JOURNAL OF MEMORY AND LANGUAGE 24, 179–199 (1985)

The Representation of Scripts in Memory

VALERIE ABBOTT

University of Illinois at Chicago

JOHN B. BLACK

Yale University

AND

EDWARD E. SMITH

Bolt, Beranek and Newman, Inc.

Previous research has shown that people possess stereotyped knowledge about common events; that is, people have scripts containing information about the actions that comprise these events and about the temporal order of these actions. They use this knowledge in making inferences that help them to fill in gaps found in narratives or predict information to follow. The memory representation of common events was investigated by studying the pattern of inferences people make when they read descriptions of these events. People always made inferences that generalized the information presented. When a detail was presented, a more general concept of which it was a part was inferred. Readers often inferred items at the level of abstraction corresponding to scene headers when a sentence embodying a more abstract concept was stated. These results indicated that scripted events are represented in memory as hierarchically and temporally organized information packets. The connections between packets in the network are arranged in such a way that useful generalizations and predictions are available to aid people in understanding events.

People use what they know about the real world to understand both actual events and events in stories. That people have such knowledge is hardly controversial. What is debatable is how this knowledge is organized in memory. Since events are themselves sequential, it would seem that their representations should also be temporally organized. The research reported here indicates that knowledge about events, that appears to be entirely sequential is, in fact, also hierarchical.

This research was supported by a grant from the Sloan Foundation, U.S. Public Health Service Grants MH-19705 and MH-36586, and by the National Institute of Education under Contract US-HEW-C-400-81-0030. Experiments 1 and 2 are based on a master's thesis presented to the University of Illinois at Chicago in partial fulfillment of the requirements for the master's degree. We thank Lawrence Birnbaum and Brian Reiser for helpful discussions of the ideas presented in this paper and insightful comments on earlier drafts. Robert Abelson was of great assistance in the statistical analysis of the results reported here. Requests for reprints should be addressed to John B. Black, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, Conn. 06520.

PRIOR KNOWLEDGE AND UNDERSTANDING

When reading a text, people utilize their prior knowledge of the subject matter covered in the text. This prior knowledge plays an important part in facilitating the understanding processes of the reader. People read much faster and understand better than would be possible if each unit of information were processed in isolation without referring to prior knowledge to help decide what connection the item has with

0749-596X/85 $3.00

the rest of the passage. Consider, for example, what happens to reading speed and comprehension when one studies a difficult article in an unfamiliar field. The actual words used may not be unfamiliar and the sentences not particularly complicated, but reading is much slower than normal, and later recall is difficult.

This point is supported by a substantial body of experimental research. Dooling and Lachman (1971), and Bransford and Johnson (1972) found that subjects remembered a text much better when they knew what topics were being discussed; that is, subjects remembered more when they knew what prior knowledge to use in understanding the text. Similarly, Black and Bern (1981) found that memory was better when readers could use their prior knowledge to make inferences providing causal connections between the statements in a story. Bransford and Johnson (1973) and Thorndyke (1976) found that subjects claimed on a memory test to have previously read statements that were in fact inferences they had made, using real world knowledge, to understand the text. Graesser, Hoffman, and Clark (1980) found that the familiarity of the subject matter in a text was an important determinant of reading speed, and Miller and Kintsch (1981) found that parts of a text that were predictable from previous knowledge were read faster than those that were unpredictable. Thus, effects of knowledge on reading have been found in studies measuring amount remembered, memory distortions, and reading time. (See Black (1984) for a survey.)

One particular kind of knowledge that adult members of our society possess is knowledge about commonplace events such as going to restaurants and visiting doctors. Typically, a written account of such an event is incomplete. So much is left out that the account would be incomprehensible if the written description were the reader's only source of information. People use their knowledge both to fill in the gaps found in these narratives, and to anticipate what will come next at each point in the narrative. The contrast between having knowledge applicable to understanding an account of an event and not having such knowledge is illustrated by the difference in the comprehensibility of the following two sequences:

(1) George entered the department store. He picked out some shoes. He paid the cashier.
(2) George entered the doctor's office. He ordered a salad. He took careful notes.

The statements in the first episode all refer to the stereotyped series of activities typically performed when one goes shopping. The second episode does not refer to any body of knowledge shared by most people. In the first story, it is easy to fill in the gaps between the actions explicitly stated with other activities that are part of the same sequence. Each sentence can be anticipated on the basis of the preceding sentences. For example, George must have found the shoe department and put on the new pair of shoes; probably someone said "Cash or charge?" to him. In the second story, however, it is much more difficult to find sensible actions to connect the stated actions and there is no way to correctly predict what actions might come next.

The information used for filling in gaps and expecting further inputs does not come from the individual sentences but from the overall bodies of knowledge to which they refer. Some quite different actions would be expected to follow the act of picking out a pair of shoes if it were a part of getting dressed in the morning rather than part of buying apparel. This example illustrates the point made by Schank and Abelson (1977) and Schank (1982) among others, that people must use the information they receive in reading to make available from memory more general information that will guide processing of further inputs. Examples like the sequences above indicate the importance of accounting for the ability to accomplish this task in any valid language understanding theory. Memory must be organized so that relevant general information can be accessed from the input material.

At the same time, the memory organization must provide for inferences on a useful level of abstraction. Very abstract information, while providing a general framework for interpreting input, does not provide specific enough information to aid in understanding a detailed input. It should be possible to use the knowledge accessed from information in a text to find more specific predictions for processing later input.

An efficient memory organization for processing events should also represent temporal relationships among pieces of information. This would limit the possibilities for what is expected next. When one is reading a story about a shopping trip, if the shopper has found the item she desires, then the reader would anticipate information about paying for the item rather than entering a store.

Schank and Abelson (1977) have characterized a knowledge structure called a "script" that is an organization of information that allows access of relevant information during reading while ignoring irrelevant information. Scripts are intended to represent knowledge about events that are so well practiced in everyday life that their performance is stereotyped. Eating in restaurants, grocery shopping, and visiting doctors' offices are three situations about which we could expect people to have knowledge in the form of a script. A script for a commonplace event consists, in part, of the ordered sequence of actions and the standard characters and objects involved in the event. If people's knowledge about stereotypic situations is standardized, this would be an ideal sort of knowledge for experimental work on the use of information from long term memory during reading, because anyone coming into an experiment could be expected to possess the same information in his or her memory. This information would have been naturally acquired over the person's lifetime and, hence, would more likely be representative of the usual knowledge the person uses to interact with the world than would information acquired in the course of an experiment.

Previous Research on Scripts

Recently, the knowledge that people possess about commonplace events has been explored experimentally. In this section and the next, we selectively review those experiments concerning scriptal knowledge that are directly relevant to our own experiments. For a more extensive review, see Abelson (1981).

Investigations of people's knowledge of commonplace events have concluded that the actions in scripts are linked together in memory as sets; that is, when some of the actions in the set are accessed, so are the others. Bower, Black, and Turner (1979) provided two kinds of evidence for this conclusion. First, when they asked a group of 30 or so people to list the typical actions that make up such events as *going to a restaurant* or *getting up in the morning*, they found that only three or four of every hundred actions mentioned were mentioned by only one person and that many actions were referred to by more than half the subjects. These results show that there exists a commonly agreed upon group of actions that comprise these situations. Second, when given a memory test on stories that contained some of the actions from a script, subjects falsely remembered having seen the other actions in the script. Thus the script actions tended to be evoked together both when subjects were asked about what typically happens in a situation and when they read a story about that situation.

Smith, Adams, and Schorr (1978) also found that subjects falsely remembered omitted script actions after having read some of the actions in the script. In an additional study of time taken to recognize previously presented sentences, Smith et al. found that increasing the number of unrelated actions that subjects had to memorize slowed recognition time for each item. However, increasing the number of script actions subjects had to remember did not slow down recognition of the items. An explanation of this result is that items in the same script are linked in such a way that they are accessed together. Requiring subjects to remember a few more script items does not cause them to expend extra effort for retrieval.

Scripts are sets of actions, but they are not unstructured sets. In particular, scripts are ordered sets of actions. Bower et al. (1979) found that when subjects were asked to remember stories that presented some actions out of order, they tended to recall the stories with those actions shifted closer to their proper places in the sequence. Lichtenstein and Brewer (1980) obtained the same results when subjects viewed videotapes with actions that deviated from the standard order. Galambos and Rips (1982), using similar bodies of knowledge, have shown that when subjects try to determine the order in which two actions normally occur in a sequence, the decision takes longer if the two statements are close together in a temporal ordering than if they are far apart. All of these results indicate that temporal ordering is reflected in the memory representation for such events.

In addition to being ordered sets, actions in scripts seem to cluster together into closely interrelated subsets. Schank and Abelson (1977) termed these clusters "scenes." Bower et al. (1979) provided some evidence for a division of scripts into scenes. In particular, they found that when subjects had to divide a story into groups of statements that "go together," they divided the stories into chunks corresponding to scenes. Subjects agreed to a great extent on the location of the chunk boundaries, so there is evidence that the substructures involved are consistent across subjects. However, unlike the other characteristics of knowledge about scripted events described above, Bower et al. did not provide any evidence for the use of scenes in understanding stories. The experiments we report below provide some evidence that this characteristic of scripts is also used in understanding.

Serial versus Hierarchical Memory Structures

There is more than one way that information concerning commonplace events may be arranged in memory so that it fits the constraints of being accessed as a set, having temporal ordering, and being divided into scene substructures. One possibility is that actions involved in the event may be arranged using a simple serial ordering with markers to indicate scene boundaries. Another possibility is that people's knowledge of commonplace events may be arranged hierarchically in memory. In this section we indicate some processing consequences of these two hypotheses that will allow us to distinguish them experimentally.

An essential characteristic of a serially ordered scheme for representing events in memory is that each item in the representation be linked with the one preceding and the one succeeding it in time. Since knowledge about the event is linked together this way, and thus is set apart as a unit from other information in memory, the entire representation should be dominated by a header that denotes the unit–event as a whole.

Serial organization can explicitly maintain temporal ordering in memory and connect related information into a set that can be accessed as a whole. Each action sets up the conditions necessary for the next action, and these enabling relations link actions in a particular serial order. A serial organization does not provide a natural way of representing the substructures that seem to be characteristic of people's knowledge of commonplace events, but markers indicating substructure boundaries can be used for this purpose.

Another way in which information in

memory may be organized so that it is accessed as a set with substructures is as a hierarchical structure instead of an ordered list. At the top of this hierarchy for an event is an action that summarizes the whole event (e.g., *Visit Restaurant*), which we will call a *script header*. The overall event is broken into superordinate actions, which we will call *scene headers* (e.g., *Eating, Ordering*). Each superordinate node is then broken down into a detailed set of *scene actions*. These actions are linked to the rest of the hierarchy through their scene headers. The scene actions dominated by different scene headers are not directly linked with each other, but are indirectly linked via the superordinate network of scene headers. Figure 1 illustrates a representation of this kind.

A strict hierarchical organization connects related information into a set that can be accessed as a whole and that provides a natural way of representing scene substructures, but this kind of organization is not able to represent the knowledge people have about the temporal order of events. The modification of the strict hierarchical organization shown in Figure 1 would, however, allow items under a superordinate to be organized temporally with respect to each other. This modification adds temporal or enablement links between the various scene headers in a script, and between the actions subordinate to a particular scene header, but would not do violence to the basic notion that actions dominated by different scene headers are linked through only those scene headers. The experiments reported here provide evidence that the organization of common events in memory is such a modified hierarchy rather than a serial ordering.

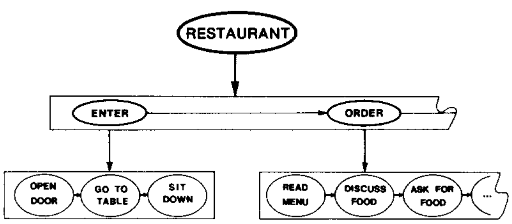

FIG. 1. Hierarchically organized representation of the Restaurant script.

As support for the hierarchical organization in Figure 1, consider the mixed history of studies attempting to demonstrate that people use the temporal information in a script when reading a script-based story. These experiments (e.g., Bower et al., 1979) focused on the time taken to read a target line in a script-based story, and studied this time as a function of whether the activities instantiated in the target and the line preceding it are far apart in the underlying script. To illustrate, consider the following vignette.

(1) John went to a restaurant.
(2) John ate his meal with gusto.
(target) (3) John paid the check and left.

There is relatively little "gap" between the script activities in the target line and the one preceding it—"pay the check" and "eat the meal." In contrast, there is a substantial gap between the target and preceding lines in the following:

(1) John went to a restaurant.
(2) John deliberated about his order.
(target) (3) John paid the check and left.

The basic prediction tested in these studies was that it should take longer to read a target line the greater the gap between the script activities mentioned in it and the preceding line (e.g., it should take longer to read "John paid the check and left" in the second vignette than in the first). This was the prediction because (i) using a script to understand a story presumably involves searching through the script for activities that match the story lines, and (ii) if the script activities are represented temporally, the duration of this search should increase with the number of script activities that intervene between the last activity matched and the one matching the present story line (Cullingford, 1978).

Early experiments failed to find the predicted effect of gap size. In Bower et al. (1979), for example, while reading time increased with initial increases in gap size, further increases in gap size led to a decrease in reading time. A close examination of these studies, however, suggested that the failure to find the expected results was due to a failure to appreciate the hierarchical structure of scripts. In some of the Bower et al. stories, different lines seem to describe actions at two different hierarchical levels, the scene header and scene action levels. Successive lines in these stories, therefore, sometimes required the reader to switch levels, which might have obscured the desired effect.

To overcome this problem, Smith (1981) varied gap size in stories that included either only scene headers (e.g., "Frank wanted to go to the movies. He bought tickets. He watched the movie."), or only scene actions (e.g., "Allan was going to buy tickets for a movie. He walked over to the ticket counter. He waited in line."). For the scene header stories, the time to read the target line increased monotonically with gap size. For the scene action stories, while reading time increased with an initial increase in gap size, a further increase in gap size led to a small and nonsignificant decrease in reading time. Thus, for scene actions, while there was a failure to find a strictly monotonic effect of gap size, there was no evidence for real nonmonotonicity.

In sum, that there was a monotonic effect of gap size for scene header stories, and a gap size effect with no real nonmonotonicity for scene action stories, implies that the temporal order of events is a part of the representation that people use in reading stories about events. This is evidence for the presence of some serial organization in these representations. However, that gap size effects are found only when the level of story events is held constant supports the hypothesis that these representations are also hierarchically structured.

Experiment 1: Memory for Single-Script Stories

The story comprehension process utilizing a serially organized representation is quite similar to one utilizing a hierarchically organized representation. Each story statement is read and the script structure is scanned for a path connecting the currently matched action with the preceding one. The script actions that comprise the connecting path are then inferred to have occurred even though they were not mentioned explicitly in the story. However, due to a difference between the pattern of inferences expected with the hierarchy and with the serial chain, hierarchical organization allows the processing to be more efficient. With the hierarchical representation, the scene header for each scene is always inferred in reading a story about a scripted event because it is on any path through the script, whereas the scene actions need not always be inferred. With the serial chain, all of these actions are always inferred because there is only one path through the script—namely, the one including all the actions. Therefore, our hierarchical representation for scripts predicts an asymmetry in inferencing of scene headers and scene actions: processing scene actions causes their scene headers to be inferred, but not vice versa. A serial chain representation, on the other hand, makes no such prediction, so we have an empirical distinction between the two representations that can be tested in a recognition memory task.

In this experiment, we varied whether the scene actions and scene headers were present or absent in target scenes of script-based stories. We compared the rate of falsely recognizing test items that were scene headers of presented scene actions with the false recognition rate for items that were scene actions of presented scene headers. In particular, we were interested in two effects:

(1) The effect of presenting a scene action on subjects' ratings of likelihood that they have seen its scene header.
(2) The effect of presenting a scene header on subjects' ratings of likelihood that they have seen its scene action.

A hierarchical network hypothesis predicts that the effect in (1) would be greater than that in (2). A serial ordering hypothesis does not predict any difference.

Method

Subjects. Subjects were 32 undergraduate students who participated to fulfill a course requirement.

Materials. We constructed four stories from scripts in such a way that we could manipulate independently the presence or absence of a header and action in at least one scene. The four stories were about *going to a restaurant, attending a lecture, visiting a doctor's office,* and *grocery shopping.* The target scene (i.e., the scene in each story chosen for variation) was moderately important; that is, statements corresponding to its header and actions had moderately high production frequencies in script norms (Bower et al., 1979). Each target scene was at least five statements long; this length further insured that the target scene was an important part of the story. The target scenes appeared in the stories in one of four forms:

(1) Nothing from the scene was mentioned.
(2) Only the scene header was mentioned.
(3) Only one scene action was mentioned.
(4) Both the scene header and one scene action were mentioned.

These forms provided the four conditions of the experiment. In particular, they represent the orthogonal manipulation of the presence or absence of the scene headers and scene actions. The rest of each story was composed by stating all the nontarget scene actions and scene headers produced by more than 25% of subjects in the Bower et al. norms. Some additional material was also included to make the stories more readable. One of the stories used, *Going to a Restaurant,* is shown in Table 1.

TABLE 1
TEXT OF *GOING TO A RESTAURANT*

When they got home after shopping, Mary and her friend were hungry, so they decided to go to a restaurant. They called to make a reservation and drove to the restaurant.

Mary opened the door of the restaurant. They went inside. Mary gave the reservation name to the hostess. The table was not ready yet, so they had to wait to be seated. In a few minutes they went to their table. They were seated.

Scene action:	They discussed what they wanted to eat.
Scene header:	They ordered their meal.

They put their napkins on their laps. The waitress brought them their appetizers and refilled their water glasses. They drank water and ate their appetizers. When their meal arrived, they ate their food. When they were finished, they decided to forget their diets and order dessert. They ate the dessert. It was getting late, so they asked for the check. The check came. They figured out the tip and left it on the table. Mary paid the check. They went to the checkroom and got their coats. They left the restaurant.

Note. The target scene is enclosed in a box. Different versions of the story included the scene action alone, the scene header alone, both or neither.

Regardless of which version of a story the subjects received, they all took the same recognition test. There were 20 recognition items for each story, 9 of which were always old (i.e., appeared in the story), for example, "In a few minutes they went to their table"; and 9 of which were always new, for example, "They heard the telephone ringing." The 9 new items were statements that could have been in the story, but were not, and were also not implied by the story. The other 2 items were the scene action and the scene header from the target scene; "They discussed what they wanted to eat" and "They ordered their meal." Whether these items were new or old depended on the experimental condition.

Procedure. Each subject read one story in each of the four versions. Across subjects, each version was read by eight subjects. Subjects were given 2 minutes to read each story, and were then asked to write one sentence describing what they expected to happen next in the story. They were given 30 seconds to make this extrapolation. (This extrapolation provided a cover task and was designed to encourage the subjects to process the material meaningfully.) Subjects then performed an unrelated intervening task for 20 minutes. After this, subjects were presented with an unexpected recognition test for the stories. The recognition items were blocked by story title. Subjects were asked to rate each item on a seven point scale according to how likely it seemed that the item was stated in the story, where "7" indicated certainty that the item had not been stated and "1" indicated certainty that it had been stated.

Results and Discussion

As predicted by the hierarchical hypothesis, there was an asymmetry in inferencing favoring the scene headers over the scene actions. Figure 2 gives the mean recognition ratings for the scene headers and scene actions. The data are collapsed over story topic, but are given separately for each story version.

The "NEITHER" condition consists of subjects' ratings for scene headers and scene actions when neither had been stated in the story. The "OTHER" condition consists of ratings for the scene action when only the scene header had been stated in the story, and for the scene header when only the scene action had been read. In the "SAME" condition were subjects' ratings for items when the identical item had actually been presented, and the other possible item from the same scene had not. The "BOTH" condition contains subjects' ratings for each item when both scene header and scene action were read.

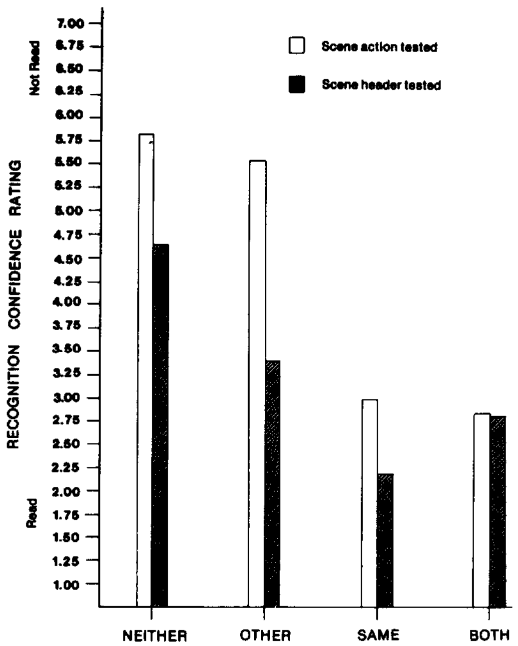

FIG. 2. Subjects' ratings of recognition items for Experiment 1. Here "1" means certainty the item was read, "7" means certainty the item was not read.

A comparison of ratings in the NEITHER and OTHER conditions in Figure 2 indicates that presentation of a scene header did not affect ratings for the scene action from the same scene, but presentation of a scene action did affect ratings for the scene header from the same scene. This asymmetry argues strongly against serial organization and for hierarchical organization. The difference between ratings in the NEITHER and OTHER conditions for scene action test items was negligible—only 0.28 point on the scale. Ratings for scene actions when none were presented were well below the 7.0 maximum possible on the rating scale, so a ceiling effect for these items cannot be used to explain this result. For scene headers the difference was a much larger 1.24 points. Thus, scene headers show a tendency to be inferred by subjects when they read an action from the same scene. For scene actions, such a tendency was not evident.

The interaction between test item type (scene header or scene action) and condition was significant by subject ($F(2,56) = 4.97$, $p = .010$), but not by story ($F(1,3) < 1$). This interaction was mostly due to the results for the OTHER condition, in which ratings for scene headers and scene actions diverged compared to the NEITHER, SAME, and BOTH conditions in which both test item types were given more similar ratings. This is exactly the pattern of ratings we would expect if subjects infer scene headers when presented with scene actions, but fail to infer scene actions when presented with scene headers. A linear × quadratic comparison to test whether ratings diverge significantly in the OTHER condition was significant by subjects ($F(1,28) = 5.88$, $p = .021$). This comparison was not significant by stories ($F(1,3) < 1$) because we were only able to use four stories, but three of the four stories used in this experiment showed the expected pattern.

The problem with the other story may have been our unfortunate choice of a scene action for the critical scene. The subjects objected to this statement, rating it as probably not seen even when it was presented explicitly in the story. The predictions of either structural hypothesis rest on the assumption that when an item is presented there is a reasonable chance that it will be processed and remembered. If it is not, it is not clear that the item can be expected to have an effect on the processing and remembering of other items.

Also more consistent with a hierarchical then with a serial ordering hypothesis is the fact that the scene headers received higher confidence ratings (3.21) than the scene actions (4.25). There was also a significant main effect of test item type ($F(1,28) = 32.48$, $p < .001$ by subjects; $F(1,3) = 6.42$, $p = .086$ by materials). According to the hierarchical hypothesis, scene headers are in a privileged position determined structurally. Processing has to pass through a scene header whenever one of its scene actions is mentioned. Also, as was seen in the Smith (1981) study described above, scene headers are temporally linked, so a scene header is on a direct path through a representation between its two neighboring scene headers. These superordinate items should have more chances to be processed than the more subordinate scene action items and consequently should be more likely to be included in the memory representation for an event. In order to account for this scene header superiority, the serial ordering hypothesis would have to maintain additionally that some actions have memory attributes that make their base level of response higher. There is no reason to believe such an ad hoc claim, so the serial ordering hypothesis cannot parsimoniously account for the results. In an overall analysis of critical items, there was a main effect of presentation condition ($F(3,84) = 23.69$, $p < .001$ by subjects; $F(3,9) = 13.93$, $p < .001$ by materials). Not surprisingly, both scene headers and scene actions were rated as more likely to have been seen when they were presented (SAME and BOTH conditions) than when they were not (NEITHER and OTHER conditions) ($F(1,28) = 65.10$, $p < .001$ by subjects; ($F(1,3) = 28.46$, $p = .016$ by materials).

An alternative hypothesis about our results that must be considered is that they are due to a general strategy on the part of the subjects to remember only general information. This alternative can be rejected because the scene headers and scene actions were rated essentially the same in the BOTH condition. If subjects were simply biased against responding positively to detailed test statements or in favor of translating all input into general terms, this bias should not have vanished when one scene header and one scene action from a scene were presented.

The results for the noncritical items (which were the same in all the conditions) did not interact with the experimental conditions. Subjects gave average ratings of 2.03 to noncritical veridical items and ratings of 6.28 to noncritical distractor items. Ratings for these items did not vary with which version of the story the subject received (for veridical items $F(3,93) < 1$; for distractor items $F(3,93) < 1$). In the NEITHER condition (the condition in which critical items were not presented and hence were always distractors), noncritical distractor items were rejected by subjects with more certainty than critical items ($F(1,31) = 10.93$, $p = .002$). Thus, the critical items showed more of a tendency to be inferred than did noncritical distractors. Recall that the noncritical distractor items were designed not to be inferable from the stories, whereas the scene headers and scene actions were inferable using the script. Also, as is to be expected, the critical items in the NEITHER condition were more likely to be rejected than noncritical veridical items that were actually stated in the story ($F(1,31) = 87.41$, $p < .001$).

EXPERIMENT 2: MULTISCRIPT STORIES

The results of Experiment 1 indicate that people have memory representations for events that permit them to infer abstract knowledge (scene headers) from input details (scene actions). This allows them to use general bodies of knowledge to provide a framework for understanding what they read. We found no evidence in Experiment 1 that people infer more detailed information when they read general statements. However, surely there are limits to this process of generalization. If contact is made only with more abstract knowledge, inferences quickly become so general as to be useless. There must be some way to facilitate inference on a useful level, one which is specific enough to provide useful expectations, but not so specific that the expectations are likely to be wrong.

In studies that bear on this problem, Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) found that for natural object categories, there is a basic level of classification at which people prefer to describe and think about objects. For example, in most contexts people prefer to classify an object as an *apple* rather than more generally as a *fruit* or more specifically as a *McIntosh apple*. This idea of a basic level may carry over to knowledge about situations like going to a restaurant or visiting a doctor (Abbott & Black, 1980). If there is a basic level of abstraction for situation knowledge, then people might prefer to infer actions on that level when actions at a more abstract or a more specific level are presented.

Experiment 2 was designed to test this possibility. It essentially repeated Experiment 1, but with everything moved up one level of abstraction. Instead of varying presence versus absence of scene headers and scene actions as Experiment 1 did, Experiment 2 varied presence or absence of script headers and scene headers. Recall that a script header is a reference to a whole script (e.g., "Joe ate at a restaurant"), while a scene header is a reference to a scene (e.g., "Joe ordered a meal"). If there is in fact a basic level of representation for scriptal situations it would likely correspond to the scene header level because inferences on this level would yield fairly specific predictions about what was to follow in a story without generating inferences so specific that there would be a high probability that they were wrong. Thus, presentation of a script header might cause its scene headers to be processed and possibly be included as part of the memory representation.

Generalization from input material would always facilitate linking that material with more abstract information in a subject's memory. That is, inferring from a statement such as "Joe ordered a meal" that Joe was in a restaurant would make available general information about restaurants that could be useful in processing the rest of the story. Therefore, when a scene header is mentioned, we expect script headers to be inferred and included as part of the memory representation.

At this higher level, then, we did not expect to find the superordinate–subordinate asymmetry in inferencing that we found in Experiment 1. Generalization of input information would still occur, and if the scene header level was the most useful or basic level for situation knowledge, then people would show a tendency to infer information about events at that level.

Method

Subjects. Subjects were 32 Yale undergraduate students who participated to fulfill a course requirement.

Materials. We wrote two multiscript stories. One was about the activities of a student during a day; the other was about the experiences of a man going to a city to meet a friend who was arriving by plane. These stories were entitled *The Day* and *The Trip*. The stories were constructed so that each would contain references to eight stereotyped situations (scripts) in addition to sentences added to flesh out the stories and make them more readable. Two statements were chosen from an account of each of the eight target situations, one a script header and the other a scene header. For example, in the restaurant situation, the reference to the script header was "He ate dinner at a fancy restaurant" and the scene header was "He ordered a gourmet meal and some wine." The story versions varied according to whether both or neither the script header or the scene header, was presented for a given situation. Two scripts were presented in each condition in each story, the order of conditions counterbalanced over subjects using a Latin square. One of the stories used, *The Day,* appears in Appendix A.

The recognition test for each story contained the same items for all subjects, and the order of items was randomized for each subject. There were 48 items on the test for each story, 24 of which had been present in the story and 24 of which were new. Both the script header and the scene header from each situation appeared on the recognition test. Thus, there were 16 items that were always old, 16 items that were always new, and 16 items that were old or new depending on the condition. This last group of 16 was made up of the eight script headers and eight scene headers for events in the story.

Procedure. Each subject was presented with one version of each story. Half the subjects read *The Day* first and the other half read *The Trip* first. Subjects were instructed to read each story carefully so that they could answer questions about it later, and were allowed to read at their own rate. They were then given an unrelated intervening task to perform for 20 minutes, after which they were given the recognition test for the first story, and were asked to rate each recognition item on a 1–7 scale according to how likely it seemed to them that the item was in the story ("7" indicated certainty that the item had not been stated and "1" indicated certainty that it had been stated). The recognition test for the second story followed immediately. Subjects were self-paced throughout the recognition portion of the experiment.

Results and Discussion

Results. As we expected, the results of the recognition tests show that presentation of scene headers led to false recognition of script headers, and presentation of script headers led to false recognition of scene headers. In other words, there was no inferencing asymmetry at this level as there was at the lower level of abstraction.

The mean confidence ratings for the recognition test of critical items in each event condition are presented in Figure 3. The open bars in the figure present the ratings for the scene headers in the various event conditions and the dotted bars present the ratings for the script headers.

The difference in ratings of test items between the NEITHER and OTHER conditions (from 5.19 to 4.26) shows that false recognition increased for these items when another item from the script was mentioned ($F(1,31) = 36.25$, $p < .001$ by subjects; $F(1,15) = 18.11$, $p = .001$ by materials). This result, along with the lack of an interaction between test item type and condition for these two conditions ($F(1,31) < 1$ by subjects; $F(1,15) < 1$ by materials), indicates not only that more general knowledge structures were sought for interpretation of input material, but also that people infer more detailed information than they have actually read. Not only did false alarming occur for script headers when their scene headers were read, but there was also false alarming to scene headers when their script headers were read in the story. This lends

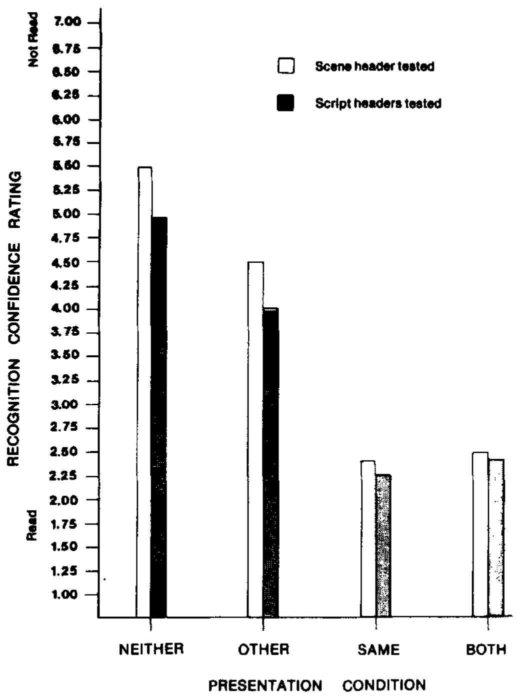

FIG. 3. Subjects' ratings of recognition items for Experiment 2. Here "1" means certainty the item was read, "7" means certainty the item was not read.

credence to our claim that scene headers occupy a privileged position in the representation of knowledge about situations, in that the inferencing asymmetry favoring more general items found in Experiment 1 does not occur when script headers are the more specific items.

Overall, the script header statements were rated slightly more likely (3.35) to have been seen than the scene header statements (3.72). Although rather small, this difference was statistically significant ($F(1,31) = 21.10$, $p < .001$ by subjects; $F(1,15) = 11.64$, $p = .004$ by materials). However, this difference disappears in the "SAME" and "BOTH" conditions, which indicates that subjects were as likely to recognize a scene header as a script header when they were stated in the story.

The effect of presentation conditions was significant overall ($F(3,93) = 113.54$, $p <$ .001 by subjects; $F(3,45) = 74.76$, $p < .001$ by materials). Items were rated more likely to have been read when they were actually presented (SAME and BOTH conditions) than when they were not (NONE and OTHER conditions) ($F(1,31) = 217.45$, $p <$ .001 by subjects; $F(1,15) = 120.24$, $p <$ .001 by materials).

For noncritical items, subjects gave average ratings of 2.14 to items they had seen, and average ratings of 5.41 to items they had not seen. These ratings are in an acceptable range and are comparable to the ratings of critical items in the NEITHER and BOTH conditions.

EXPERIMENT 3: MEMORY FOR THREE LEVELS OF SCRIPT ACTIONS

The combined results of Experiments 1 and 2 indicate that people have an overall strategy of inferring more general knowledge structures (i.e., inferring scene headers from scene actions, and script headers from scene headers). Added to this is the tendency to infer, whenever possible, the basic-level information or scene headers. However, it is possible that it is only under the special conditions of Experiment 1 that there is an asymmetry of inferencing between scene actions and scene headers; or the equality of inferencing between scene headers and script headers might only be seen under the conditions of Experiment 2. Consequently, we conducted an experiment that combines these conditions.

Method

Subjects. Subjects were 32 undergraduate students at Quinnipiac college who volunteered to participate.

Materials. We wrote four stories, each of which contained references to four scripted situations. All stories were about the activities of a main character during a single day. These stories were entitled *The Secretary's Saturday, Winter Vacation, An Ordinary Day,* and *A Stormy Day.* To make the stories coherent and at least moderately interesting, filler material was inserted between references to each scripted situation. Three statements were chosen from an account of each of the four script situations in a story, a script header, a scene header, and a scene action. For example, in the classroom situation the reference to the script header was "She attended her afternoon class," the scene header was "She listened to the teacher's lecture," and the scene action was "She took careful notes." There were eight possible presentation conditions for each target situation:

A the scene action only,
B the scene header only,
C the script header only
D all three,
E the scene action and scene header,
F the scene action and script header,
G the scene header and script header,
H nothing.

The order of scripts in a given story was always the same: that is, the main character of *An Ordinary Day* always went to the laundromat first, then to the bank, and so on. The order in which presentation conditions occurred was counterbalanced over subjects.

All subjects took a recognition test for each story composed of the same items in one of four random orders. There were 36 items on the test for each story, 18 of which had been presented in the story and 18 of which were new. The script header, the scene header and the scene action from each target situation appeared on the test. There were 12 items that were always old, 12 that were always new, and 12 that were old or new depending on the condition. This last group of 12 was made up of the four script headers, four scene headers, and four scene actions for situations in the story.

Procedure. Each subject was presented with one version of each story. Subjects were instructed to read each story carefully so that they could answer questions about it later. They were allowed to read at their own rate. Subjects were then given an unrelated intervening task for 20 minutes, and then the recognition tests. Subjects were asked to rate each recognition item on a 1–7 scale according to how likely it seemed to them that the item was in the story they had read. The number "7" indicated certainty that the item had not been stated and the number "1" indicated certainty that it had been stated. The scale was the same as in Experiment 1, and subjects worked at their own pace.

Results and Discussion

Figure 4 presents the results for the critical items. The striped bars present results for script header items, the dotted bars present results for scene header items, and the open bars present results for scene actions. The six possible presentation conditions for a tested item form the horizontal axis of the figure. These are

1. NONE—no item from the target situation was presented;
2. OTHER1—one item from the target situation was presented, but not the tested item;
3. OTHER2—the tested item was not presented, but both other items from the target situation were;
4. SAME—only the tested item was presented;
5. SAME+—the tested item plus one other item from the target situation were presented; and
6. ALL3—all items from the target situation were presented.

Note that these are presentation conditions *for a particular tested item.* To illustrate, when the item tested is a script header, the OTHER1 condition consists of ratings to script headers when only the scene header or only the scene action from its script appeared in the presented story.

As can be seen in the OTHER1 and OTHER2 conditions in the figure, subjects falsely recognized scene headers and script headers when other items from the same situation were mentioned. They were less likely to falsely recognize scene actions, whatever other items were presented. This results in a significant interaction between the effect of presentation condition and test item type ($F(10,390) = 5.33$, $p < .001$ by subjects; $F(10,150) = 4.61$, $p < .001$ by materials).

The critical issue was whether recognition ratings for scene actions were influenced less by presentation of other items in the same script than were scene headers or script headers. This would lend credence to the hypothesis that subjects always draw generalizing inferences, but draw specifying inferences only if, by doing so, they can obtain information at the scene header level. If this is true, ratings for scene headers and script headers should diverge from ratings of scene actions in the OTHER1 and OTHER2 conditions because the former are affected by presentation of other items from the same script, but the latter are not. This divergence was tested using a Linear × Quadratic comparison between two test item types at a time.

The comparison was significant between test items at the scene action and scene header levels ($F(1,39) = 6.54$, $p = .014$ by subjects; $F(1,15) = 7.24$, $p = .016$ by materials). This indicates, as can be seen by inspection of Figure 4, that while recognition ratings for scene headers and scene actions differ somewhat in the NONE and SAME conditions (by about 0.5 points), they differ much more in the OTHER1 and OTHER2 conditions (by about 1.3 points). This result reinforces the result of Experiments 1 and 2, showing that when a scene action or a script header is stated, subjects infer the scene header, but that scene actions are not inferred when scene headers or script headers are stated.

The same comparison between the scene action and script header level test items was also significant ($F(1.39) = 21.48$, $p < .001$ by subjects; $F(1,15) = 16.88$, $p = .001$ by materials). Recognition ratings for script header and scene actions diverge in the OTHER1 and OTHER2 conditions (a 1.6-point difference) as compared to the NONE and SAME conditions (a 0.5-point difference). This result illustrates again the generalizing inferences subjects make, and their failure to make specifying inferences to the level of scene actions.

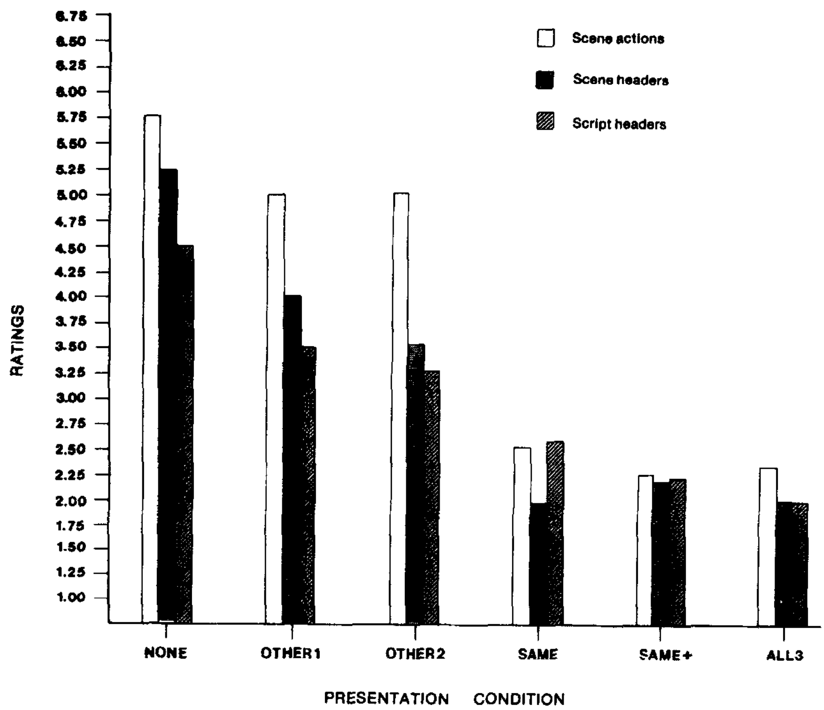

FIG. 4. Recognition ratings for items in Experiment 3. Here "1" means certainty the item was read, "7" means certainty the item was not read.

Between test items at the scene header and script header level, there is no such divergence in the OTHER1 and OTHER2 conditions, so the comparison is not significant ($F(1.39) = 1.26$, ns by subjects; $F(1,15) = 1.11$, ns by materials).

These results implicate the scene header level of abstraction as a particularly useful level at which to process information. Subjects tend to infer information at this level no matter at what level of abstraction information is actually presented. Also, subjects tend to make generalizing inferences from presented information at all levels tested.

Overall, script headers were more likely to be recognized than scene headers, and scene headers were more likely to be recognized than actions. So there was a significant effect of test item type ($F(2,78) = 34.53$, $p = .001$ by subjects; $F(2,30) = 14.69$, $p < .001$ by materials). Subjects were more likely to recognize a test sentence when it had been presented than when it had not, so there was also a significant effect of presentation condition ($F(5,195) = 78.50$, $p < .001$ by subjects; $F(5,75) = 59.01$, $p < .001$ by materials).

The average ratings subjects gave for noncritical veridical items (2.42) were similar to ratings for critical items that were presented (2.23). Their ratings were not significantly different from critical items in the SAME condition ($F(1,39) = 1.92$, ns). The average ratings for noncritical distractor items (5.76) indicated that they were less likely to be recognized than critical items (5.25) when no other items from their scripts were presented (the NONE condition) ($F(1,39) = 11.16$, $p = .002$). This indicates that the activities referred to in the critical items might have been more predictable in the context of the story than the distractor items.

General Discussion

The Case for Hierarchical Representations

We can now combine the results from Experiments 1 through 3, along with previous work, to yield a coherent picture of how scripts are represented in memory and how such representations are used to understand stories.

Previous work has shown that people utilize temporal ordering information in representations of events to guide their inference processes. The Smith (1981) study in particular demonstrated that when subjects do not have to switch hierarchical levels in reading stories, they take longer to read a sentence the more separated it is from the preceding sentence in the temporal sequence of the script. Thus, serial organization is indicated in representations of scripts, since the length of the search necessary to connect two items on the same level of abstraction depends on their distance apart in the script. However, since this effect of temporal order seems to occur only in those experiments where subjects read stories at a constant level of abstraction, serial organization alone cannot describe people's representations of events; rather, the representation must distinguish between different levels.

These results suggest a hierarchical arrangement for information about events in memory. Additional support for this suggestion comes from a number of other sources in both cognitive psychology and artificial intelligence. Many theories of memory for stories propose a hierarchical structure for stories (e.g., Rumelhart, 1977; Thorndyke, 1977; Mandler & Johnson, 1977). The basic finding related to these theories is that people remember events high in the story hierarchies better than those at lower levels. Thus event hierarchies have proven to be a useful representation for predicting the probability of remembering the statements in stories. Black and Bower (1980) have characterized stories in terms of problem solving hierarchies. They found that actions that lead to successful resolution of problems, or that are high in the action hierarchy, were better remembered than unsuccessful attempts and actions low in the hierarchy. Lichtenstein and Brewer (1980) found that such a hierarchical representation is also consistent with people's memory for videotaped events. Their study suggested that actions are more strongly linked to their superordinates than to the actions temporally contiguous with them in the sequence.

The question of why scripts should be organized hierarchically has been addressed specifically in the artificial intelligence literature. Sacerdoti (1974) argued that a hierarchical network of actions is the most efficient representation for use in planning how to accomplish a goal. In particular, he was concerned with endowing a robot with enough problem solving ability to enable it to interact with a simple world. He found that using a hierarchical representation in planning actions was more efficient for a problem solver than arranging the planning actions at a single level of detail because it allowed the robot to avoid wasting time worrying about details of plan execution when there was a flaw in the plan at a more general level. So, in addition to aiding a person's understanding of the actions of others, hierarchically structured information can be used to guide the person's own actions.

Experiment 1 turned up an asymmetry in inferencing. When a scene action was mentioned in a story, its scene header was likely to be inferred and stored as part of the memory for the story; on the other hand, mentioning the corresponding scene header did not lead to the subordinate scene action being inferred and stored in memory. We claim this asymmetry occurs because the scene actions are linked to the rest of the script only via their scene headers. Thus whenever a path is constructed between two story statements involving actions from different scenes of the script, these scene headers must be part of the path. This is further evidence for hierarchical organization in representation of scripts. A result from Experiment 2, that a script header is inferred when a scene header is mentioned in a story, shows that the generalization expected from a hierarchical representation extends to the script header level.

Combining the results of previous work with the results of Experiments 1 and 2, we can see that an appropriate representation for scripted events must combine an overall hierarchical organization with temporal connections between items under a given superordinate. Such a representation is illustrated in Figure 1. In this representation the script header, *Restaurant,* is related as a superordinate to the scene headers, *Enter, Order, Eat,* and so on. These scene headers are connected serially with each other. Each scene header stands in a superordinate relation to its scene actions. These, in turn, are interconnected serially. This representation would allow generalizing inferences to be made, while allowing the effect of serial order seen in the Smith (1981) study.

The results of Experiment 2 indicated that the asymmetry in inferencing does not occur if one moves up a level in the event hierarchy. Specifically, mentioning script headers in a story leads to their scene headers being inferred and stored in memory and mentioning scene headers in a story leads to their script headers being stored. This result suggests that people are able to infer more detailed information from general information when the detailed inference is at the scene header level. Experiment 3 confirmed that these results were not peculiar to the conditions presented in Experiments 1 and 2. It reaffirmed our contention that inferences are most likely to be made at the scene header level.

Given the results of these experiments, it would be possible to maintain that this is not the only level at which these inferences are made. We have not given subjects the opportunity to falsely recognize script headers when only more abstract items have been presented. Under most circumstances it is difficult to imagine that a restaurant experience would be inferred on the mention of a social occasion. However, situations could probably be constructed in which this could happen. The general point of these investigations is that people are capable of using hierarchical structures in their representations of events in order to generalize detailed input or specify abstract input so that expectations can be generated at a useful level. The degree of flexibility people have in deciding which level of abstraction will yield the most useful expectations is left to future research.

Nevertheless, mention of a script header evokes the scene headers, and this evocation is neither in the direction necessary for generalization of input material nor essential for connection with earlier stated actions. Its functionality, we believe, stems from the need to supply expectations that facilitate connection with subsequently stated action. For this purpose, scene headers appear to be the level that people prefer or find most useful to think about script events, at least when they read stories about the events and are not specifically encouraged to concentrate on any one level. In this way it is similar to the basic level of abstraction that Rosch et al. (1976) found for natural object categories. This similarity is explored in the next section.

A Basic Level for Situation Knowledge[1]

While most objects can be categorized at various levels, Rosch et al. (1976) showed that one level in a taxonomic hierarchy is basic. For example, a particular object may be categorized as a *piece of fruit*, an *apple*, or a *Jonathan apple*, but only *apple* is a basic level category as witnessed by the facts (among others) that an appropriate object can be categorized most quickly as an *apple* and is most likely to be called "apple" in a free naming situation. In short, we tend to describe the object world in terms of basic level categories. As for what makes a category "basic," numerous researchers have lined up with Rosch et al. in arguing for the importance of the number of distinctive attributes, where a "distinctive" attribute is one that is shared by many category members and few nonmembers (e.g., Abelson, 1981; Murphy & Smith, 1982). This criterion clearly works for our *apple* example: instances of *apple* share many attributes with one another but few with instances of competing categories such as *orange;* in contrast, *piece of fruit* has few distinctive attributes since instances of this category share few attributes; and *Jonathan apple* has few distinctive attributes because most of the attributes its instances share are also shared by members of competing categories such as *McIntosh apple*.

How well does this kind of analysis apply to scripts? At first blush, quite well. A familiar situation, like going to a restaurant, can be described at various levels of a hierarchy including the levels of script header, scene header, and scene action. And the present experiments suggest that the scene header level is the basic one in that it seems to be the level at which people prefer to describe, that is, draw inferences about, the situation. Further, as with basic object levels, scene headers can be described with a single word, for example, "eat," "order," while other levels require several words. Given these similarities, the obvious question is, are scene headers "basic" because they have many distinctive attributes?

The scene header does fulfill the requirement that an instance share many attributes with other instances of the same scene header, and share few attributes with instances of other scene headers. Any instance of *ordering* involves a decision concerning what object is to be requested, the necessity to gain the cooperation of an agency with the capability of conveying the object to the actor, and the communication of the results of the decision to that agency. This is the case whether *ordering* is done in a restaurant or from a Sears catalogue.

[1] The ideas presented in this section are derived, in part, from conversations with R. Abelson, E. Markman, and B. Tversky.

Few attributes are shared between *ordering* and *eating,* a competing category. Thus, scene headers do have the internal cohesiveness and external distinctiveness required of basic level categories.

The situation becomes more complex, however, when one considers the attributes of script headers and scene actions. An instance of *visiting a restaurant* involves going in, asking for the desired items, getting them, paying, leaving, and so on. As such it has many common features with other instances of *visiting a restaurant.* Thus, the scene header level of abstraction does provide for internal cohesiveness in a way not found in superordinate categories in object taxonomies. Notwithstanding the fact that script headers are not externally distinctive, *visiting a department store* also has the attributes described above for *visiting a restaurant,* the situation is clearly different than that for object taxonomies. Even worse is an examination of the attributes of a scene action such as *reading the menu.* At the corresponding level in an object taxonomy (*Jonathan apple*) an item should have many attributes in common with competing categories, as well as many common attributes with other instances of the same category. However, although different instances of *reading the menu* are similar to each other, they are different from *asking for crab bisque.* Thus, according to this analysis, the scene action could be the basic level. However, scene actions are not the level at which subjects prefer to make inferences about scripted events.

The problem here may be the difference in the relationship between levels in the hierarchy and items at the same level in object taxonomies and scripts. An apple is a type of fruit, but ordering is not a type of visiting a restaurant. That is, the connection between nodes at different levels in a script is not *isa* but *partof.* Also, there is no causal or temporal connection between items in an object taxonomy as there is for items in a script. The fact that *ordering* occurs leads to the expectation that *paying* will follow. There is nothing about an apple that leads us to expect oranges.

What are the ramifications of considering a script as a partonomy (rather than a taxonomy) for the existence of a basic level for scripts? To get some leverage on the question, consider how people seem to hierarchically represent the object partonomy of body parts (e.g., Miller & Johnson-Laird, 1976). The top level of the hierarchy contains a single node that depicts the *entire body;* the latter node is connected by *partof* relations to second level nodes that depict the *head, trunk, arms, and legs;* each second level node is in turn connected by *partof* relations to third level nodes, for example, *leg* has as parts *hip, thigh, knee, ankle, and foot;* and so on, for example, *foot* has as parts *heel, arch, sole, and toes.* Is there a basic level here?

One sense of "basic level" that can be defined for object partonomies and perhaps generalized to scripts is the most general or inclusive one at which "good parts" are first defined, for example, the level including *head, trunk, arms,* and *legs* in the hierarchy of body parts. But what exactly is a "good" part? In a recent paper, Tversky and Hemenway (in press) suggest three determinants of "goodness." First, a good part is one that is perceptually salient (see Anderson, 1975, and Brown, 1976, for some supporting anthropological evidence). This criterion seems promising for object concepts chiefly because perceptual salience may be reducible to known perceptual factors (e.g., *arm* and *trunk* are perceptually salient because they are so discontinuous from one another). But the perceptual salience criterion seems less relevant to scripts where the parts are temporally defined events whose salience may not be readily reducible to familiar perceptual factors.

A second suggestion of Tversky and Hemenway (in press) is that a good part is one that is functionally significant (again, see Anderson, 1975, and Brown, 1976). Functional "significance" may reflect not only the frequency and utility of the function but also the modularity of the function. Thus, *head, trunk, arms,* and *legs* have modular functions in that they can be stated relatively independently of one another and can appear in numerous other object partonomies. This aspect of functional significance seems to transfer nicely to scripts. In the restaurant script, *entering, ordering, eating,* and *paying* are modular in that their functions can be defined relatively independently of one another and that these parts can appear in many other scripts (R. Abelson, 1983, personal communication).

The third suggestion due to Tversky and Hemenway (in press) is that a good part is obligatory. Thus, every person has a *head* and *trunk,* and the vast majority have *arms* and *legs;* but it is more common (though still infrequent) for a person to be missing body parts that are defined at a lower level, like *fingers* or *toes.* This notion of obligatoriness corresponds well with our idea that scene headers are basic because they provide a useful level of expectation (Abbott & Black, 1980). If we know that someone went to a restaurant, the events corresponding to the scene headers—for example, *entering* and *ordering*—seem obligatory, while those corresponding to the scene actions—*getting a menu* and *looking over a menu*—seem more optional. This difference in predictability is a part of the explanation of the critical asymmetry we observed in our experiments: a reader is more likely to infer, say, *ordering* from *looking over a menu* than to infer *looking over a menu* from *ordering.* Our claim is that *Ordering* is the most detailed level at which it is useful to make inferences because more detailed inferences are likely to be wrong or irrelevant.

In sum, because of the difference in domains, one should be cautious in importing details of the analysis of object taxonomies to deal with event knowledge. A more useful analogy may be with object partonomies. However, the utility of the "basic level" concept applied to object taxonomies, partonomies, or scripts is the same. Items at the "basic level" should provide useful inferences for the task at hand. The behavior of subjects in our experiments leads to the conclusion that scene headers are basic for event knowledge, in that they provide inferences that are specific enough to provide useful information, but not so detailed as to provide information that is possibly erroneous and probably irrelevant.

Concluding Comments

Our finding that only part of a script gets inferred during reading and stored in memory is different from previous conceptions that assumed all actions in a script were inferred (e.g., Bower et al., 1979). An approach compatible with our results is the memory organization packet framework proposed by Schank (1982). In particular, Schank proposed that large scale knowledge structures like scripts are constructed when needed from smaller packets of information. In these terms, the representation of scripts we have described in this paper proposes that scripts are a hierarchically organized network of memory packets with a top level packet corresponding to a script header that indexes the scene header dominated packets. These lower level packets are composed of the scene actions that further detail the scene headers. Each packet contains information about the temporal ordering of its component parts. These packets are assembled during understanding into whatever combination is needed to understand the particular story being read.

Thus our results seem most consistent with a more flexible, generative representation for scripts than the rather rigid structures previously proposed. We have argued that a hierarchical network of organized information packets that can be combined in appropriate sequences is the best event representation for use in comprehension of events.

APPENDIX A

This is the text of one of the stories presented to subjects in Experiment 3. The script header and scene header from each script that were critical items in this experiment are surrounded by square brackets. In a given version of the story read by a subject, one, both, or neither of these items might have been included.

The Day

John went to bed early in preparation for his busy day. However, he had trouble falling asleep so when his alarm went off he still felt very tired. Getting ready for the coming day, he went into the bathroom. He took a long leisurely shower and put on the clothes he planned to wear. He had breakfast. He poured some cereal and drank his coffee.

As an afterthought, he put some stale pieces of bread in a bag and brought it with him. [He went to the bank where he had an account. He deposited some money in his bank account.] He hurried to the park and opened the bag, scattering some crumbs for the birds. Soon he was surrounded by a flock of pigeons. When it was time to go, he carelessly left the bag on the grass. [Then he attended his morning class for which he liked the professor. He listened carefully in class as the professor lectured.] He had a few minutes free, so he stopped by to see if his friend Paul was in his room. Paul had already left, so John continued on his way. [He exercised at the gym. He swam twenty laps in the indoor pool.] John was almost hit by a car as he crossed the street. He started cursing at the driver but stopped when people stared at him. [He ate lunch at a table in the school cafeteria. He carried some food to the table on a cafeteria tray.] He tried to get a newspaper from a machine, but it stole his money. Fortunately, just then someone else got a newspaper and held the door open so that John could get one too. [He needed a book so he went to the library. He checked out a library book.] He sat down on a bench outside and started to read his newspaper. He read the headlines, but got depressed quickly and turned to Dear Abby. He got a laugh out of that and turned to the sports section. Seeing an ad for an auto shop he remembered his car badly needed a tune up. He made a mental note of it as he got up and folded his paper. [He attended an afternoon music concert. He enjoyed listening to the music at the concert.] He went back to Paul's room and this time Paul was there. He offered John a cup of coffee and John gratefully accepted. He felt like he was getting a lot accomplished, but he was a little tired. When he mentioned the tune up to Paul he gave John the number of a gas station that had done some good work for him. John left feeling refreshed. [Afterward he went to see his doctor. His doctor gave him a physical examination.] When he got home, John called the gas station Paul had recommended. He made arrangements to take his car in on Wednesday morning. He wrote it down on the calendar by the phone so he wouldn't forget. [In the evening, he went to a cocktail party. He got drunk at the evening party.] Later, he went to a coffee house. He and his friends discussed philosophy. They argued about the meaning of life but did not come to any resolution. A folk singer provided background music. She had a good voice. However it was difficult to hear from their table.

Finally, John went home and watched TV. He checked the channels for something interesting. He rejected a Mary Tyler Moore rerun and ended up watching a war movie. He started getting sleepy. Settling comfortably in his chair, his head began to nod.

John woke up the next morning with a splitting headache.

REFERENCES

ABBOTT, V., & BLACK, J. B. (1980). *The representation of scripts in memory*. (Tech. Rep. No. 5) Cognitive Science Program, Yale University.

ABELSON, R. P. (1981). Psychological status of the script concept. *American Psychologist*, 36, 715–729.

ANDERSEN, E. S. (1975). Lexical universals in body-part terminology. In *Universals of human language,* Stanford, CA: Stanford Univ. Press, 1975.

BLACK J. B. (1984). Understanding and remembering stories. In J. R. Anderson and S. M. Kosslyn (Eds.), *Essays on learning and memory,* San Francisco: Freeman.

BLACK, J. B., & BERN, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior,* **20,** 267–275.

BLACK, J. B., & BOWER, G. H. (1980). Story understanding as problem-solving. *Poetics,* **9,** 223–250.

BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in text comprehension and memory. *Cognitive Psychology,* **11,** 177–220.

BRANSFORD, J. D., & JOHNSON, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior,* **11,** 717–726.

BRANSFORD, J. D., & JOHNSON, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.), *Visual Information Processing,* New York: Academic Press.

BROWN, C. H. (1976). General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist,* **3,** 400–424.

CULLINGFORD, R. E. (1978). *Script application: Computer understanding of newspaper stories.* (Tech. Rep. No. 116) Department of Computer Science, Yale University.

DOOLING, D. J., & LACHMAN, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology,* **88,** 216–222.

GALAMBOS, J. A., & RIPS, L. J. (1982). Memory for routines. *Journal of Verbal Learning and Verbal Behavior,* **21,** 260–281.

GRAESSER, A. C., HOFFMAN, N. L., & CLARK, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior,* **19,** 131–151.

LICHTENSTEIN, E. H., & BREWER, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology,* **12,** 412–445.

MANDLER, J. M., & JOHNSON, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology,* **9,** 111–151.

MILLER, G. A., & JOHNSON-LAIRD, P. N. (1976). *Language and perception.* Cambridge, MA: Harvard Univ. Press.

MILLER, J. R., & KINTSCH, W. (1981). Knowledge-based aspects of prose comprehension and readability. *Text,* **3,** 215–232.

MURPHY, G. L., & SMITH, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior,* **21,** 1–20.

ROSCH, E., MERVIS, C. B., GRAY, W., JOHNSON, D., & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology,* **8,** 382–439.

RUMELHART, D. E. (1977). Understanding and summarizing brief stories. In *Basic processing in reading, perception, and comprehension,* Hillsdale, NJ: Erlbaum.

SACERDOTI, E. (1974). Planning in a hierarchy of abstraction spaces. *Artificial Intelligence,* **9,** 119–135.

SCHANK, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people.* New York: Cambridge Univ. Press.

SCHANK, R. C., & ABELSON, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.

SMITH, E. E. (1981, May). *Studying on-line comprehension of stories.* Paper presented at the Cognitive Science Colloquium, Yale University.

SMITH, E. E., ADAMS, N., & SCHORR, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology,* **10,** 438–464.

THORNDYKE, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior,* **15,** 437–446.

THORNDYKE, P. W. (1977). Cognitive structures in the comprehension and memory of narrative discourse. *Cognitive Psychology,* **9,** 77–110.

TVERSKY, B., & HEMENWAY, K. (in press). Objects, parts, and categories. *Journal of Experimental Psychology: General.*

(Received March 1, 1983)
(Revision received December 12, 1983)