals in life (a self category).

Procedure. The procedure used was the same as in Experiment 1, except that no mention was made in the instructions or in the test booklets about recalling the information in any particular order. Also, subjects were given only 3 min for information retrieval from each knowledge structure rather than the 4 min given in Experiment 1.

Results

The results of Experiment 2 are summarized in Table 2.

Amount of information recalled. An ANOVA on the number of items recalled from each knowledge structure showed knowledge structure to be significant [$F(7,315) = 46.44, p < .001, MSe = 29.1$], as was the knowledge structure \times session interaction [$F(7,315) = 2.13, p < .05, MSe = 6.1$]. A Tukey HSD indicated that any knowledge structure means differing by more than 3.4

items were significantly different. Inspection of the means for the two sessions indicated that from Session 1 to Session 2 the number of items for the ad hoc category of objects that could fall on one's head increased from 17.4 to 18.7, whereas the number of attributes recalled about one's best friend decreased from 13.7 to 12.8. These represent the most extreme increases and decreases found and probably are the reason for the marginally significant interaction.

The amount of information retrieved from the knowledge structures varied somewhat from previous experiments, but the values were comparable. The number of actions retrieved for the restaurant script was 16.4, compared with 19.8 of Experiment 1. This decrease may have occurred because 3 min were provided for retrieval in Experiment 2, whereas 4 min were provided in Experiment 1. Also, in Experiment 2 no suggestion was made that at least 20 actions should be retrieved.

Reliability of retrieval. An ANOVA of the overlap scores showed knowledge structure to be significant [$F(7,315) = 35.17, p < .001, MSe = .021$]. A Tukey HSD test indicated that any overlap means differing by .09 or more were significantly different. Thus the overlap mean for fruits was significantly larger than for physical appearance and life goals. These, in turn, were larger than the overlap means for objects that can fall on one's head, one's own good qualities, qualities of one's best friend, and Ronald Reagan's qualities. The overlap mean for the restaurant script is closer to the second than to the third group of overlap scores.

The overlap means were similar to those found in earlier experiments. The value of .53 found for the restaurant script was similar to the value of .54 found in Experiment 1.

Organization of recall. An ANOVA on the output-output correlations (Bellezza, 1984a) shown in Table 2 indicated that type of knowledge structure was significant [$F(7,315) = 6.01, p < .001, MSe = .245$]. A Tukey HSD test indicated that any two correlations differing by .31 or more were significantly different. Inspection of the correlations in Table 2 reveals that the output-output correlation for the restaurant script was significantly greater than the output-output correlations for the other

Table 2
Characteristics of the Information Retrieved From the Eight
Knowledge Structures Used in Experiment 2

Knowledge Structure	Number of Items Retrieved	Within-Subject Overlap	Output-Output Correlations
One's Own Good Qualities	12.7	.41	.32
One's Own Physical Appearance	10.3	.55	.46
Objects That Could Fall On One's Head	18.1	.45	.35
Qualities of Ronald Reagan	8.5	.38	.43
Things One Does When Going to Eat at a Restaurant	16.4	.53	.88
Qualities of One's Best Friend	13.3	.40	.34
Instances of the Category Fruit	18.7	.77	.42
One's Life Goals	10.3	.55	.47

seven knowledge structures, which were not significantly different from each other. The correlation of .88 found for the restaurant script was similar in value to the .92 found in Experiment 1. The important conclusion to be drawn from these data is that special instructions to the subjects to order items retrieved for the restaurant script are not necessary for actions to be retrieved in the order in which they normally occur. The natural order is the order in which script events are experienced, so it is not surprising that the retrieval order reflects the order in which they are experienced (Barsalou & Sewell, 1985).

GENERAL DISCUSSION

The results found here indicate that information from scripts is not more reliably retrieved from memory than is information from many other kinds of knowledge structures. The mean overlap score used to measure reliability of retrieval was found to be .50 over all 10 scripts in Experiment 1. This value is very similar to the value of .55 found for definitions of concrete nouns (Bellezza, 1984c), the value of .46 found for definitions of category labels (Bellezza, 1984c), the value of .55 found for characteristics of famous people (Bellezza, 1984b), and the value of .56 for retrieving information about one's self (Greenwald, Bellezza, & Banaji, in press). In fact, the overlap value for the restaurant script in Experiment 2 was .53, which was not significantly different from the value of .45 found for the ad hoc and presumably not well-formed category of objects that can fall on one's head (Barsalou, 1983).

The knowledge structure that possesses the highest reliability of retrieval seems to be common categories, when instances of those categories are requested. These values hover around .70 (Bellezza, 1984a, 1987). Nevertheless, there is something special about scripts. First, any information that is repeatedly retrieved from scripts tends to always be retrieved in the same order. Thus, although script information is not as reliably retrieved as category instances, the information that is retrieved is highly ordered. The results of Barsalou and Sewell (1985) also demonstrate this point. Second, Bellezza (1984a) found that the larger the common category, the lower the reliability of retrieval from that category. However, the results of Experiment 1 showed the opposite relation for scripts. The more actions a subject gave for a script, the greater the reliability of retrieval. One possible reason for this relation is that script actions tend to be experienced together, and, therefore, each experience improves memory for both the actions and their relative order. Most category instances, such as animals, tend not to be experienced together.

Finally, it should be mentioned that the category of one's life goals did not provide information that was any more reliably retrieved from memory than information in other self categories. College students' behavior may be greatly influenced by their immediate goals (Schank & Abelson, 1977), but their long-term life's goals do not seem to be well enough organized to guide their behavior on a day-to-day basis.

REFERENCES

- ALBA, J. W., & HASHER, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203-231.
- ANDERSON, R. C. (1978). Schema-directed processes in language comprehension. In A. Lesgold, J. Pellegrino, S. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 67-82). New York: Plenum Press.
- BARSALOU, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211-227.
- BARSALOU, L. W., & SEWELL, D. R. (1985). Contrasting the representation of scripts and categories. *Journal of Memory & Language*, *24*, 646-665.
- BELLEZZA, F. S. (1984a). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society*, *22*, 324-326.
- BELLEZZA, F. S. (1984b). Reliability of retrieval from semantic memory: Information about people. *Bulletin of the Psychonomic Society*, *22*, 511-513.
- BELLEZZA, F. S. (1984c). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society*, *22*, 377-380.
- BELLEZZA, F. S. (1987). The reliability of retrieving information from knowledge structures in memory: Self information. *Bulletin of the Psychonomic Society*, *25*, 407-410.
- BELLEZZA, F. S., & BOWER, G. H. (1982). Remembering script-based text. *Poetics*, *11*, 1-23.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.
- CHARNIAK, E., & McDERMOTT, D. (1985). *An introduction to artificial intelligence*. Reading, MA: Addison-Wesley.
- GARDNER, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- GREENWALD, A. G., BELLEZZA, F. S., & BANAJI, M. R. (in press). Is self-esteem a central ingredient of the self-concept? *Personality & Social Psychology Bulletin*.
- KIRK, R. E. (1982). *Experiment design: Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Wadsworth.
- MANDLER, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: Erlbaum.
- MINSKY, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.
- NORMAN, D. A. (1982). *Learning and memory*. San Francisco: W. H. Freeman.
- PAIVIO, A. (1986). *Mental representation*. New York: Oxford University Press.
- RUMELHART, D. E. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.
- SCHANK, R., & ABELSON, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- WYER, R. S., JR., & SRULL, T. K. (1984). *Handbook of social cognition* (Vol. 1). Hillsdale, NJ: Erlbaum.