COGNITIVE PSYCHOLOGY 11, 177–220 (1979)

Scripts in Memory for Text

GORDON H. BOWER
Stanford University

JOHN B. BLACK
University of Illinois at Chicago Circle

AND

TERRENCE J. TURNER
Department of Education, Queensland, Australia

These experiments investigate people's knowledge of routine activities (e.g., eating in a restaurant, visiting a dentist) and how that knowledge is organized and used to understand and remember narrative texts. We use the term *script* to refer to these action stereotypes. Two studies collected script norms: people described what goes on in detail during familiar activities. They largely agreed on the nature of the characters, props, actions, and the order of the actions. They also agreed on how to segment the low-level action sequences into constituent "scenes," suggesting a hierarchical organization in memory of the activity. Other studies investigated memory for a text narrating actions from a script. Subjects tended to confuse in memory actions that were stated with unstated actions implied by the script. This tendency increased as more related script instances were studied. Subjects also preferred to recall script actions in their familiar order; a scrambled text that presented some script actions out of order tended to be recalled in canonical order. We also investigated whether the reading time for adjacent statements in a text varied with their distance apart in the underlying script. A statement at a one-step distance was read faster than one at a two- or three-step distance; statements in the second half of a script were read faster than those in the first half. A final experiment found that goal-relevant deviations from a script were remembered better than script actions. The role of script knowledge in text memory was discussed, as was the relation of scripts to schema memory in general.

We are interested in how people understand and remember narratives since this seems a promising way to investigate cognitive processes. A persistent problem for theories of narrative comprehension is to specify how people use their knowledge to expand upon what they are reading or hearing. Texts are usually elliptical and abbreviated, suggesting far more

We wish to thank Larry Meyers, Gail Meyers, Justine Owens, and Michael Gardner for assistance with various portions of this research. This research was supported by Grant MH-13905 to the first author from the National Institutes of Mental Health. Requests for reprints should be mailed to Dr. Gordon Bower, Psychology Department, Stanford University, Stanford, CA 94305. Dr. Turner was supported by a fellowship from the Australian Educational Council.

0010-0285/79/020177–44$05.00/0

than they say explicitly. A conversational postulate to "avoid prolixity and boring redundancy" may force such brevity. To understand a text fully, then, a model of comprehension must have methods for expanding upon an abbreviated text. Further, the model needs an organized knowledge store surrounding the topic of the text, which serves as a base for the elaborations.

Schank and Abelson (1977) proposed their "script theory" as a partial solution to the elaboration problem. They propose that part of our knowledge is organized around hundreds of stereotypic situations with routine activities. Examples are riding a bus, visiting a dentist, placing an operator-assisted telephone call, asking for directions, and so on. Through direct or vicarious experiences, each person acquires hundreds of such cultural stereotypes along with his idiosyncractic variations. Schank and Abelson use the term "script" to refer to the memory structure a person has encoding his general knowledge of a certain situation-action routine. The script theory is a specific elaboration of the frame theory of Minsky (1975).

The parts of a script are illustrated by the restaurant script in Table 1. As with other scripts, the restaurant script has standard roles to be played, standard props or objects, ordinary conditions for entering upon the activity, a standard sequence of scenes or actions wherein one action enables the next, and some normal results from performing the activity successfully. The information surrounding any one of these roles, props, or actions is assumed to be stored at varying levels of abstraction. For example, the Server Role in the restaurant must be a human, can be a male or female, and is usually dressed "appropriately" (e.g., is not wearing a suit of armor), and so on. The Server Role may be thought of as a list of alternative feature packages, with some features obligatory (e.g., must be alive), some optional (e.g., male or female), and some with weakly-bound ranges (e.g., age and style of dress).

A person's scripts are supposedly used in several ways. First, they aid planning and execution of conventional activities. The entering conditions and normal outcomes of scripts are examined during planning; the planner selects from memory a script whose normal result matches the current goal (e.g., satisfy hunger), then tries to bring about the entering conditions so she can perform the script. Second, scripts enable understanding when the person observes or reads about someone performing another instance of a conventional activity. We shall focus on this second use of scripts.

Whenever a text mentions a script-header (e.g., "The Restaurant") or a few lines that match parts of the memory script, the reader can "instantiate" the general script by filling in its variables (or "slots") according to the details mentioned. To illustrate, consider this vignette:

John went to a restaurant.
He ordered lasagna.
Later, he paid and left.

TABLE 1
THEORETICAL RESTAURANT SCRIPT (ADAPTED FROM SCHANK & ABELSON, 1977)

Name: Restaurant	
Props: Tables	*Roles:* Customer
Menu	Waiter
Food	Cook
Bill	Cashier
Money	Owner
Tip	
Entry Conditions: Customer hungry	*Results:* Customer has less money
Customer has money	Owner has more money
	Customer is not hungry

Scene 1:	*Entering*
	Customer enters restaurant
	Customer looks for table
	Customer decides where to sit
	Customer goes to table
	Customer sits down
Scene 2:	*Ordering*
	Customer picks up menu
	Customer looks at menu
	Customer decides on food
	Customer signals waitress
	Waitress comes to table
	Customer orders food
	Waitress goes to cook
	Waitress gives food order to cook
	Cook prepares food
Scene 3:	*Eating*
	Cook gives food to waitress
	Waitress brings food to customer
	Customer eats food
Scene 4:	*Exiting*
	Waitress writes bill
	Waitress goes over to customer
	Waitress gives bill to customer
	Customer gives tip to waitress
	Customer goes to cashier
	Customer gives money to cashier
	Customer leaves restaurant

The first line activates the restaurant script. John instantiates the role of the customer and lasagna is the food ordered. Because the brief text calls forth the full script, the reader can elaborate many objects and connections that are implied but not stated. The availability of these connections is suggested by the reader's ability to answer such questions as: Did John eat? What did he eat? Did he talk to a waiter or waitress? Did he receive a bill? What for? Such elaborated connections are frequently useful in understanding later parts of the story. For example, if later in the story John is found to have tomato stains on his shirt or professes not to be hungry, readers can guess how this came about from the restaurant scene.

Our experiments investigate some psychological implications of Schank and Abelson's script theory. Experiments 1 and 2 examine the organization of people's knowledge about stereotyped activities. What actions, roles, and props do people mention and how do they group or cluster these into subscenes? Experiments 3 and 4 ask whether, in remembering a text mentioning a subset of script actions, people tend to remember numerous unmentioned parts of the underlying script. Experiment 5 examines whether in recalling a text people will tend to recall the script actions in their stereotypic order even though the actions are mentioned in another order in the text. Experiment 6 asks whether the reading of earlier actions in a script speeds up the reading and comprehension of later actions in that script. Finally, Experiment 7 examines memory for occasional events, inserted into script-based stories, which interfered with or deviated from the smooth-running of the script.

EXPERIMENT 1: SCRIPT GENERATION

In this experiment, we collected "free association norms" for common scripts. If subjects did not agree about the essentials of a script, we would doubt the "cultural uniformity" assumption of script theory. However, it is not a foregone conclusion that everyone would describe a continuous action stereotype in the same way, using terms at the same level of specificity. There may not be a culturally uniform level of "basic action" description for scripts, as Rosch, Mervis, Gray, Johnson, and Boyer-Braem (1976) had found for basic object categories. Groups of undergraduates were asked to write scripts about common activities. The exact instructions to them turned out to be critical. If subjects were told to "write a completely common, boring story about a lecture or doctor visit" (which is one description of a script given by Schank & Abelson, 1977), their replies were quite variable. Our subjects in fact tended to write interesting stories about boring lectures or to describe frustrations caused by boring delays in waiting to see a doctor. Therefore, we revised the instructions to emphasize the subject's task of listing the central actions in a cultural stereotype.

Method

Materials. Each student generated events or actions for one situation. The five situations were attending a lecture, visiting a doctor, shopping at a grocery store, eating at a fancy restaurant, and getting up in the morning and getting off to school. Each subject received a blank sheet with appropriate instructions at the top. For example, the instructions for eliciting the lecture script were as follows:

> "Write a list of actions describing what people generally do when they go to a lecture in a course. We are interested in the common actions of a routine lecture stereotype. Start the list with arriving at the lecture and end it with leaving after

the lecture. Include about 20 actions or events and put them in the order in which they would occur."

Subjects. The subjects were Stanford undergraduates fulfilling a course requirement for their Introductory Psychology class. We handed out different scripts to differing numbers of subjects during one mass testing session with a group of 161 students. The numbers of subjects filling out and turning in the various scripts were as follows: grocery 37, getting up 35, restaurant 33, lecture 32, and doctor 24. The data were edited and tabulated according to frequency of citation of specific events, and their associated roles and props. Paraphrases and synonyms were lumped together.

Results

The issue is whether people agree in the actions they mention. The maximum diversity would be if all subjects generating a particular script mentioned once 20 or so completely unique events. But what is surprising is how much agreement there is in the "basic action" language that people use to describe the activities. This uniformity is reflected in how few of the events were mentioned by only one person. For example, in the restaurant script, of 730 actions mentioned in total (types times tokens), only four were completely unique (given by a single person). Similarly, the ratio of unique mentions to total events was 4/704 for Lecture, 26/814 for Grocery, 26/770 for Getting up, and 36/528 for Doctor (which had the fewest subjects and least chances for overlap). So there is at least someone who agrees with nearly every action that any subject writes for a script.

Furthermore, there is high reliability in the frequency with which particular actions of a script are mentioned. We divided each group in half and correlated the frequencies of mentioning specific actions by the two halves. The Pearson correlations were Restaurant .88, Lecture .81, Grocery .85, Getting up .87, and Doctor .80. Thus, the frequency norms are reliable, at least with a homogeneous group like Stanford undergraduates.

Each subject mentioned a sample of very common actions along with some less common ones, presumably reflecting his experiences and speaking style. Across subjects, there was a continuous gradation in frequency of mention of particular events. We may arbitrarily designate the group's stereotype or script to be those events mentioned by more than some criterion percentage of subjects. Examining the distributions of how many actions were mentioned by varying numbers of respondents, the distributions had similar shapes with a distinct gap near 25% agreement. So we selected 25% mention as a lenient criterion for inclusion of an action in Table 2. Table 2 reports for the five situations each action mentioned by at least one-fourth the respondents. The actions are listed in the serial order in which they are usually mentioned. Two other criteria were selected at natural breaks in the percent-agreement distribution for each situation. The items in italics were more popular, falling above a criterion of 40–50% mention; actions in capitals were the most popular, having been mentioned by 55–75% of subjects (varying with scripts).

TABLE 2
EMPIRICAL SCRIPT NORMS AT THREE AGREEMENT LEVELS

GOING TO A RESTAURANT	ATTENDING A LECTURE	GETTING UP	GROCERY SHOPPING	VISITING A DOCTOR
Open door	ENTER ROOM	*Wake up*	ENTER STORE	*Enter office*
Enter	*Look for friends*	Turn off alarm	GET CART	CHECK IN WITH RECEPTIONIST
Give reservation name	FIND SEAT	Lie in bed	Take out list	SIT DOWN
Wait to be seated	SIT DOWN	Stretch	Look at list	Wait
Go to table	Settle belongings	GET UP	Go to first aisle	Look at other people
BE SEATED	TAKE OUT NOTEBOOK	Make bed	*Go up and down aisles*	READ MAGAZINE
Order Drinks	*Look at other students*	*Go to bathroom*	PICK OUT ITEMS	*Name called*
Put napkins on lap	*Talk*	Use toilet	Compare prices	Follow nurse
LOOK AT MENU	Look at professor	*Take shower*	Put items in cart	*Enter exam room*
Discuss menu	LISTEN TO PROFESSOR	*Wash face*	Get meat	Undress
ORDER MEAL	TAKE NOTES	Shave	Look for items forgotten	*Sit on table*
Talk	CHECK TIME	DRESS	Talk to other shoppers	Talk to nurse
Drink water	Ask questions	Go to kitchen	Go to checkout counters	NURSE TESTS
Eat salad or soup	Change position in seat	Fix breakfast	*Find fastest line*	Wait
Meal arrives	Daydream	EAT BREAKFAST	WAIT IN LINE	Doctor enters
EAT FOOD	Look at other students	BRUSH TEETH	*Put food on belt*	Doctor greets
Finish meal	Take more notes	Read paper	Read magazines	Talk to doctor about problem
Order Desert	*Close notebook*	*Comb hair*	WATCH CASHIER RING UP	Doctor asks questions
Eat Desert	*Gather belongings*	*Get books*	PAY CASHIER	DOCTOR EXAMINES
Ask for bill	Stand up	Look in mirror	*Watch bag boy*	Get dressed
Bill arrives	Talk	Get coat	Cart bags out	Get medicine
PAY BILL	LEAVE	LEAVE HOUSE	Load bags into car	Make another appointment
Leave Tip			LEAVE STORE	LEAVE OFFICE
Get Coats				
LEAVE				

Items in all capital letters were mentioned by the most subjects, items in italics by fewer subjects, and items in small case letters by the fewest subjects

The 22 to 25 entries for the different scripts appear to capture common experiences in our culture. The actions mentioned the standard characters in the script, the usual "props" and locations for the actions. The fewer actions at the more stringent criteria also seem to be the more central or important ones of the script, with the actions in all capitals being the most important. In recent pilot work, Masling, Barsalou, and Bower collected "importance ratings" for events within scripts. The most frequently mentioned events of Table 2 were also rated as centrally important to the script. Such a high frequency-high importance event seems to correspond to what Schank and Abelson (1977, p. 45) call a "main conceptualization," an event which is essential within its scene in that subordinate actions within that scene depend upon it, often enabling or resulting from it. These main conceptualizations are likely to appear in a summary or synopsis of a script-based text. Furthermore, mention of one of these central events should act as a powerful probe to call up the script from the reader's memory.

Discussion

The script norms in Table 2 are the main outcome of this study. There was considerable agreement in the way subjects described events. For example, people wrote "He ate his soup" rather than "He picked up his soup spoon, dipped it into the cup of soup, lifted it out, blew on it to cool it, raised it to his lips, put the spoon in his mouth. . .". Presumably people use action-summary terms or "basic-level action" terms (see Rosch *et al.*, 1976) to describe a continuum of events because they share a conversational postulate which says one ought to speak or write so as to be informative but not overly redundant (Grice, 1975). Because people know how to eat soup, it is ordinarily unnecessary and ill-mannered for a speaker to describe the steps in such detail. Perhaps it is these habits of speaking that lead our subjects into event descriptions at roughly the same basic level. Of course, a script itself is a familiar routine whose recital would normally be redundantly boring and gauche. But the experimental setting and our instructions for the generation task conveyed our interest in the subject's telling us all these normally boring details within a script.

We conceive of a standard activity like eating in a restaurant or visiting a dentist as a fuzzy concept for which there are many characteristic features but few if any defining features (see Smith, Rips, & Shoben, 1974; Rosch & Mervis, 1975; Zadah, 1965). Different instances of an activity seem to bear a "family resemblance" to one another, but they may possess no common features. For example, although eating would seem to be a general event of the restaurant script, one can be "in the restaurant script" without eating just as one can eat in many contexts besides the restaurant script (e.g., tasters, food judges). Events or actions are more or less diagnostic of a given activity or script. A diagnostic event or feature is statistically valid in the sense that its presence or absence correlates highly across situations with the presence or absence of the scriptal activity. A given script-instance should be judged as prototypical of the activity to the extent it combines highly valid features. As Rosch and Mervis have shown, the most prototypical exemplar is judged to resemble on average more other instances of the category and to bear less resemblance to instances of alternate categories. These properties of fuzzy concepts seem particularly appropriate to the notion of scripts.

The scripts of Table 2 should be useful for further experimental investigations. We have used them in investigations of memory for scripts and of comprehension of script-based texts. The scripts can be used like high-frequency association norms, much like the Battig and Montague (1969) category norms have been used in studies of word perception, priming, semantic judgments, and memory. As one example, an experiment in progress is checking to see whether a given action can be classified more rapidly as "fitting" a given script header if it is a high associate of the script. Further, one could check whether high-frequency action associates of a script are better recalled but more poorly recognized than low-frequency actions of the script. High frequency actions, if not mentioned in a text, may later attract many false-positive recognitions because they were implicitly aroused during reading the text.

EXPERIMENT 2: CONSTITUENT STRUCTURE OF SCRIPTS

The events within a lengthy ordered script appear to be segmented naturally into several major chunks or constituents. The script is not an undifferentiated, linear chain, but rather seems organized into major scenes, with those composed of subsequences of actions. Thus, eating in a restaurant may have as major scenes entering, ordering, eating, paying, and leaving. But ordering requires being seated, getting and reading a menu, having a waitress come to take your order, and so on. To check our intuitions we asked subjects if they thought there was a "natural segmentation" of the lower-level action sequences in a script. If they segmented the actions into chunks, we were then interested in whether they agreed on the location of the chunk boundaries.

Method

Ten texts were written by converting all of the actions in 10 underlying scripts into actual story statements. The texts ranged in length from 148 to 254 words. The scripts used were going to a restaurant, getting up in the morning, attending a lecture, going to a birthday party, going swimming, grocery shopping, making coffee, visiting a doctor, attending a movie, and playing football.

Thirty Stanford undergraduates were given a booklet containing the 10 stories, each on a different page. They were told that some people felt that each story had several natural parts; they were to read the stories, decide whether a story had some parts and, if so, identify these parts by placing a slash line in the texts marking the end of each part. They were given no hint as to how many slashes (chunk boundaries) to place in each text, if any.

Results

The main issue is whether people agree in their chunking judgments. That is, are there locations in the stories where most people put slashes to indicate a part boundary, and other locations where very few people put slashes? We first divided the story into script action statements, since subjects only placed slashes at such clause boundaries. We tabulated the frequency of slashes (chunk boundaries) at each site. Table 3 displays these slash frequencies in brackets after each clause for two of the ten texts. To measure the agreement of the location of slashes, we divided the

TABLE 3
THE CHUNK JUDGMENTS OF EXPERIMENT 2

The Doctor

Diane was feeling very bad today, [0] so she decided to go see the family doctor. [7] Therefore she had her husband take her to the doctor's office. [17] / When she arrived at the doctor's office, [1] she went into the waiting room. [1] Once inside she walked over and checked in with the receptionist [0] and then sat down to wait her turn. [4] As she waited she read some old medical magazines that were on the table [0] and looked at the colorful medical posters that adorned the wall [17] / Finally the nurse came in and called her name, [0] so she went into the examination room with the nurse. [7] The nurse closed the door and asked her to take off her clothes. [0] The nurse then weighed her and took her blood pressure. [0] When these preliminaries were completed the nurse left [7] and a short while later the doctor came in. [5] The doctor was very nice to her, so she calmed down a little bit. [15] / As the doctor started to make various examinations of her [3] she wondered what he was doing [0] and what he was finding. [5] Finally he looked directly at her and told her that she had the flu and could expect to be laid up in bed for a few days. [0] Then he wrote a prescription for some pills [0] and left [22]. / She got dressed [0] and made another appointment with the receptionist [0] on her way out of the doctor's office.

The Restaurant

David noticed that his stomach was emitting hunger pains, [1] so he decided to go out to a restaurant to eat. [6] Therefore he drove to the local French restaurant. [19] / He arrived at the restaurant a little before the dinner rush hour, [1] so as he entered the restaurant [0] he noticed that there were plenty of empty tables. [0] He decided to sit at a window table, [5] so he went over [0] and sat down. [20] / A waitress came up [0] and gave him a menu. [1] He carefully perused the menu [0] and decided what he wanted. [4] The waitress came back [0] and he gave her his order. [13] / After a short wait [0] during which he nibbled on bread and butter [1], his dinner arrived. [13] / He proceeded to eat the dinner with gusto. [2] The food here was really excellent [0] and not too expensive either. [16] / Finally he finished [0] and asked the waitress for the check. [9] He left the waitress a tip on the table [0] and went over to the cashier. [2] He paid the cashier [1] and went home quite satisfied.

The numbers in brackets indicate the number of subjects (out of 30) who marked those locations as boundaries between parts of the story. Slashes mark the major constituent boundary locations (i.e., those marked by 10 or more people).

group in half and correlated their slash frequencies at the clause boundaries for the several story locations. The Pearson correlations for the different scripts were as follows: getting up .98, coffee .94, lecture .99, doctor .99, birthday .96, swimming .95, movie .98, football .98, shopping .96, and restaurant .98. Next, we performed chi-square tests to determine whether the distributions of slashes differed from a uniform distribution. The chi-squares for all scripts were significant, with all p's $< .001$.

We considered how to describe the group's agreement on number and location of chunk boundaries. We examined the number and location of story constituents selected by varying numbers of people. All 10 of the scripts yielded a roughly bimodal distribution: there were a few sites between action clauses where most of the subjects placed slashes (e.g., 18 to 20 subjects); then there were many locations where few subjects placed slashes (e.g., three, four, five, and six subjects), finally, there were very few sites selected by a moderate number of subjects (e.g., 10 to 12). In nine of the 10 scripts, the one-third percentage (10 subjects) fell in a natural gap in the slash distribution, so we selected one-third as the cutoff point. By this criterion the 10 scripts each have three to five constituent boundaries. Slashes locate these major constituent boundaries in the texts of Table 3.

One guide to determining the constituents of a sentence is that "a constituent is a group of words that can be replaced by a single word without change in function and without doing violence to the rest of the sentence" (Clark & Clark, 1977, p. 48). We find that we can substitute summary-actions for our empirically derived constituents in scripts. For example, in the Doctor story given in Table 3, we can replace the first constituent by "Diane went to the doctor's office," the second by "She waited," the third by "She went through the preliminaries," the fourth by "The doctor examined her," and the fifth by "She left the doctor's office." Note that if we try to move any boundary to encompass adjacent text actions, the one-action summaries no longer fit. The other script boundaries show similar properties, so our empirically derived script boundaries fit the summary criterion of a constituent.

Discussion

Clearly, our subjects agreed with our intuitions that a continuous script activity can be segmented into chunks or scenes. And they agreed with one another where the scene boundaries were located in the event sequence. Thus, the script is not a linear chain of actions at one level but rather a hierarchically organized "tree" of events with several levels of subordinate actions. That activities are decomposable into a hierarchy of subactions is a recurrent theme in cognitive psychology (e.g., see Bower, 1975; Goodman, 1970; Miller, Galanter, & Pribram, 1960; Pew, 1974).

Actions may be identified with their intended goals or subgoals. From this perspective, the activity hierarchy is really a goal tree, wherein a top goal is decomposed into a series of subgoals. Thus, the top goal of eating in a restaurant script can be decomposed into the subgoals of getting inside a restaurant, sitting down and ordering, eating the food received, paying the bill, and leaving. In turn the subgoal of ordering the food can be decomposed into events of getting a menu, reading it, getting the waiter to your table and telling him what food you want. The relevance of such goal hierarchies is that they are frequently useful for answering questions about specific actions or events (see e.g., Schank & Yale A. I. Project, Note 4; Winston, 1977, p. 301 ff.). The following strategies often work:

> *Why*-questions (e.g., why did you speak to the waiter?) can often be answered by moving one level up from the queried action in the hierarchy and stating that goal (e.g., "Because I was ordering"). Repeated *Why's* may force the respondent to the top of the tree, where sits the script-entry reason (i.e., I ate to reduce hunger; and I did that to stay alive).
>
> *How* questions can often be answered by moving one level down in the hierarchy from the queried action, and listing its subordinate actions. Thus, "How did you order?" can be answered by "I translated the French menu, called over the waiter, and pointed to what I wanted." If the queried action has no subordinates in the tree, an acceptable answer is "I just did it," (as in answer to "How did you swallow your food?").
>
> *When* questions (e.g. when did you read the menu?) can be answered by referring either to the goal-activity one level up and using *while* ("While I was ordering"); or by referring to preceding and succeeding actions at the same level. Thus, we can reply "I read the menu just *after* the waiter left it on my table, and *before* I gave him my order."

Besides helping answer questions, the action hierarchy can be used by the person generating summaries of script-based texts. Any sequence of subordinate actions within a given chunk can be summarized by the superordinate action (Rumelhart, 1977; Schank & Yale A. I. Project, Note 4). Thus, "John read the menu, decided on his selection, and told the waiter what he wanted to eat" can be summarized as "John ordered."

Given this information about chunk structures in scripts, several psychological experiments would seem called for. Our own efforts along this line have been limited to date. One pilot study is that mentioned earlier by Masling, Barsalou, and Bower which had subjects rating the importance of various scenes to the goal of the script. The most important scenes within a text should be those most likely to survive in memory over a retention interval. Another pilot study of ours investigated the number of false alarms that occurred in recognition memory for unstated actions within a chunk for which either one or two other actions of that chunk were stated in a text the subject read. Contrary to expectation, the number of false alarms was lower to an unstated action when two actions within that chunk were also stated than when only one of those actions was stated. Other studies using the chunk-boundary norms could examine all-or-none recall of chunks, probe latency for recall of the "next successor" by a within-chunk vs a prior-chunk action (see Ammon, 1968), and the process of purging or clearing out of short-term memory as a sequence of script-action sentences passes over a chunk boundary (see Jarvella, 1970, 1971). Clearly more research on the topic is needed.

EXPERIMENT 3: SCRIPT RECALL

Next we investigate recall of texts composed of selected lines from an underlying script.[1] The question was whether in recalling such a text subjects will use the underlying script to fill in gaps of intervening actions not explicitly mentioned in the text. We might expect some such intrusions in recall since they correspond to the familiar phenomenon of changing the working of a story when we retell it "in our own words."

The script idea gives us a way to think about paraphrases of a story. If a sequence of actions calls up an underlying script from memory which is then filled out according to the particulars of the text, then the "same story" can sometimes be retold by mentioning a different selection of its actions. If a b c d e f g represent the underlying script events and Text 1 comprises sentences instantiating a c e g, then Text 2 comprised of congruent instantiations of b d f could be judged to be an acceptable paraphrase of Text 1 despite having no events in common. Such a possibility is a consequence of conceiving of script instances as only bearing a family resemblance to one another. Similarity of two texts will probably be a systematic function of the amount and importance of their overlapping as opposed to their distinguishing features (see Tversky, 1977). It is an empirical matter to explore in detail which parts of a script-based text can be replaced or altered without changing the intuition that the second text is a paraphrase of (or closely similar to) the first. The importance of similar paraphrases for studies of memory is that we can expect the person's memory to be confused between two paraphrases of the same script, sometimes substituting or intruding in recall actions that were implied but not stated in the text.

We may expect subjects to remember for some minutes the events explicitly stated in the text. But as their "surface memory" of the text

[1] Note our terminology. A *script* refers to a generic memory structure in a person's head (e.g., visiting a Health Professional). A script-based *text* or *story* is a list of sentences, read by the subject, most of which denote actions of an underlying script. A given script may be instanced by different texts (e.g., Diane visiting her doctor or dentist). An *instantiated script* is an episodic memory structure set up in the reader's head to encode and remember a particular script-based text; it is the "memory trace" of reading the story of Diane visiting her doctor.

fades, they should intrude more assertions into recall which represent implications and which in theory were used to fill-in the gaps between the script actions read originally. A model for this might suppose that after reading the person has both a veridical memory for the actually read statements (in "episodic memory") and an activated and completely filled out underlying script. In immediate recall, the person merely reproduces his veridical memory. But this memory fades over time and he relies then upon the fully-completed script, which leads to unstated script actions being intruded into recall.

We were also interested in a second phenomenon which, if found, may prove harder to explain. We wondered whether we could increase the subject's belief that an unstated action had occurred in a script-based text by having him read *related* script stories in which the analogous or parallel actions were explicitly mentioned. The kind of parallel stories used are illustrated in Table 4, one for John visiting a doctor, the other for Bill visiting a dentist. These are different instances of an abstract script for visiting a health professional.

The actions enclosed by parentheses in Table 4 were not included in the texts read by the subjects, but are part of the Health Professional script. Notice that some of the actions left out in the Doctor text have their parallel actions explicitly stated in the Dentist text, and vice versa. For

TABLE 4
SAMPLE TEXTS USED FOR EXPERIMENTS 3 AND 4

	"The Doctor"		*"The Dentist"*
Entry:	John was feeling bad today.	Entry:	Bill had a toothache.
1.	John decided to go see the doctor.	1.	(Bill decided to go see the dentist.)
2.	(John arrived at the doctor's office.)	2.	Bill arrived at the dentist's office.
3.	(John entered the doctor's office.)	3.	(Bill entered the dentist's office.)
4.	John checked in with the doctor's receptionist.	4.	(Bill checked in with the dentist's receptionist.)
	. . .		. . .
7.	John looked at some medical magazines.	7.	(Bill read some dental magazines.)
	. . .		. . .
12.	(The nurse checked John's blood pressure and weight.)	12.	The dental hygenist x-rayed Bill's teeth.
	. . .		. . .

Sentences in parentheses were presented only during the recognition test in Experiment 4.

example, the Dentist text mentions nothing about Bill checking in with a receptionist whereas the Doctor text mentions that John checked in with the Doctor's receptionist. We predicted that in a memory test people would tend to recall gap-filling events left out of a script story, and they would tend to do this more if they also read other instantiations of this script which did mention the analogous actions. The experiment below tests this specific prediction, that recall intrusions of gap-fillers will increase if other instances of the same script are also read.

Method

Materials. There were nine basic scripts each with about 20 actions (and thus similar to the 25% cutoff level in Experiment 1). Each script was chosen to have at least three distinctive versions (e.g., the Health Professional script had the Doctor, Dentist, and Chiropractor versions). Hence the materials were a total of 27 specific script versions, so it was impractical to collect and analyze script norms to serve as a basis for all these materials (one would need around 30 subjects per action list for 27 action lists). Therefore these scripts were created by one of the experimenters (JB) and a colleague. We individually composed a list of common actions for each generic script, then compared our versions and reached a compromise prototype. Three stories were then written based on each script; each story was about eight lines long and represented a different instance of the script. Hence, the story pool contained nine clusters of three stories per cluster: attending a symphony, play, or movie; visiting a doctor, dentist, or chiropractor; playing a game of football, baseball, or golf; going swimming, skin diving, or surfing; getting ready in the morning to go to school, church, or work; attending a class lecture, sermon, or speech; shopping for bread, a coat, or a toaster; going to a birthday party, New Year's Eve party, or Halloween party; and making coffee, tea, or hot chocolate. Each subject read one story each for three of the nine scripts, two stories each for another three scripts (i.e., six stories), and three stories each for the other three scripts (i.e., nine stories). Hence, each subject read a total of 18 stories, presented in random order. Which story fell into which category was balanced across subjects. Each story was also given a title (e.g., The Doctor, Football, etc.).

To clarify complex relations among instantiations, the three texts for Visiting a Health Professional are given here.

The Doctor

John was feeling bad today so he decided to go see the family doctor. He checked in with the doctor's receptionist, and then looked through several medical magazines that were on the table by his chair. Finally the nurse came and asked him to take off his clothes. The doctor was very nice to him. He eventually prescribed some pills for John. Then John left the doctor's office and headed home.

The Dentist

Bill had a bad toothache. It seemed like forever before he finally arrived at the dentist's office. Bill looked around at the various dental posters on the wall. Finally the dental hygienist checked and x-rayed his teeth. He wondered what the dentist was doing. The dentist said that Bill had a lot of cavities. As soon as he'd made another appointment, he left the dentist's office.

The Chiropractor

Harry woke up with a bad pain in his back again. He decided to go see a chiropractor that very day. He had to wait a long time. Finally the chiropractic assistant finished and left him, and the chiropractor himself came in. The chiropractor carefully examined Harry by feeling all the bones in his back. Eventually Harry left the chiropractor's office.

Note, the stories are about seven or eight lines long, and they share only the common entering and leaving conditions. The middle six or so actions of each story were selected without replacement from the 20 actions comprising the master Health-Professional script. A third of the subjects would read all three of these texts scattered through the study phase; a third read a randomly selected two of them; and a third read a randomly selected one of them. The same routine was followed with each of the nine master scripts.

Procedure. The subjects were run in groups of two to eight. As a warmup, subjects first read for 10 min a coherent narrative that was completely unrelated to the material of this experiment. They were then given reading instructions, given their story booklet for the present experiment, and had 10 min to read and study it. This study time was sufficient for all subjects to read through their 18 stories. After reading, the subjects performed an intervening task for 20 min. This task involved their recalling in writing the narrative that they had read at the beginning of the session. After that they were instructed regarding recall of the 18 scripts of this experiment. The script titles were read as cues, and subjects wrote their recall of each corresponding text on a new sheet of paper. They were given 1 min to recall each of the eight-line stories. They were asked to be accurate and to reproduce each text verbatim insofar as possible, but they were to recall the gist of an event if they could not remember it verbatim.

Subjects. The subjects were 18 students at California State University at Sacramento who participated to fulfill a service requirement for their Introductory Psychology course.

Results

The recalls were scored for the presence or absence of the underlying script actions. We had no trouble deciding to which script action a given sentence in recall referred. We classified the actions in each person's recall into stated script actions, unstated script actions, and other actions that did not fit these two categories. The numbers of recalled actions falling into these three categories are displayed in Table 5. The rows of the table correspond to the conditions of one, two, or three story instances of a script.

Before listing statistical tests, let us recognize the current controversy over whether investigators using language materials should use fixed or mixed effect analysis of variance (ANOVA) models to test hypotheses (Clark, 1973; Wike & Church, 1976; Clark, Cohen, Smith, & Keppel, 1976). In our case, the question is whether to treat the scripts as a random sample of all scripts and thus as a random effect in the ANOVA model, or to limit the conclusions to these 18 scripts and thus treat them as a fixed effect in the ANOVA model. We report statistical tests for both models.

TABLE 5
AVERAGE NUMBERS OF ACTIONS RECALLED PER SCRIPT VERSION IN EXPERIMENT 3

		Number of stated script actions	Number of unstated script actions	Other actions	Total actions recalled	Percent of total that are unstated script actions
Number of script versions	1	3.03	0.80	0.39	4.22	19
	2	2.27	1.26	0.35	3.88	31
	3	2.56	1.16	0.36	4.08	28

Happily, only rarely in this paper does the test used influence the pattern of significances, and we will note whenever it does.

A first feature of the data in Table 5 is that stated script actions are recalled more than unstated script actions, which in turn are more frequent than other actions. These pairwise comparisons are statistically significant at .01 for the three- and two-instance rows, and at .05 for the one-instance row. Stated actions exceed unstated ones significantly considering the differing materials exemplifying a condition as either a fixed or random effect (using an F test for a fixed effect ANOVA model or a quasi-F, F', test for a random effect ANOVA model). So, while we may conclude that unstated script actions appear in recall in appreciable amounts, subjects nonetheless display considerable "surface memory", at least at a 20-min retention interval, since they are producing two to three times more stated than unstated actions. The data in Table 5 are raw frequencies. To convert them to percentage of script actions given, one must divide the left hand entries by the number of stated actions (eight) and the middle entries by the number of unstated actions (arbitrarily 12 here, though indeterminate in reality). Thus the proportionate reproduction of stated vs unstated actions is even more pronounced than the raw frequency results.

Next consider the influence of having read two or three script instances upon intrusion of unstated script actions (the "gap fillers" in Table 4). Since recall levels varied somewhat across conditions, we expressed the intrusions of unstated script actions as a percentage of the total recall, and entered this in the last column of Table 5. Clearly the percentage of unstated script intrusions increase when other story instances of the same script are present. Comparing the percentages of gap-fillers to all actions recalled, scripts with two story versions have many more than do scripts with one story version [$F'(1,18) = 10.68, p < .01$]; similarly, three story versions lead to more gap-fillers than does one story version. However, three and two do not differ in intrusion rate.

So, we may conclude that having another version of the same script mention an action increases the probability that the unmentioned analogous action is intruded in recalling a related story. We reserve theoretical discussion of this priming effect until after the results of Experiment 4 are presented. Unfortunately, the priming effect does not increase steadily as the number of script versions is increased. We have no explanation for this leveling-off.

Considering recalls of the stories having one, two, and three instances of the underlying script, it is clear that they do not differ in the number of "other actions" recalled nor in the total recall. However, the conditions differ in recall of stated actions with the one instance stories being higher. This effect is significant considering materials as a fixed effect [$F(2,30) = 4.42, p < .05$], but not when materials are considered as a random effect ($p > .05$).

If reliable, the lower recall of stated actions with multi-instance texts suggests several interpretations. First, because multi-instance texts create more intrusions, there are more opportunities for "output interference," whereby output of an intrusion causes forgetting of a not-yet-recalled stated action (see Roediger, 1974). Second, if people normally avoid reporting redundantly close (and therefore uninformative) actions from a script, then the increased intrusions of implied actions in the multi-instance conditions would cause increased editing out and nonrecall of nearby stated actions.

Finally, examination of the "other actions" category in recall reveals that 40% of them are superordinate actions or terms summarizing lower-level script actions. For example, if the text said "She took out the coffee pot, filled it with water, put it on the stove, . . . ," a summarizing "other" statement would be "She heated some water." Such summarizing statements are expected, considering the underlying script as a hierarchical tree with the summarizing terms referring to nodes at the higher levels of the tree.

An unexpected result is that the proportion of superordinate actions among the "other actions recalled" category is over twice as great (62.5%) for the recall of the one-instance stories as for the averaged recall of the two- and three-instance stories (28.3%). This effect is statistically reliable [$F(1,5) = 19, p < .01$]. However, we have no explanation for this curiosity.

EXPERIMENT 4: SCRIPT RECOGNITION

The recalls in Experiment 3 showed the effect of interest, namely, a fairly high intrusion rate of unstated actions from the underlying script. Moreover, this intrusion rate when recalling a story could be raised by presenting the counterpart actions in a parallel story that is a different version of the same underlying script. Our next experiment aimed to demonstrate these effects more clearly using recognition memory. In the recognition test, the subjects saw a test sentence from the underlying script and assessed how well it matched anything within the specific subset of sentences stored in memory from when they had read the text. To the extent that studying a sentence describing an action activates associated actions in the underlying script, those associated actions will later appear to have been read. This prediction arises because, during testing, the person allegedly only knows the level of activation in memory corresponding to the test proposition, but is confused about the source of this activation, whether it is due to an explicit or implicit presentation. This failure to discriminate the source of activation is one supposed cause of false-positive recognition judgments.

Method

Materials. The learning materials were the same titled script stories used in Experiment 3. The presentation of the stories and the counter-balancing of materials across experimental conditions were the same as before.

Procedure. Subjects first read two unrelated, 600-word stories for 5 min. These stories were unrelated in content to the scripts of this experiment. Subjects then read the 18 script stories of this experiment, each with a title. They had 5 min to read these stories. They then spent 30 min writing recalls of the first two stories. Finally they were given the recognition test over the 18 script stories. The recognition test contained 16 test statements for each of the nine generic scripts (or $9 \times 16 = 144$ in total). These were blocked by story title. For each generic script the recognition test contained eight stated script actions, four unstated script actions, and four other actions that were false though not implausible. Typical false items mispaired actors and actions, or locations and actions, from the stories studied. When two or three story versions of the same generic script were used, the 16 recognition items were chosen evenly from the two or three stories. The subjects rated each test statement on a 7-point scale, with 1 denoting "Very sure I did not read this sentence" and 7, "Very sure I did read it."

Subjects. The subjects were 45 Stanford University students who were fulfilling an Introductory Psychology course requirement. They were tested in groups of two to eight.

Results

The primary results are the average recognition ratings for the various item types. These are shown in Table 6 for stories in the one, two, and three script version conditions. A high number indicates that the subjects believed the statement had been in the texts read a half hour earlier.

In general, ratings are highest for stated actions, intermediate for unstated script fillers, and lowest for novel false statements. These pair-wise comparisons are statistically significant at .01 within each row of Table 6; they are significant considering learning materials as either a fixed or random effect.

Comparing conditions with varying numbers of instances, recognition ratings for stated items are equivalent. So the recognition accuracy measure does not uphold the small difference between one vs two or three versions found earlier with the recall measure. There is a slight difference in ratings of False items which is significant treating learning materials as a fixed effect [$F(2,84) = 5.33, p < .01$] but not as a random effect [$F'(2,28) = 1.29, p > .10$].

The most salient difference between the different instance conditions occurs with unstated script fillers. As predicted, false-positive recognition ratings for an unstated script filler increase as more instances are put into memory. The differences in ratings on unstated script fillers for one, two, or three instances are highly significant; fixed effect ANOVA yields [$F(2,84) = 31.3, p < .001$]; random-effect ANOVA yields [$F'(2,17) = 5.35, p < .05$]. Individual comparisons show that conditions one and two differ, and one and three differ by both F and F' tests. However, although two and three differ at the .05 level by a fixed-effect F test, they are not significant by a random-effect F'.

The ratio score in the last column of Table 6 scales ratings of script fillers relative to the hit rate on Stated actions and the false positive rate on Falses. The ratio score is 100 (U−F)/(S−F), where S, U, and F stand

TABLE 6
AVERAGE RECOGNITION RATINGS PER TEST ITEM IN EXPERIMENT 4[a]

		Stated script actions (S)	Unstated script actions (U)	Other actions (O)	Ratio (R)
Number of script versions	1	5.46	3.91	1.71	54.6
	2	5.40	4.62	1.76	66.3
	3	5.59	4.81	1.86	68.6

[a] 7 means "Sure Old" and 1 means "Sure New."

for Stated trues, Unstated script fillers, and False items, respectively.The ratio scores differ reliably, mainly because the one-instance ratio is below the other two. The two- and three-instance conditions do not differ reliably in this ratio measure.

After the data were collected, we noticed that our test items for unstated script actions for two or three script instances were actually of three types. The types are illustrated by the example fillers in Table 4. One type of action confusion is shown in lines 1, 2, 4, and 7 of Table 4; it involves mixing up, say, John's arriving at the Doctor's office with Bill's arriving at the Dentist's office. False alarms to such gap fillers would reveal the person mixing up actions, objects, and actors in a simple manner. Line 3 of Table 4 illustrates the second type of action confusion. These items are script actions that were not stated in any of the story versions read. The third type of action confusion is shown in Line 12 of Table 4; here the similarity of the sentences in the two stories stems from their exemplifying the same role or function in the abstract script. For example, the abstract health-professional script has an action in which an assistant performs some preliminary procedures on the patient. An instantiation of this role for the Doctor script is "The nurse checked John's blood pressure and weight;" for the Dentist story an instantiation is "The dental hygenist checked and x-rayed Bill's teeth." Clearly, taking John's temperature has little direct similarity to x-raying Bill's teeth. Rather they are similar by virtue of the roles they play in the abstract script.

We wondered whether the effect of number of instances in increasing false alarming to gap fillers held true for all three types of recognition items. There were many more items of the first type than of the second and third; furthermore, the counterbalancing of story versions into the one-, two-, and three-instance conditions caused a marked loss of observations on any specific abstract inference. Putting aside these problems, however, we computed the average recognition ratings for the second and third types of unstated script actions mentioned above, for the one-instance condition vs the combined two- and three-instance conditions. For the direct inferences not stated in any text (second type above), the ratings for one vs two and three instances were 3.49 and 4.75, respectively (one vs two [$F(1,28) = 28.06, p < .01$]; one vs three [$F(1,28) = 18.15, p < .01$]; two vs three is not significant). For the abstract inferences (third type), the ratings for one vs two and three instances were 2.73 and 4.73, respectively (one vs two [$F(1,28) = 7.84, p < .01$]; one vs three [$F(1,28) = 12.11, p < .01$]; two vs three is not significant). There is a clear effect of other text instances on these two classes of confusion errors. The effect is as large as that obtained for simple actor-action mix-ups which comprise the majority of cases in the composite averages in Table 6. The fact that the third type of inference yields an equal-sized effect is remarkable since the basis for the similarity of two action-statements in this case is not their meaning per se but rather the abstract role they fulfill in the two scripts. In closing we would repeat that the comparison of the effect of number of instances for the three inference types is not balanced across materials, and we would be more confident if we had designed our experiment for this comparison. However, the current results suggest optimism that a properly balanced experiment would replicate the effects found here.

Discussion

Two major results of Experiments 3 and 4 require explanation. The first is that subject intrude in recall and falsely recognize actions that are in the underlying scripts but were not stated in the texts. The second result is that these intrusions and false alarms increase if the subjects read more than one text that instances the same script. We will now contrast two models of memory for script-based texts and note whether they can account for these results. The models describe the way a script-based text interacts with a generic script in memory to create an episodic memory trace of the text. We will first discuss the representation of generic scripts and of episodic instances of them.

Figure 1 schematizes the knowledge structure for the "Visit Health Professional" script. At the top level the script has a Name (or Header), a list of Roles, Props, Entry Conditions, and Actions. The Roles would be Patient Receptionist, Professional Person, and Professional's Assistant. The Props would be an Office, Waiting Room, Magazines, Examination Room, etc. The Entry Conditions would be state descriptions such as "Patient feels bad." The Actions would be verb-based conceptualizations like "Patient Arrives at Office," "Patient Checks In With Receptionist," etc., or some representation even more abstract than these conceptualizations, such as the primitive actions suggested by Schank and Abelson (1977). The entries for the Roles, Props, Entry Conditions, and Actions in Fig. 1 are enclosed in angle brackets to indicate that they are lists of constraints on whatever fills the role rather than actual instantiations of the variables.

```
Generic Visit Health Professional Script in the Knowledge Store

Name = Visit Health Professional

          R1 =< Health Professional >
          R2 =< Receptionist >
Roles =   R3 =< Health Professional's Assistant >
          R4 =< Health Professional >

          P1 = <Office >
          P2 = <Waiting Room >
Props =   P3 = <Chair >
          P4 = <Magazines >
                  .
                  .
                  .

Entry =   C1 = <Patient feel bad >
          C2 = <Patient have Money >

          A1 = <Patient decide to go see Health Professional >
          A2 = <Patient arrive at office >
Actions = A3 = <Patient check in with Receptionist >
                  .
                  .
                  .
          A23 = <Patient leave office >
```

FIG. 1. Generic visit health professional script in the knowledge store.

As a person reads a script-based text that fits this Health Professional script, we assume that she sets up a new episodic memory structure to encode that particular story. We will contrast two different models of what the episodic memory contains, the Full Copy and the Partial Copy models. To illustrate the contrast, suppose the person reads a script-based text that begins as follows:

> John was feeling ill. He arrived at Doctor Smith's office. As he waited, he read a *Family Health* magazine. Then he went into an examination room where a nurse checked his blood pressure and weight . . .

In the Full Copy model, the reading of this text allegedly causes the person to set up an episodic memory structure which is a fully instantiated version of the complete script up to and including the point of the last script statement in the text. Thus, from the sample text, the script-applying mechanism knows how several "variables" of the abstract script in Fig. 1 are to be "bound" or specified. Thus, the Patient (R1) is set equal to John; the Health Professional (R4) is set equal to "Doctor Smith;" and the Assistant (R3) is set equal to "Nurse." The Receptionist (R2) is not specified but will be filled with the prototypical value (e.g., a young woman in a white uniform). The actions of the script are then filled in as fully as possible according to these bound variables or according to their prototypical values. If a script action is explicitly mentioned in the text, then that action instantiates an action variable. If the action is not mentioned, then an appropriate default value or prototypical action congruent with explicit bindings is entered. This process is shown in part by Fig. 2 where we have noted the text sentences plus a few surrounding script actions. (This figure omits the role and prop variables for simplicity). To the right of each statement is a tag or marker, which is ** (two asterisks) if that script-line was in the text and which is * (one asterisk) if it was not in the text but was derived by filling in the corresponding generic action with the prototype congruent with the bound role and prop variables of this particular text. The asterisk count is the sole discriminative trace regarding which script lines were stated in the text vs which lines were derived.

A nice feature of the Full Copy model is that it can readily answer questions about script-based inferences from the stated text. If we suppose that the occurrence markers fade with time, then the person will eventually be unable to discriminate whether a script-based statement was or was not stated in the text. This explains the false positive recognition results that we found.

What about the influence on the "John at Doctor" memory of other texts such as "Bill visiting the Dentist?" According to the Full Copy theory, the "Bill visiting Dentist" text would cause another instantiation

```
Name = Visit Doctor

Prototype = Visit Health Professional

Entry =   C1 = John feeling bad                          **
          C2 = John has money                            *

          A1 =  John decided to visit Dr. Smith          *
          A2 =  John arrived at Dr. Smith's office       **
          A3 = John checked in with Receptionist
          .                .
          .                .
          .                .
          A7 =  John read "Family Health"                **
Actions   A8 =  John went to Dr.'s exam room             *
          .                .
          .                .
          .                .
          A12 =  Nurse checked blook pressure and
                        weight of John                   **
          A13 =  Dr. Smith came into exam room           *
                          with John
          .                .
          .                .
          .                .
```

FIG. 2. Full copy model episodic memory.

of the Health Professional script to be fully copied into episodic memory, comparable to the episodic structure in Fig. 2. The new structure would have different variable bindings (e.g., "R4 = Dentist" rather than "Doctor Smith"). Unfortunately, within the Full Copy model, we have found no natural or parsimonious mechanism for creating confusions between memories of two script instances: they exist in distinct episodic compartments and do not interact. Consequently, the Full Copy model has no way to explain our second result, that other text instances of a script increase memory confusions about a given text.

A more adequate model is the Partial Copy model schematized in Fig. 3. In this case, the reading of a script-based text sets up an episodic memory that leaves traces at two locations. First, the specific propositions of the texts are recorded into episodic memory structures, shown as boxes to the right in Fig. 3. The unstated script actions are not instantiated at this time. Second, the abstract action corresponding to a text statement is marked in the knowledge store (to the left in Fig. 3) as having been accessed and used during understanding of a text. This activation is indicated in Fig. 3 by two asterisks attached to the memory node representing the abstract action. On the other hand, unstated actions of the script will be weakly activated (one asterisk) because they are part of the overall Health Professional script which has been aroused and is activating all its subordinate actions. This "top-down" activation from the script name (or header) to its subordinate actions corresponds to unconscious expectations regarding what conceptualizations the language understander is

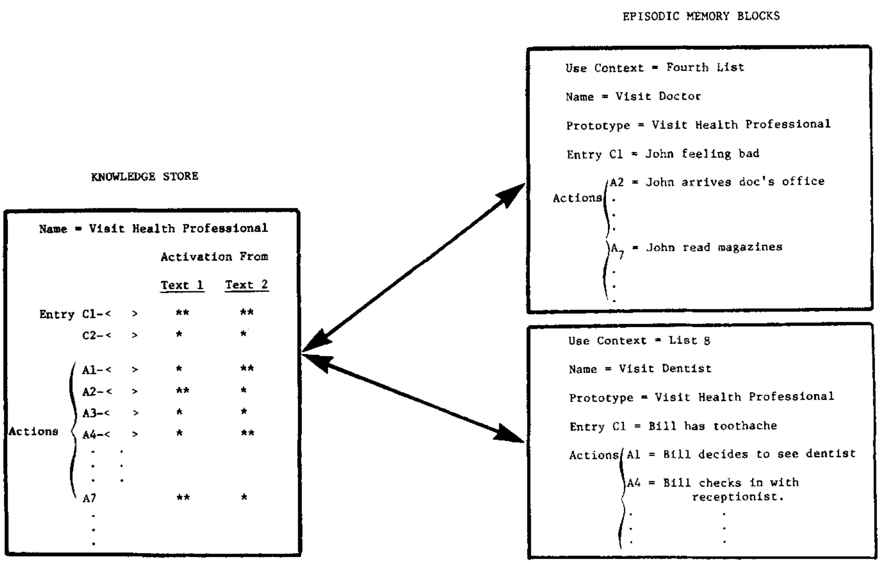

FIG. 3. Partial copy model.

likely to encounter as the script-based text is read. In any event, stated actions are distinguished in memory from unstated script actions because (1) a copy of the stated action is in the episodic trace, and (2) the level of activation is higher on the abstract action-node corresponding to a stated action.

In such a theory, the person would be able to answer questions about the plausible truth of unstated but implied events by deriving the answer at the time of the question using the generic script and instantiating it according to the variable bindings in the episodic memory for the text. Such derived answers would take longer than directly stored answers, a result found by Reder (Note 1) and Kintsch (1974) among others. However, this difference in reaction time could be predicted by the Full Copy model if actions filled in by implication are at a weaker strength than stated actions. Therefore, the major distinguishing feature of the two models is that the Full Copy model has no simple way to account for some interfering "cross-talk" between different instantiations of the same abstract script.

In the Partial Copy model, forgetting occurs in two ways. First, the propositions represented in the episodic trace are simply "erased" or "fade away," the loss of any one propositional trace occurring in an all-or-none manner with a fixed small probability in each unit of time. This process leads to exponential decay (or spottiness) over time of the propositional traces in the text's memory block. Second, and independently of loss of the episodic propositions, the activation on the superordinate action-node in the abstract script also decays over time. The effect of the first decay process is that over time the subject will forget exactly what was stated in the script-based text, though he may be able to derive approximately what was stated from the difference in activation levels on the script-action nodes. The effect of both decay processes is to cause the person eventually to forget everything about the text he read.

In the Partial Copy theory, recognition memory is a three-stage process which is diagrammed in Fig. 4. First, the memory scanner searches the appropriate episodic memory block (one of the episodic boxes in Fig. 3), where it checks the test sentence for a match to the memories of the stated propositions remaining in this episodic memory block (which may be spotty from decay). If a match occurs here, then a fast, confident recognition judgment ("Yes, Old") ensues.[2] If no match occurs here, the scanner

[2] A possible earlier stage not elaborated in Fig. 4 would have the text probe first analyzed for its theme or script; and if the script of the probe is not among the active script headers, then the probe might be rejected outright. The logical complications with such a process are several: first, not all probes will activate a determinate script, so a branching decision tree must be articulated farther; second, distinctive distractions (e.g., witnessing a murder while eating) may be highly memorable parts of a text yet not be related at all to any active script from the experimental texts. Since we have no data to inform the complications created by such a "script-checking" stage, we have not inserted it in the decision process of Fig. 3.

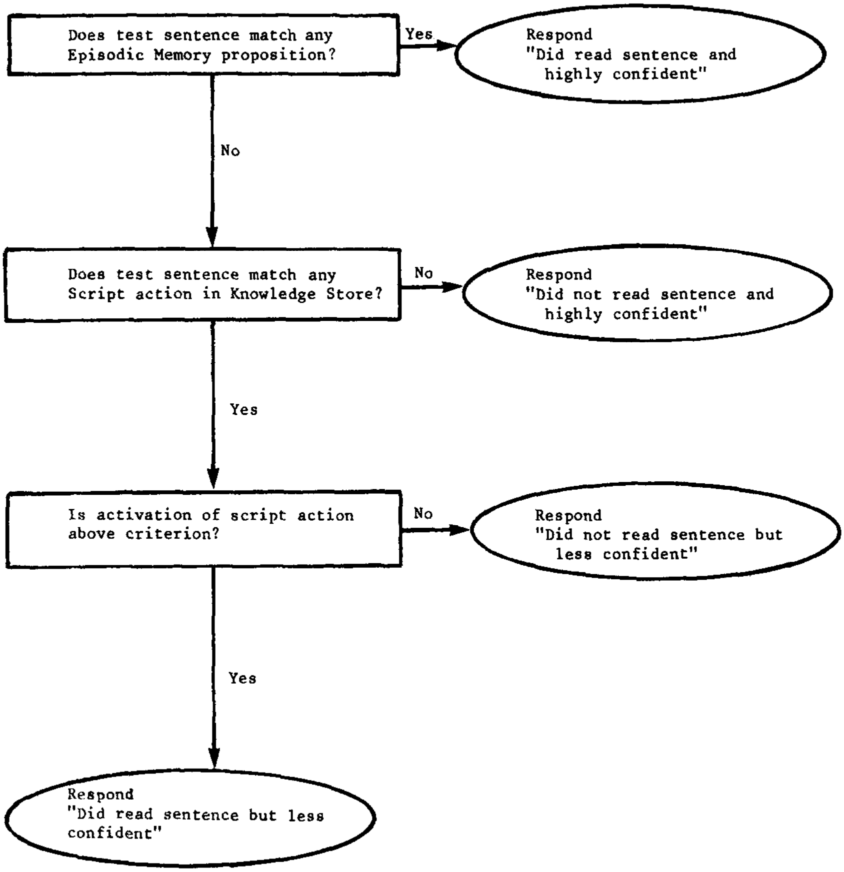

FIG. 4. Partial copy model's recognition process.

then goes up to the superordinate script-action corresponding to the test statement. If no superordinate script action corresponds to the probe, then the decision maker issues a confident "Did not read before" judgment. If a superordinate match does occur, then the scanner checks the level of activation on this node (the total number of asterisks on that line in Fig. 3). If the activation here is high, that means that an instantiation of this abstract expression recently occurred in some context. Therefore, it is plausible that the text probe in fact corresponds to an action stated earlier but now forgotten. On these grounds, the decision process is likely to "accept" or "recognize" the test statement. On the other hand, if the activation on the superordinate node is low, then the test sentence should be rejected, judged to have been "not stated" in the text. The judgments given by the third stage are less confident than the judgments from the first two stages. This decision model resembles that of Atkinson and Juola

(1973) except ours has more stages and its episodic and semantic memories are assumed to be searched in different orders. Hopefully there will be simplifying special cases of this complex decision model.

Let us now examine how this Partial Copy model deals with our results on memory for several instantiations of the same abstract script. Figure 3 shows an episodic memory block for the "Bill visiting Dentist" text as well as the block for the "John visiting Doctor" text. The stated propositions of two texts are recorded separately with the Prototype variable referring each instance back to the general script in the knowledge store. Note, however, each action in the Dentist text causes activation (asterisks) to accumulate at the superordinate script level in Fig. 3. We suppose this activation on an action-node summates regardless of its source. This accumulation of activation occurs not only for abstract actions stated in the second instance text but also, by spreading activation, for actions not stated in either instance of the text. (These correspond to the several types of script inferences distinguished in the Results section). It is this accumulated activation on the superordinate action-node which, according to this theory, causes the increase in false positive responses to unstated actions as the number of related texts increases. An unstated test sentence will fail to match in the topical episodic store, will then be referred to the corresponding superordinate action, and the activation on that node is likely to be above the "acceptance" criterion if one of the other texts had recently instantiated this action. At this point, the decision maker knows only the level, not the source, of activation on the generic action node, so false positive recognitions may occur. These later-stage responses would be expected to come out more slowly and with less confidence than early-stage responses. This model predicts our result that unstated inferences receive lower recognition ratings than do tests on stated propositions.

The Partial Copy model seems to provide a reasonable account of the salient features of our results. It also makes further predictions about verification latencies that seem consistent with existing data. We will, therefore, tentatively maintain the Partial Copy model. It is regrettably complex. But the data to be accounted for seem to warrant some complexities.

EXPERIMENT 5: SCRIPT ACTION REORDERING

Some scripts have strongly ordered actions, whereas others have none. The scripts used in the previous experiments (the Restaurant, the Doctor, etc.) are temporally ordered and most actions either cause or set up the preconditions for ("enable") the next action to occur.

The memory script that represents ordered actions must record this order in some manner. If this structure is used to record any text mentioning these actions anew, the process would be biased towards storing this listing in the canonical order of the script. But suppose the person must remember an arbitrarily reordered version of the Restaurant or Doctor scripts. That should prove difficult because the canonical order in memory should serve as a source of negative transfer and proactive interference against learning and recall of the reordered version. Further, the interference during recall should be in a specific direction. In trying to remember, the person should rearrange the misordered actions of the arbitrary text so that they are reproduced closer to their canonical order in the underlying script.

This ordering prediction for scripts is a special case of the general idea that stock narrative schema have ordered constituents. Thus the general story schema may have an ordering like Setting, Initiating Event Establishing Goal, Plan, Attempts, Consequence, and Resolution. Experiments have suggested that after reading misordered stories, readers are likely to remember them in their natural schematic order (see Kintsch, Mandel, & Kazminsky, 1977; Mandler, 1978; Stein & Glenn, 1978).

For comparison to learning of shuffled ordered scripts, we had subjects also learning arbitrary orders for what are naturally unordered scripts. An "unordered script" is a known set of actions or events for which there is no canonical order (in a particular subculture)—or if there is, the subject does not know it. For example, for most of us the order of events passing in review in a Veteran's Day parade is haphazard; all such parades have bands, children on bikes, dogs, policemen on horses, ladies auxiliary marchers, and so on, but to our knowledge there is no conventional order. The same is true for the scheduling of circus performers in the main ring at a circus. Since these events have no canonical order, the learning of any arbitrary order for them should be equally difficult, and there is no predictable direction for the misordering errors. Therefore, comparing memory for an arbitrary sequence of an unordered script vs an ordered script, the best recall for order (or serial position) should be for actions from ordered scripts whose position in the arbitrary sequence correspond to their position in the canonical script. Intermediate serial-position recall should occur for any actions from unordered scripts. Least accurate serial-position recall should occur for actions from ordered scripts whose position in the arbitrary sequence differs from their position in the canonical script.

A first experimental attempt to create this effect produced negative results. We scrambled all 12 actions in ordered scripts, had subjects study them, and then reconstruct their serial order from memory. Their reconstructions were almost random, with none of the predicted drift toward the canonical order. The subjects told us what was wrong with this procedure, namely, they remembered that the script actions were "all mixed up" in the to-be-learned orders, so if they forgot the presented random order they still tried to reconstruct an order which preserved the "all mixed up" property.

We avoided this task strategy in the experiment reported below by presenting the scripts predominantly in their canonical order but with a few events rather far out of order. The hope was that if subjects forgot the order, they could still use the fact that the listed actions were predominantly in their canonical order to guide their reconstruction. This reconstructive guide would help rather than hinder the directional errors of interest.

Method

Subjects. The subjects were 12 Stanford undergraduates participating to fulfill a service requirement for their Introductory Psychology class. They were tested in groups of four, three, three, one, and one, sitting around a large table. The experiment took about 45 min.

Materials. We composed five ordered scripts and five unordered scripts, each 12 sentences long. The ordered scripts were Birthday Party, Shopping for a Coat, Playing Football, Going to a Movie, and Attending a Lecture. The unordered scripts were Cleaning House, Shopping at the Supermarket, Veteran's Day Parade, Main Ring at the Circus, and Mechanic's Check-up on Car. The Cleaning House and Birthday Party scripts were used alternately as primacy and recency buffers in the memory task, and data from them were discarded by plan. Each ordered script was paired with an unordered mate and the two scripts were yoked together in the sense of appearing close together in all input schedules. Also, an out-of-order action in the ordered script was compared in reconstruction to its yoked action presented in the same position in the unordered script.

The subjects read the 12 ordered actions in each of the 10 lists, then later tried to reconstruct all their orders. Setting aside the two lists assigned as primacy and recency buffers, the input position of the eight middle lists was counterbalanced over subjects by using four different input sequences. Thus any given list (e.g., circus) appeared in Positions 2 or 3, 4 or 5, 6 or 7, and 8 or 9 for three subjects at each position.

The position of the 12 actions within the ordered scripts was assigned by the following procedure. From the canonical order four actions were selected at random for each script to be interchanged in their positions (the other eight remained fixed in their normal positions). A Latin Square was used to produce four orders of these selected four actions: one was the normal order and the other three became unique misorderings. Each of the three misordered versions was used with four different subjects. The summed number of steps required to move all of the four interchanged actions within a story back to their canonical positions averaged 13.5. The to-be-learned order of the unordered scripts was determined by random numbers; only one order was learned by all subjects.

The 12 action statements were typed on 12 3 × 5 in. cards and a title card was placed at the front of the deck. The experimenter arranged the cards in each deck in the order to be learned by the subject using that deck.

Procedure. Subjects were seated and instructed that they were to study 10 ordered lists of 12 actions per list, each typed on a deck of cards with a title at the front. They were told to remember the specific order in which the cards were arranged, so that they would be able to reconstruct that order exactly when they were given the scrambled deck of cards during later testing.

Ten decks of cards were then placed before each subject in the sequence he was to study them. They were in four colors to help him keep his place as he moved from one deck to another. The subject looked at the title card at the front of a given deck for 5 sec, then turned over the cards within that deck, one at a time, studying each statement for 5 sec paced by the experimenter tapping on the table using a stopwatch. Memory for the order of the statements was emphasized.

After all 10 lists had been studied, the experimenter thoroughly shuffled each deck of 12 action statements while instructing the subject on the serial reconstruction task. The subject was to take the shuffled cards of a deck, lay the 13 cards (12 plus a title) on the table before her, rearrange the cards into the serial order she believed she had studied (guessing when necessary), and finally record on a protocol sheet her reconstructed order using random numbers on the back of the cards for identifying them. The 10 lists were reconstructed in the same sequence as they had been studied. Testing took about 30 min.

In the statistical analysis, subjects and the four scripts within each condition served as random effects, whereas the type of script (ordered or unordered) and the sequence of presentation were fixed effects. We scored each sentence on the basis of its absolute position in the reconstructed order, comparing this to its position in the studied list and, if possible, to its canonical position in the underlying script.

Results

Comparing the probabilities that items are remembered in the same absolute position as their presentation, the predicted pattern emerges. Within ordered scripts, actions presented in their canonical location were remembered 50% at exactly that location, actions presented away from their canonical location were remembered only 18% at the presented location. For comparison, actions of the unordered (control) scripts were remembered 30% of the time at their presented position. For statistical analysis, the unordered script actions were divided into one set of four yoked to the four misordered actions in the ordered scripts and the remaining set of eight. We tested the interaction of ordered vs unordered scripts with correctly located vs mislocated actions. As the percentages above indicate, this interaction was highly significant [$F'(1,7) = 19.3, p < .005$]. That is, in ordered scripts, misordered actions were recalled less accurately than their yoked actions (in unordered scripts), but ordered actions were recalled more accurately than their yoked actions.

Next, we examined whether misordered actions of ordered scripts drifted back towards their canonical position during reconstruction. Each reconstructured item was given a score indicating the number of steps (serial positions) it had moved from its presented position in the direction toward its canonical location. (Movements in the opposite direction received negative signs). The summed displacements of the four misordered items towards their canonical positions averaged 6.44 steps (out of 13.5 possible), whereas the comparable sum for the unordered control items was 0.92 steps. These differ reliably [$F'(1,5) = 8.74, p < .05$]. Thus, the misordered items move about 48% of the distance towards their normal location in the underlying script.

We may conclude that the canonical order of a memory script helps people learn the order of any new textual instantiation which preserves the canonical order but hinders them if the new instantiation has misorderings. The script order in memory acts as a source of proactive interference or as a source for guessing when the presented order has been forgotten.

Texts that relate events out of their natural order, as we have done here, are not uncommon in literature. A storyteller is not constrained to preserve natural order in his narrative rendering. Some styles of storytelling, such as sports reporting (football, baseball), even have a standard misordering; one first tells the final score of the game, then recounts the significant events (e.g., big plays, penalties) in order of their importance, then perhaps recites the temporal succession of scoring plays. Thus, in telling some stories, we sometimes describe events in a manner other than reciting their temporal order, much as was done in the present study.

EXPERIMENT 6: SCRIPT EXPECTATIONS AND COMPREHENSION TIME

We have assumed that when an abstract script is aroused, it activates the memory nodes representing the script actions. Each such memory node is similar to a logogen (see Morton, 1969) which accepts and accumulates activation ("evidence") from prior context and from present sensory input. Activation of the overall script brings the activation level of script actions close to the firing threshold. Hence, relatively little sensory evidence directed to an action node is required in order for it to be perceived. Also, expected stimulus patterns should be identified rapidly because their logogens have been brought near firing threshold by the context alone.

The script-to-subactions activation just mentioned is a mechanism for creating generalized expectations; that is, the person expects other script actions to occur. A more refined proposal hypothesizes more specific activation local to an action node that has just been instantiated (i.e., read in the text). Specifically, this hypothesizes that the local activation caused by reading a given script action spreads to neighboring script actions in a forward direction, causing them in particular to be expected. We will call this the Local Spread hypothesis.

One method for testing this hypothesis compares the time a person takes to comprehend a sentence that is expected to one that is not. The method relies on the assumption that a sentence is perceived and understood more readily if it is expected (e.g., Haviland & Clark, 1974). The Local Spread hypothesis implies that a script line in a text should be read and comprehended more quickly if it is preceded in the text by a statement referring to an action that just precedes it in the underlying script. The text statement just preceding the target statement will be called the "priming statement" or "prime" for the target. The Local Spread hypothesis predicts that target comprehension time should be shorter the closer in the underlying script is the preceding action given as a prime and the greater the number of preceding primes.

This hypothesis and derivation was first thought of by Abelson and Reder, and they performed experiments (unpublished; Note 1) to investigate it. One experiment was inconclusive and the other gave mainly negative evidence against the Local Spread hypothesis.

We were unaware of these experiments when we planned and conducted the experiment below, looking for a distance effect. Our procedure differed from that of Abelson and Reder, and was somewhat more successful in getting the subject "farther into the scripts" before the target probe occurred.

Method

Materials. The script versions used were 18 from Experiments 3 and 4, which were three instances of six different abstract scripts, each text being eight statements long. Within each text the fourth and the seventh statements served as critical test sites. These were edited to be 12 words long. The construction of the texts and the test sites relative to the underlying script are illustrated in Fig. 5. Every three subjects were tested with a different selection of statements (version) from each script, comprising the columns of Fig. 5. The plus marks in the columns indicate the statements presented, always in the first-to-last script order. The symbols A_1 through A_{12} represent events of a hypothetical underlying script. The selected test sites in this instance are the sixth and eleventh elements of the underlying scripts. For different text versions the critical test sites are immediately preceded by a statement which, in the alleged underlying script, was either one, two, or three steps back ("distance"). Within a list, the priming distance at the second test site was different than that at the first site. Across the 18 lists for a given subject, the first and second sites were preceded six times each by priming sentences one, two or three steps back. The last test site was never the last line of the text or the script.

After composing these texts we used the chunk (scene) norms of Experiment 2, to score whether the priming and target actions were within the same chunk, or were one or two chunks apart. The average chunk distances for Steps 1, 2, and 3 were 0.50, 0.50, and 1.08,

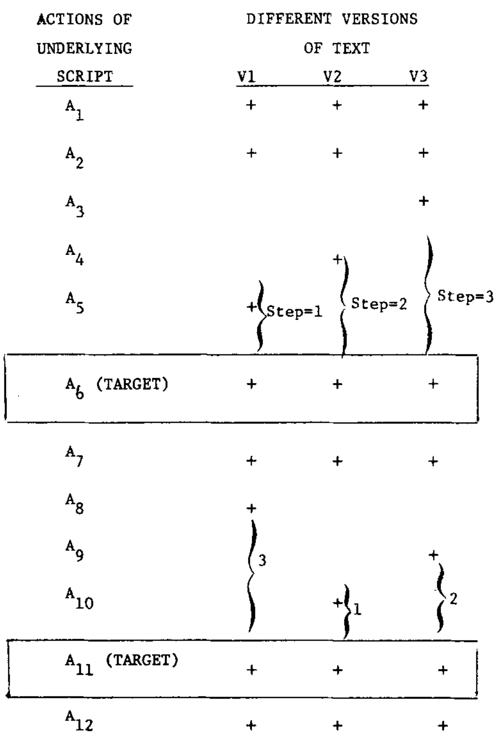

FIG. 5. Schema for materials used in experiment 5.

respectively. Thus, chunk distance increases monotonically along with number of actions between prime and target statements.

Procedure. The texts were presented one line at a time on a 9.5 × 7 in. cathode ray tube (Hazeltine Model 1) connected to a Nova 820 laboratory computer. The subject pressed a button as soon as she read and understood each sentence. The button press removed the sentence and 1 sec later the next sentence was displayed. Comprehension time was defined as the time from the display of a sentence to the button press removing it. Subjects were unaware that their reading time per sentence was being recorded. To provide a reason for reading the sentences, subjects were told we were interested in how people make up titles to characterize stories. Our subjects thus made up an interesting title for each text as it was read. They wrote down their title when a message signaled the end of the text. This procedure enforced a delay of 15 to 30 sec between scripts. After writing their title, to start the next text, subjects pressed the button which displayed a warning asterisk and 1 sec later the first sentence appeared. Subjects read the texts combined with the step sizes of the primes in one of 18 different orders over the experiment, with three subjects per order.

Subjects. The subjects were 54 Stanford University undergraduates, half fulfilling an Introductory Psychology requirement and half recruited for pay. They were run in groups of one to three at separate computer terminals. Each subject sat at a table with her CRT screen about 18 in. away.

Results

The primary result is that a Step-1 prime produced faster comprehension of the target sentence than did Step-2 and Step-3 primes. The reaction times in msec are 2600 for Step-1, 2730 for Step-2, and 2681 for Step-3. The difference in the means is statistically significant [$F(2, 106) = 3.43, p < .05$]. (Materials were nested within subjects by design so that F is the correct test here even with materials as a random effect). There is a significant difference between Step-1 and Step-2 [$F(1,53) = 5.93, p < .05$]; but not between Step-3 and the other two. Inspection shows the distance effect to be small and nonmonotonic. We will return to discussing this later.

The other significant effect in the experiment is that comprehension time is slower for the later (seventh line) location (2818 msec) than for the early (fourth line) location (2424). This is highly significant [$F(2,53) = 51.39, p < .001$]; however, the location effect does not interact with the priming distance effect reported above. Although the target sentences in these two locations, had been equated for number of words, we noted that they unfortunately differed in the number of syllables per sentence. In fact, when the reading time per syllable is calculated, the location effect reverses, with the first location having a comprehension time of 170 msec per syllable and the second location 152 msec per syllable [$F(1,53) = 16.71, p < .001$]. If syllables rather than words are the proper units for reading, this faster comprehension of the second location provides some support for the accumulation of activation across script lines.

The results show facilitation of reading at only the immediately adjacent action, without a graded effect of distance in the underlying script. While there is doubtless general activation of the whole script as individual events are read, the extra "local spread" of activation from the just-read action appears extremely local. The absence of a graded effect cannot be explained by appeal to a "chunk-distance" measure since that increases monotonically with our measure of number of intervening actions.

Discussion

Our failure to find a graded distance effect can be interpreted in several ways. One view suggests that the failure can be laid to uncontrolled variations in literary style or conversational fittingness of the texts with increasing prime-to-target distances. Some gaps are simply too large to bridge meaningfully; others convey the wrong impression of what is going on; others sound stilted or silly. Although we tried to avoid such stylistic blemishes in our materials, we cannot guarantee that style effects were inconsequential relative to the distance effects of interest.

An alternative interpretation is that there is no graded distance effect. After hearing an action, the reader expects its immediate successor more than others, but his within-script expectations are not further differentiated than that. Apparently, the latest version of the SAM program for applying scripts to texts (see Cullingford, Note 2; Schank & Yale A. I. Project, Note 4) has a predictive mechanism that operates in just this way.

A third view of the results is to suppose that a script is such a complex web of interconnected parts that the representation of temporal distance between actions bears little relationship to the node distance between them in the network. For example, scripts often have some causal or contingent connections which bridge large temporal gaps between early events and much later events. Thus, the size of a tip the customer leaves depends upon the quality of service he received much earlier in the restaurant script; what specifically the dentist does to you depends on your specific reason for visiting him. Further, two temporally distant actions may be close in node distance because they involve the same props and actors, whereas temporally close events may be farther apart in node space in that they refer to different roles, props, and within-script locations. Taking this view of scripts, then, it is misguided to expect local spread of activation to be indexed simply by temporal distance between two events.

EXPERIMENT 7: REMEMBERING DEVIATIONS FROM SCRIPTS

Although we are investigating script-based texts, we should remember that a script does not make an interesting story; in fact, it is miserably dull, containing all the predictable details of an activity. Script recital violates a conversational postulate that enjoins speakers and writers to be informative and not overly redundant. People read such boring texts only to assuage some psychology experimenter.

While actions within a script may be referred to in real stories, they serve only as a background or context within which something more intesting happens in the foreground. Thus, we may recite a restaurant script to note that while waiting for our food, a man at the next table had an epileptic seizure; or we note that when the soup arrived it had a fly in it, etc. Let us call these "interruptions" in the predictable flow of the normal script. The next experiment asks how well people remember such interruptions in comparison to the script actions.

Schank and Abelson (1977) noted several types of script interruptions called *obstacles, errors,* and *distractions.* In obstacles, some enabling condition for an imminent action is missing (e.g., You can't read the French menu), so some corrective action is taken (e.g., ask the waiter to translate for you). In errors, a script action leads to an unexpected or inappropriate outcome. For example, you order a hamburger, but the waitress brings a hot dog. The standard corrective action is to repeat the action: order the hamburger again. Distractions are unexpected events or states which set up new goals for the actor, taking him temporarily or permanently outside the script. For example, the waiter may spill soup on the customer, initiating a visit to the restroom for cleaning up.

Our intuition is that such interruptions will be remembered better than the routine script actions because they will appear subjectively more important and so will occasion more attention or deeper processing. The interruption is an unpredicted event, and so a script-based "von Restorf" effect (i.e., better memory for the surprising event) could be predicted. Further, from the viewpoint of the reader, the interruption seems to be the only "point" of the story.

Besides interruptions other extraneous statements can occur in script-based texts. These include irrelevant statements about attributes of the props or characters of the script, or the thoughts and feelings of the character which have no essential place in the causal flow of events. An irrelevancy would seem to be something that can occur in parallel with essential actions without impeding the flow of events. Thus, while looking at the menu the central character may notice the print-type, or notice the waitress' red hair, etc. Since such irrelevancies neither aid nor block the goal-directed actions of a script, we would normally expect them to be remembered very poorly. Now, there surely are some irrelevancies that refer to events or properties which, if they were experienced, would have a vivid impact, and would be well remembered (e.g., "The waitress was stark naked."). The experiment below side-steps this issue by having the irrelevancies as well as the script interruptions be relatively pallid and routine.

To summarize, then, we predict that interruptions will be remembered better than script actions and that irrelevancies will be remembered less than either.

Method

Materials. Six script-based stories were written about making coffee, attending a lecture, getting up in the morning, attending a movie, visiting a doctor, and dining at a restaurant. The texts were 22 to 26 sentences long. Setting aside the first few and last few sentences, the remainder of each text was divided into three groups of five to seven actions (avoiding divisions within a chunk). In each group we inserted one irrelevant remark and one interruption. Across the three groups within a text, there was one obstacle, one error, and one distraction. Across scripts, the six possible orders of the three interruption-types occurred once each. Thus, each text contained one each of an obstacle, error, and distraction, three irrelevancies, and about 20 relevant script sections. Each story was titled, and all six were stapled into a booklet.

Procedure. The subjects read the booklet of six stories for 5 min total, with only general comprehension instructions. They engaged in an intervening task for about 10 min as part of another study that involved rating the comprehensibility of unrelated sentences. After the rating task, subjects were asked to recall the stories in writing as close to verbatim as possible. They were cued with each story title and given as much time as needed (maximum of 5 min per story).

Subjects. The subjects were 24 Stanford undergraduates either fulfilling an Introductory Psychology requirement or receiving payment. They were run in groups of two to eight.

Results

The recall protocols were scored for presence of each text proposition (basically, clauses). The average percentages recalled were: interruptions 53%, script-actions 38%, and irrelevancies 32%. The interruptions are recalled reliably more than are the script-actions, with fixed effect [$F(1,21) = 55.26$] and mixed effect [$F'(1,12) = 21.30$, both p's $< .001$]. The script-actions are recalled somewhat better than the irrelevancies [with fixed $F(1,21) = 6.36, p < .05$ but mixed $F'(1,8) = 1.64, p > .10$]. The results are as predicted; interruptions were remembered best as though they were the point of the story, and irrelevancies were remembered poorest.

Calculating the percentages recalled of the three types of interruptions, obstacles were highest (60%), then the distractions (56%), with the errors recalled least (42%). Obstacles and distractions do not differ reliably but both are significantly higher than recall of the errors when materials are considered as a fixed effect [$F(1,21)$ for obstacles is $15.11, p < .001$, and for distractions is $5.86, p < .05$]. However, the differences are not significant when different scripts are considered as a random effect [obstacles $F'(1,10) = 4.83$, $p > .05$; and distractions $F'(1,13) = 2.69$, $p > .10$]. Consequently, we are uncertain whether the recall profile of types of interruptions will generalize across variations in learning materials. Although the interruptions as a whole are recalled better than the script actions as noted above, the recall of the errors does not differ reliably from the recall of the script actions [$F(1,21) = 1.56, p > .10$].

Discussion

Recall of interruption types was ordered in the way we expected, with obstacles and distractions being remembered better than errors. In our texts, the errors were "minor": simply incorrect outcomes of events. Obstacles, on the other hand, involved real blocks to the progress of the script; the obstacle stopped the flow of events and had to be dealt with. Distractions were interesting incidents that suspended the script's goal, and temporarily replaced it with a more pressing goal for the character.

We can speculate about how such interruptions would be recorded in the episodic memory trace of the script-based text. The main trace would be the appropriate "Partial Copy" of the underlying script (see Fig. 3) with the variables bound as specified by the text. An error is an outcome with an unexpected value inserted into a standard slot in the script. It is as though a prediction of an outcome (a "slot filler") has to be replaced by a different object or value. An obstacle, and the corrective actions it causes, would also be recorded in the episodic script memory, at the script location where it occurred to frustrate a subgoal. A distraction, on the other hand, could be recorded on a separate "goal chart" (see Schank & Abelson, 1977, Ch. 5), with its actions and outcome together with a pointer to the location in the episodic script where the distraction occurred. The timing of most distractions is not linked causally to any specific point in the script (e.g., the person at the next table could have a seizure at any time). Therefore, we might expect subjects to remember the location of distractions within a script-based text more poorly than they would the location of obstacles or errors. A pilot experiment found no tendency for memory to err by mislocating distractions nearer to (rather than farther from) a scene boundary.

GENERAL DISCUSSION

We may view our results from two perspectives. The first considers them as empirical investigations of a previously unexplored domain of semantic knowledge; the second assesses their articulation with the general script-theory of Schank and Abelson (1977).

From the empirical perspective, we have explored the properties of scripts considered as concepts about routine activities. Just as concepts like *birds* or *weapons* have culturally agreed upon attributes and instances, so do activities like *eating in a restaurant* or *visiting a doctor*. These activities have conventional roles, props, event-sequences, standard entering conditions, and standard outcomes. Not only did our subjects largely agree on what these are but also on how to segment the event sequences into constituent scenes or chunks. We found that in remembering script-based texts subjects confused what was said with what the script strongly implied. Further, subjects preferred to learn event sequences that preserved the scriptal order. In remembering a script-based text, subjects were best at recalling brief obstacles or distractions which blocked or temporarily suspended pursuit of the script goal, whereas properties or events irrelevant to that goal were least recalled. Such results serve as an opening into empirical explorations of the organization and use of script knowledge in text understanding and recall.

From the theoretical perspective, our results are generally consistent with current script theory (see Cullingford, Note 2; Schank & Abelson, 1977). Some results are not specifically addressed by script theory (e.g., how different instantiations of a script interfere with one another in memory, or how people remember texts with out-of-order actions). Further, our results do not address many problems of language processing which script theory was proposed to solve. The theoretical writings and computer-simulation programs (see Cullingford, Note 2) mainly concern the way scripts act as a predictive context for processing single sentences, for tying together sequences of sentences, and for answering questions about a text (see Lehnert, 1977). For example, for understanding single sentences, an active script will suggest the relevant meanings of ambiguous words, will help establish referents of terms, will merge referents, will fill-in unspecified roles, and will predict likely conceptualizations to follow. The psychological counterparts of such processes probably occur in real-time as sentences are comprehended, and variations in factors affecting them (e.g., the ease of establishing referents) could affect comprehension time. However, except for Experiment 6 which examined within-script gradients of expectancy, we have not investigated how scripts influence comprehension in real-time.

Our results also do not address the extensive theorizing concerning goals and plans that underlays the script concept (see Meehan, Note 3; Schank & Abelson, 1977). Goal issues arise quickly in answering *why* questions about human actors (e.g., "Why did he leave the restaurant before he'd eaten?"). It is not obvious how the methods of experimental psychology can provide much relevant information on such issues beyond that provided by intuition and common sense. In any event, we have not tried to do so here. Because we have not examined issues of real-time comprehension or of goal-based question answering, our results make contact with only a small part of script theory.

Unresolved Issues About Scripts

Although our results have advanced the case for scripts and we find many attractive features of script theory, we would be remiss not to balance the picture by pointing out a few of the unresolved issues or critical questions about script theory. These represent conceptual puzzles that are on the research agenda.

A first question requiring some answer concerns how to elicit script knowledge. In Experiment 1 we instructed people in detail how to tell us about the major events of a routine activity and we obtained reasonable results; but the results surely will vary with the way instructions are phrased. Rather than recall, our subjects could have been asked to identify or recognize script-relevant actions from a large pool presented to them. There is little guarantee that all methods will yield the same conclusions. A problem is that recall or report methods can give only script knowledge that is accessible to conscious introspection. But just as a fish fails to report that he is surrounded by water, so do people surely have much tacit, nonintrospectible knowledge about stereotyped procedures and activities that they do not or cannot report (see Goffman, 1959 for details of unconscious social conventions). For example, traditional Japanese rarely think to mention that one takes off his shoes before entering a Japanese restaurant, though that feature immediately strikes a Westerner as unusual. Similarly, our Restaurant script does not mention when it is proper to sit at a table with a stranger, or how close the waiter stands while taking your order. In reciting a script, people assume and do not report tacit conventions regarding physical layouts and interpersonal commerce, even though such conventions are clearly exhibited when they enact the script.

This latter split between enacting vs verbalizing a script raises the more general question of how we are to decide whether someone is behaving according to a script and, if so, which script it is. Clearly someone can go through the motions of a script enactment without having the knowledge that customarily underlays its performance. Thus, an uneducated person will appear to follow eating protocol at a formal state dinner by imitating those around him, or by following instructions from his dining companion. He is using immediately available cues or rules rather than memories of prior performance or prior rules as his guide. But the cues that control behavior are rarely conspicuous, so we will frequently err in inferring whether someone is following a memory script. Furthermore, we are prone to similar errors in deciding which of several scripts someone is following. Errors are introduced by the loose relation between intentions and actions. The source of errors can be illustrated by considering cases of deceit, con-games, bluffs, and sham put-ons. In competitive situations, which many stories describe, a character may present an appearance or "front" of following one script (designed to mislead his adversary), whereas he is actually following another plan which will give him an advantage. Thus, in football, the quarterback fakes a line plunge before passing, hoping that the defenders will react to the apparent intention and leave themselves open for the pass. Such common examples show the difficulty of identifying which script someone is following from their surface behavior.

A second question about scripts concerns the level of detail that is recorded with each script, and how much of this is called forth when the script is instantiated. For example, in the Restaurant script, what kind of general and specific information is stored about the Service Person? Some features are mandatory (must be a living person), some are optional (male or female), some have a range of permissible values with a prototypical value (e.g., age, with 25–45 years), and others have a range without a clear prototype (e.g., hair color). Is all this information stored with the script and brought forth when it is instantiated? Possibly not. Rather the script may only refer to an empty waiter or waitress role, which then points to a mental "dictionary" or lexicon which holds context-free information about these concepts and their prototypes. In instantiating the script, the features of the prototype will be "loosely bound" in the sense that the text can readily replace a guessed value.

A problem with separating the script from the lexicon is that we often want the two sources to interact and exchange expectations. For example, if at lunch we discover our Server is a (nude) topless waitress, we want inferences from that to propagate back to selection of a specific track (type) for the Restaurant script, namely, that we are in a Go-Go bar. Similarly, a fast-food restaurant and a haute cuisine French restaurant have different Server prototypes. It would seem therefore that each track of a script must have an associated file of prototypical values.

Schank and Abelson (1977) introduce the idea of a *track* to refer to a distinct subclass of script situations. For example, for the eating script one has a track for eating various types of meals and snacks at MacDonalds, at school, at a coffeehouse, a picnic, a hiking camp, a church benefit, an Algerian Casbah, on a train or airplane, and so on to the limits of his experience. Mention of each situation evokes memories which specify contextually appropriate prototypes for various roles, props, and events. But, one may ask, if episodic memories about track-experiences provide this information, of what value is the general Restaurant script? Presumably, the general script identifies the cluster of common or frequent features of the tracks it subsumes (e.g., there is an exchange of money for prepared food). By virtue of these clusters, to say that one ate in a restaurant is to set up many common expectations in the listener.

A third problem for script or frame theories is to account for how special or novel contexts propagate throughout the script to modify the appropriate details. For example, if we watch theater actors in a play pretend to eat in a restaurant, or if a child enacts with dolls what goes on in a restaurant (as did the preschoolers of Nelson, 1978), how does the context of that pretend world get passed along to modify this instantiation of the real-world script? We do not expect the pretend food to be hot, or the money to be real, or there to be a real kitchen off-stage, etc. The issue is to account theoretically for the way a context like "theater world" or "toy world" selectively cancels certain aspects of the script but not others. The issue is similar to how we know what properties to ascribe to a fake duck or a dollhouse.

A fourth problem for script theory (or any schema theory) is to decide at what level of abstraction the memory script is to be used and modified. For example, suppose you read a text about visiting a specific cardiologist. How are we to understand and record that text in memory? Is that an instance of the script called "visiting a cardiologist," or a doctor, or any health professional, or any professional, or any person, or the script called "going to place X and talking to person Y"? The concepts are all connected in a hierarchy, and properties true of an activity described at one level are also true of subordinate scripts in the subset tree. But when a text calls up a script from the reader's memory to begin instantiating its slots, at what level is that script?

A defensible answer is that successive clues from the text are sorted through a discrimination net to retrieve the most detailed script available in memory to encode the current text or situation. A problem with such a system is that according to schema theory the understander must commit himself to some initial schema in order to understand sentences; yet the most diagnostic information may not appear in the text until later. That is, one can be misled down "garden path" stories. When gross errors of predictions are encountered, the system must be able to discard the currently active schema, substitute another, and then try to retrieve earlier inputs and reinterpret them in terms of the newly suggested schema. Thus, what started out as an "eat-in restaurant" script may turn into an "attend political fund-raiser banquet" or a "deliver food to restaurant" script. Any strongly predictive ("top-down") understander system such as script theory must have ways to modify a current script, to reject and replace it when it becomes inappropriate.

An alternative view of the comprehension process is that the reader progressively builds up a model or image of the situation the text is about; he conjectures an initial ill-defined model from the initial sentences, then uses successive lines of the text either to fill in the empty slots of that model or to revise it dynamically (e.g., see Feldman, 1975; Collins, Brown, Larkin, 1978). In such a dynamic process, the important connections would not be specifiable in advance; the appropriate model for a text would be arrived at by successive revisions, by refinements according to constraints of the text, by applying problem solving methods rather than by selecting a preformed template from a storeroom of static scripts. This dynamic modeling approach would claim that each situation is somewhat unique, with novel combinations of features and happenings, and that it is unlikely that a limited file of scripts would "cover" or subsume many particular instances. The script-file could appear to cover many instances only by ignoring their special or deviant features, or by endlessly amending or specializing the general script to describe the unique cases (see the "weird list" of Schank & Abelson, 1977, p. 166). Perhaps it would have been more efficient to combine information from several different scripts to describe the cases directly. But then, why bother to have static scripts to do this rather than just a network of concepts?

A fifth problem for script theory is to specify an induction algorithm by which new scripts can be learned from experience. If new situations can only be understood as instances of preformed scripts (perhaps with some deviations), how then can any new script be learned? The theory now seems to encompass *specialization* as a way to learn, as when we learn the MacDonald's track as a special case of the general restaurant script. In turn, the restaurant script could be learned as a special instrument to enable the basic action of eating. Similarly, attending concerts, lectures, museums, and movies would be specialized instruments to the basic action of attending to (perceiving) something (see the semantically primitive acts enumerated by Schank, 1975, p. 40ff). This view suggests that many scripts will be clustered around the primitive action they enable. Thus, the restaurant, bar, and kitchen scripts cluster around ingestion; the bus, train, airplane, and bicycle scripts cluster around physically moving oneself, and so on. The elaboration of a given primitive action (say, eating) would seem to develop by specialization and by noting recurrent patterns of features (see, e.g., the concept learner of Hunt, Marin, & Stone, 1966). The result would be a tree or discrimination net in memory, with branches encoding different locations and styles of eating (the "tracks" of Schank & Abelson, 1977). The details of growing such a net have not been worked out.

Related to the issues of abstraction and learning, a sixth issue for script or frame theory is deciding where a new fact is to be stored and which scripts are to have access to it (see J. R. Anderson, 1976, p. 446). Suppose while at a lecture I learn that the use of saccharin sweeteners can be harmful to my health. I gain nothing by simply recording that fact in my "Lecture" script. I must record it in such manner to make it accessible to my "Bar" script to avoid drinking Thintonic, to my "Coffeehouse" script to avoid putting it in my coffee, to my "Birthday Party" script to avoid giving dietetic candies as presents to my friends, and so on. We would like the fact to be available diffusely across all relevant contexts, but that would seem difficult to achieve if it is buried in one particular script. Conceptual networks provide this sort of diffuse broadcasting of a new fact across multiple contexts. Perhaps script theory can gain some of this generality by altering facts in the lexicon (e.g., about drinks or foods that involve saccharin), so that all scripts using tokens of that lexical entry could be modified if that value is retrieved when the script is next executed. But it is not obvious how to do this in an efficient way.

A final problem for script and frame theories is that they currently have no clear way to deal with simultaneous execution of several scripts which have strong interactions. SAM and other script implementations (Cullingford, Note 2) seem to deal with one script and motive at a time, whereas people frequently act from multiple motives and within multiple constraints. To illustrate, suppose two businessmen who are chess enthusiasts are riding together on a train to a business convention. While eating in the dining car, they play a game of chess and also negotiate a business contract with one another. These men are engaging in at least five scripts more or less at the same time (or in rapid alternation): going to a convention, riding a train, eating dinner, playing chess, and negotiating a business contract. The activities have various embedding relations to one another; the train ride is the first part of the attending convention script, the dinner script is embedded within the train script, and the chess and negotiations proceed in parallel and embedded within the dinner script.

A language understander must be able to refer an incoming sentence to the appropriate script, and this could be done reasonably well by keeping a queue of foregrounded (active) scripts to which each input would be compared for a relevant match. It would not be difficult in current script programs to interrupt a script, save your location there, go do an embedded script for awhile, save your location there, return to the desired location in the first script and proceed from there for awhile, then return to the second script, and so on. What is harder to model are interactions between the goals and resource-allocations among several simultaneous scripts. Thus, the outcome of the business negotiations can influence who pays the bill for dinner, or the seller may make low-quality moves to lose in chess in hopes of influencing the business deal, or at the most delicate decision-point in the business negotiations the buyer now reveals a brilliant chess move to divert the seller's attention from the negotiations. These are cases in which a script action is performed for other than its usual reasons; they satisfy motives other than the standard ones, and along with bluffs and deceptions they comprise some of the role-taking complexities of real human commerce.

In closing this discussion, let us repeat that we consider scripts as a powerful and potentially valuable theoretical approach. The unresolved issues and theoretical puzzles raised here are not unique to script theory, but to any well-specified schema theory. We raise these issues to suggest the direction of future research.

REFERENCES

Ammon, P. R. The perception of grammatical relations in sentences. *Journal of Verbal Learning and Verbal Behavior,* 1968, 7, 869–875.

Anderson, J. R. *Language, memory, and thought.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1976.

Anderson, R. C. The notion of schemata and the educational enterprise: General discussion of the conference. In R. C. Anderson, R. T. Spiro, & W. E. Montague (Eds.) *Schooling and the acquisition of knowledge.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Atkinson, R. C., & Juola, J. R. Factors influencing speed and accuracy in word recognition.

In S. Kornblum (Ed.) *Attention and performance*. IV. New York: Academic Press, 1973.

Battig, W. R., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *Journal of Experimental Psychology*, 1969, **80**, Part *2*, 1–46.

Bower, G. H. Cognitive psychology: An introduction. In W. K. Estes (Ed.) *Handbook of learning and cognitive processes* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1975.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–339.

Clark, H. H., & Clark, E. V. *Psychology and language*. New York: Harcourt Brace, Jovanovich, 1977.

Clark, H. H., Cohen, J., Smith, J. E. K., & Keppel, G. Discussion of Wike and Church's comments. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 257–266.

Collins, A., Brown, J. S., & Larkin, K. M. Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Feldman, J. *Bad-mouthing frames*. In R. Schank & B. L. Nash-Webber (Eds.) *Theoretical issues in natural language processing*. Conference Proceedings: MIT, 1975.

Goffman, E. *The presentation of self in everyday life*. Garden City, NY: Doubleday, 1959.

Goodman, A. I. *A theory of human action*. Englewood Cliffs, NJ: Prentice Hall. 1970.

Grice, H. P. Logic and conversation. In P. Cole & J. L. Morgan (Eds.) *Syntax and semantics* (Vol. 3): *Speech acts*. New York: Seminar Press, 1975.

Haviland, S. E., & Clark, H. H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 512–521.

Hunt, E. B., Marin, J., & Stone, P. *Experiments in induction*. New York: Academic Press, 1966.

Jarvella, R. J. Effects of syntax on running memory span for connected discourse. *Psychonomic Science*, 1970, **19**, 235–236.

Jarvella, R. J. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 409–416.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.

Kintsch, W., Mandel, T. S., & Kazminsky, E. Summarizing scrambled stories. *Memory and Cognition*, 1977, **5**, 547–552.

Lehnert, W. Human and computational question answering. *Cognitive Science*, 1977, **1**, No. 1, 47–73.

Mandler, J. M. A code in the node. *Discourse Processes*, 1978, **1**, 14–35.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston, 1960.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.) *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Morton, J. The interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165–178.

Pew, R. W. Human perceptual-motor performance. In B. H. Kantowitz (Ed.) *Human information processing: Tutorials in performance and cognition*. Hillsdale, NJ: Lawrence Erlbaum Assoc., 1974.

Roediger, H. L., III. Inhibiting effects of recall. *Memory and Cognition*, 1974, **2**, 261–269.

Rosch, E., & Mervis, C. B. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, 1975, **7**, 573–605.

Rosch, E. R., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyer-Graem, P. Basic objects in natural categories. *Cognitive Psychology,* 1976, **8**, 382–439.

Rumelhart, D. E. Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.) *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge & J. Samuels (Eds.) *Basic processes in reading and comprehension,* 1977.

Schank, R. C. *Conceptual information processing.* Amsterdam: North-Holland, 1975.

Schank, R. C., & Abelson, R. P. *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.) *The psychology of learning and motivation* (Vol. 8). New York: Academic Press, 1974.

Stein, N. L., & Glenn, C. G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.) *Multidisciplinary perspectives in discourse comprehension.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1978.

Wike, E. L., & Church, J. D. Comments on Clark's "The language-as-fixed-effect fallacy." *Journal of Verbal Learning and Verbal Behavior,* 1976, **15**, 241–255.

Winston, P. H. *Artificial intelligence.* Reading MA: Addison-Wesley, 1977.

Zadah, L. Fuzzy sets. *Information and Control,* 1965, **8**, 338–353.

REFERENCE NOTES

1. Personal communications from Lynne Reder, March, 1977; and Bob Abelson, January, 1978.
2. Cullingford, R. E. *Script application: Computer understanding of newspaper stories.* Research Report No. 116, Department of Computer Science, Yale University, 1978.
3. Meehan, J. *The metanovel: Writing stories by computer.* Unpublished doctoral dissertation, Yale University, 1976.
4. Schank, R. C., & Yale A. I. Project. *SAM–A story understander.* Research Report No. 43, Department of Computer Science, Yale University, 1975.