JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 18, 565–583 (1979)

Effects of Analogy to Prior Knowledge on Memory for New Information

MIRIAM W. SCHUSTACK AND JOHN R. ANDERSON

Carnegie–Mellon University

Two experiments explored how memory for new information is affected by awareness of parallels to pre-experimental knowledge In one experiment, subjects studied brief biographies of fictional characters analogous to famous people Best recognition resulted with the analogous famous person identified at Study and Test; identification at Study or Test alone interfered relative to no identification A second experiment rejected accounts relying on simple matching between Study and Test contexts: A famous person's name was beneficial only when facts in the biography were true of that famous person Our data suggest that the benefit of prior knowledge derives from the more elaborate encodings that analogy promotes Implications for schema and depth-of-processing theories of memory are considered

Research on memory has a long (if intermittent) history of interest in the effects of pre-experimental knowledge on memory for presented information. Of the early research, the work of Sir Frederic Bartlett (1932) is probably the best-known experimental treatment of the ways in which memory for new information can be influenced by prior experience.

Recently, interest has focused on how our general knowledge about the world is organized and used, especially stereotypic knowledge about different classes of objects, people, and situations. This interest is manifested in a host of structural or organizational devices proposed within the past few years to model such knowledge. The best known of these come from the artificial intelligence literature: the "frame" theory of Minsky (1975), the "script" theory of Schank and Abelson (1977), and the "schema" theory of Rumelhart and Ortony (1977). These formulations (to be referred to generically as "schema theories") have in common the notion that knowledge about the world is organized into chunks of related information, with these chunks directing comprehension of what we experience in the present and prediction of what we will experience in the future. One feature of such representations is that a wide range of instances (differing in detail) are subsumed under a single, comprehensive schema. This schema contains the elements that are necessary or in common among instances, and allows variability in other elements; for example, in the schema for eating in a restaurant, the type of food eaten is not specified.

These knowledge structures are commonly believed to be important in the processing of prose passages, playing a role in comprehension, inference making, and memory. Many experiments have demonstrated effects attributed to the operation of schematic knowledge structures; a number of these experiments use the same general paradigm, and yield similar results (e.g., Bransford & Johnson, 1972; Dooling & Mullet, 1973; Smith, Adams, & Schorr, 1978. In this paradigm, all subjects are given a set of

This research was supported in part by NSF Grants BNS76-00959 and BNS78-17463 to John R Anderson and by an NSF Graduate Fellowship to Miriam W Schustack We would like to thank H. B Wall for assistance in generating stimuli and running subjects.
Correspondence concerning this paper should be sent to Miriam W. Schustack, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pa 15213.

0022-5371/79/050565-19502.00/0

sentences to be remembered. Each sentence, by itself, is moderately comprehensible (that is, the sentence is not anomalous, and the predicate is plausibly related to the subject). As a set, though, the sentences do not form a cohesive, comprehensible unit. Along with the set of sentences, some subjects (the "informed" group) get another piece of information in the form of an additional sentence, a picture, or a title for the set of sentences. This extra information allows the unrelated sentences to be understood as a coherent body of information about a familiar topic; the cue can be seen as the means of identifying the schema to be used in encoding the new information. All subjects are tested on their memory for the presented information, and those subjects who had the integrating cue are compared with those who did not. The usual finding is that subjects who had the integrating cue show better memory for the presented information, but also show greater difficulty in rejecting not-presented items consistent with the integrated interpretation. Examples of the first finding include Bransford and Johnson (1972) and Bower, Black, and Turner (1979). Examples of the second finding include Bower et al. (1979) and Thorndyke (1977).

The mechanisms underlying these effects on memory are still in doubt. The assumption that schematic knowledge structures are intimately involved in these results is not adequate to specify mechanisms; some still-unspecified processing assumptions must accompany the representational assumptions in order to predict the pattern of data obtained. Moreover, it has not yet been proved that invoking the notion of schematic knowledge structures is necessary; nothing precludes the possibility that subjects do not employ schemata at all, but instead use remembered individual instances of the cued event or object. The literature on prototype formation addresses the issue of whether generalized knowledge or memory for specific instances is responsible for "schematic" effects (e.g., Medin & Schaffer, 1971; Posner & Keele, 1970; Reed, 1972). That literature does, however, leave the issue unresolved.

The experiments reported here look at the effects of cues to prior knowledge about specific famous individuals. They were designed to shed some light on this issue of whether documented memory effects of cues to prior knowledge must be attributed to the use of generalized knowledge (schemata), or whether they could be due to the use of specific remembered instances. These experiments alone will not resolve the issue of whether schematic or single-instance knowledge is responsible for the effects found, but they can serve to place constraints on how that issue can be resolved, and to indicate directions for future research.

The existing literature provides little evidence on the effects of employing specific, nonstereotyped, nongeneral pieces of knowledge (for example, facts stored about a famous individual). An exception is some work of Dooling and his associates (e.g., Sulin & Dooling, 1974; Dooling & Christiaansen, 1977) involving the use of specific knowledge; these studies have repeatedly found differences only on foil items, and no effect on presented items. One way in which our experiments differ from that work is in the use of analogies; in our experiments, we provide information that allows the subjects to form analogies between prior knowledge and the information to be learned. In the earlier experiments, "informed" subjects were told that passages to be learned were about the famous figures themselves. The major difference, though, is that in the earlier work, "informed" subjects received both schematic and individual information; the biography title "Adolph Hitler" gives a cue to the military dictator schema as well as to Hitler himself. The "uninformed" subjects, though, have neither type of information (given the title "Gerald Martin"). In our experiments, all subjects receive a schematic framework for each biography (name, nationality, profession, and historical period). What differentiates our "informed" subjects is the identification of a particular famous individual who (in another time and place) performed analogous actions. All of our subjects know that the Napoleon-based character was a military dictator; they differ only in their knowledge of Napoleon as the model. While one could argue that schemata exist at various levels of specificity, we will assume that for our experiments, differences between informed and uninformed subjects are primarily, if not solely, due to the effects of knowledge about specific individuals.

Our experiments were also directed toward providing further information about the facilitating effects of prior knowledge on the acquisition of new knowledge. The studies were designed to yield more detailed information than the existing literature provides about the nature and the limits of the commonly found effects. The information about how prior knowledge is used in these experiments might also generalize to an understanding of the processes operative in situations where the prior knowledge involved seems more schematic than that used here.

In addition, this work was designed to provide some specific information about the potential of analogies to promote good memory. This would have implications for our understanding of analogies and metaphors (cf. Sternberg, 1977; Verbrugge & McCarrell, 1977), and for our understanding of memory representation and process in general. The research also has clear educational implications.

Experiment 1

In the first experiment, subjects studied 42 short biographies of fictional characters. Some of the characters had strong (but nonobvious) parallels to well-known individuals and some did not. Of the characters based on real-world models, half were studied with the appropriate model explicitly identified, and half without. Memory was tested by a recognition task in which each fictional character's name was paired with sentences from his or her biography, and with sentences from other biographies (as distractors). For each pairing, subjects judged whether the sentence had occurred in that character's biography. Half of the models identified at Study and half of the models not identified at Study were identified in the recognition task.

This orthogonal manipulation of the presensation of the model's name at Study and at Test enables determination of *when* the analogy has its effect: Does the analogy affect processes occurring at Study alone, or at Test alone, or does the effect depend on an interaction between processes occurring at both of these points? In previous experiments, the theme or topic was (at least implicitly) available at Test if it had been revealed at Study: For example, with a single story to read and remember, it is unlikely that subjects would forget the theme between Study and Test. In this experiment, however, the Study material constituted a large enough memory load (brief biographies of 42 fictional characters) that subjects were unlikely to remember the real-world model associated with any fictional character unassisted. Also, in this study the sentences themselves were stated in rather abstract terms that did not tend to evoke the real-world model.

Another issue that we wanted to address was *how* an analogy has its effect. One possibility is that subjects use a simple tagging procedure in which some previously known facts about the real-world model are tagged as being true of the new fictional character. Tagging models of memory have previously been considered both for single-word stimuli (Anderson & Bower, 1972, 1974; Yntema & Trask, 1963) and for complex script events (Bower et al., 1979). An alternative possibility is that an analogy is helpful because it allows subjects to use their knowledge about the famous model to better elaborate the facts to be learned (e.g., Anderson, 1976; Anderson

& Reder 1979; Reder, 1979) or in some way to process the material more deeply (e.g., Craik & Lockhart, 1972). In order to discriminate between tagging and elaboration, two types of facts were used in constructing the biographies of fictional characters based on real-world models. Most of the facts about each of these fictional characters were already known to the subjects as true of the famous models. In each biography, however, there was one additional fact that was not known to be true of the model (to be called the "invented" fact). If this previously unknown fact also benefits from subjects' knowing the analogy, a tagging explanation becomes less plausible, and an elaborative model is supported.

Method

Subjects

The subjects were 54 Yale University undergraduates who participated to fulfill a course requirement

Apparatus

The Study and Test materials were presented to each subject individually via VT-50 cathode ray tube (CRT) terminals (which display only uppercase alphabetic characters) controlled by a PDP 11/40 computer. All responses were recorded via the terminal keyboard.

Materials

The materials for the Study phase were a set of 42 short (four-sentence) stories, each a biography describing some events in the life of a fictional person, some of his or her personal characteristics, or both. The fictional person was identified by a unique name and by a summary description (of nationality, profession, and historical period), in the manner of a biographical dictionary entry (e.g., Francoise Jamel, Haitian musician of the 20th century). The characters' names were selected to agree with the summary descriptions along both linguistic and historical dimensions (e.g., social class, era, etc.). The stories were constructed such that, for 28 of them, there was a strong resemblance to the life of a famous person (living or dead), and for 14 of them there was not.

Rating study. Since the questions of interest to the study revolved around subjects' use of their real-world knowledge, it was important to have an accurate picture of that knowledge. To assemble materials that bore the desired relationships to the subjects' prior knowledge, a rating study was conducted with an additional 20 Yale University undergraduates, each paid $2.50 per hour for participating. Rating subjects were each given a randomly ordered set of pairings between the name of a famous person and a "fact" about some characteristic or life event. The subjects judged the extent to which they knew that the "fact" was true of the person. The six possible responses for each pairing ranged from "I'm sure I knew it was true" to "I'm sure I knew it was false," with less certain categories ("I think I knew it was true/false") as well as guessing categories ("I didn't know it—guess true/false") in between.

The "facts" were stated in rather abstract terms, with all specific reference to times, places, and people replaced by more general descriptions. For example, one fact to be judged about Robert F. Kennedy was, "He was an elected legislator, but from a geographical area other than the one which was his political base," where there is no explicit mention of New York or Massachusetts or the U.S. Senate, although all are implicitly referenced. The rationale for this decontextualization was that it permitted the sentences selected on the basis of the ratings study to be used almost verbatim in the main experiment, where the "facts" would be used to portray fictional characters of different nationalities and historical eras.

In order for a famous person to be included in the main experiment, six facts about him or her were necessary. These six facts had to consist of five facts that the subjects thought they already knew were true, and one fact they thought they did not know before as either true or false. Of the five "true" sentences, four had to be rather closely related to one another (as facts about the same period in the person's life, the same domain of his or her fame, etc.), and one had to be rather different from the other four (e.g., one personal fact with four political facts, or one political fact with four military ones). The one "invented" sentence (the one not known before as either true or false) had to be about the era or domain of the four closely related facts. To ensure enough flexibility in the selection of particular facts to fulfill these requirements, 12 sentences were paired with each of the 34 famous people included in the ratings study. These 12 sentences consisted of 6 facts from each of 2 domains or eras of the person's life; of the 6 in each domain, 4 were intended to be "true" and 2 were intended to be "invented" (although whether a sentence was used as a "true" or an "invented" fact depended only on the subjects' ratings).

Selection of facts to be used in the main experiment was made on the basis of an index computed for each fact as the sum over subjects of the number of "true" responses, minus the number of "false" responses, with "sure" responses given double weight (2 points) over "think" responses (1 point). The maximum score for any fact was therefore 40, and the minimum score −40. A sentence was considered "true" if its index value was 16 or more, and "invented" if its index value was between 4 and −4.

For each famous model, the domain chosen to provide the four related sentences (the major domain) was the one with at least four sentences meeting the "true" criterion and at least one sentence meeting the "invented" criterion. The sentence chosen as the one "true" sentence from the minor domain for each famous person was the "true" sentence from the minor domain whose rating most closely matched the average rating (for that person) of the four chosen major-domain "true" sentences. The mean index value was 30.69 for the chosen major-domain "true" sentences and 31.97 for the chosen minor-domain "true" sentences. The sentence chosen to be the "invented" sentence (if more than one met the criterion) was the one that was given the larger number of "guess" responses, or, if they were equal, the sentence with the more even distribution of "guess true" and "guess false" responses. The mean index value was 0.61 for the chosen "invented" sentences.

Of the 34 famous people included in the ratings study, 28 provided sets of sentences meeting all the criteria.

Story construction. Using the sentences selected from the rating study, stories based on each of 28 famous people were constructed. Each story consisted of four sentences: three randomly selected from the four "true" sentences of the major domain, plus the one "invented" sentence from the major domain. A sample story, as it was presented on the screen, appears in Table 1.

The four sentences of each story were ordered to make the narrative as smooth and coherent as possible. The sentences were used almost verbatim from the ratings study, except

TABLE 1

BIOGRAPHY OF ONE REAL-BASED CHARACTER AS PRESENTED IN THE STUDY PHASE (REAL-WORLD MODEL IDENTIFIED)

YOSHIDA ICHIRO
JAPANESE POLITICIAN OF THE 20TH CENTURY
(PARALLELS TO LYNDON JOHNSON)

HE WAS CHOSEN AS NATIONAL LEADER IN HIS OWN RIGHT AFTER HAVING SUBSTITUTED FOR ANOTHER. HE WAS RESPONSIBLE FOR INTENSIFYING HIS COUNTRY'S INVOLVEMENT IN A FOREIGN CONFLICT. HE DEVOTED MUCH MONEY AND EFFORT TO ERADICATING ECONOMIC AND SOCIAL INJUSTICES. HE WAS MADE AWARE OF RAPIDLY WORSENING PROBLEMS WITH HIS NATION'S ENERGY SUPPLY, BUT HE IGNORED THEM OUT OF FEAR OF POLITICAL REPERCUSSIONS.

for changes in the gender of pronouns (when the fictional character and the relevant model were of different sexes), and minor changes in wording for a few sentences (on stylistic grounds, or because a term was anachronistic or ambiguous)

The remaining two sentences about each real-world model (one major-domain "true" and one minor-domain "true") were intended for use as foils in the recognition task. To prevent pairings using these sentences from being rejected merely on the basis of the sentences never having been seen in the experiment, the sentences were presented in stories about characters (to be called "composite" characters) who were not based on real-world models.

Composite characters were made up by pooling the 56 "leftover" sentences (2 from each of the 28 real-world models), and selecting groups of four sentences that would combine into relatively coherent biographies, with the additional constraint that each composite character be made up of sentences taken from four different famous people.

Test material construction. Materials for the Test task were constructed from the sentences and fictional characters of the Study material. Each sentence appeared twice in the Test materials: once paired with the character with whom it was originally studied, and once paired with some other character. Each fictional character was tested eight times: four times with the sentences originally studied with that character, and four times with other sentences.

Of the mispairings, or foils, some were systematic and some were random. Two sentences from the composite biographies, one "true" major-domain sentence and one "true" minor-domain sentence, were systematically paired with the character based on the famous person of whom the sentences were true. Thus, the two true Abraham Lincoln facts not studied in the story modeled on Lincoln were tested paired with Lincoln's fictional analogue. These two pairings comprise the "related foils" for each real-based fictional character, to be called the "closely related" (major-domain) foil and the 'distantly related" (minor-domain) foil. All the other mispairings were random (providing the "unrelated" foils)

The sentences in the Test phase were taken verbatim from the stories, except for necessary changes in the gender of pronouns (for mispairings where the sex of the Study and Test characters differed) A complete set of the Test pairings for one real-based character is found in Table 2. The same set of 336 pairings (4 sentences per story × 42 stories × 2 pairings per sentence, or 42 characters × 8 pairings per character) was used for all subjects.

Practice materials. A set of practice materials was constructed using sentences from four famous people who almost met the requirements for the main set of materials. These stories and test pairings were constructed as analogously to the main set as was possible using a set of only four real-world models. Minor deviations should not have

TABLE 2

TEST PAIRINGS FOR ONE CHARACTER: KATHLEEN O'MEARA, IRISH POLITICIAN OF THE 20th CENTURY (PARALLELS TO GERALD FORD)

Seen pairings	
SHE WAS APPOINTED SECOND-IN-COMMAND OF HER NATION WHEN SHE WAS RELATIVELY LITTLE KNOWN	(True)
SHE TOOK OVER NATIONAL LEADERSHIP AFTER THE RESIGNATION OF HER PREDECESSOR	(True)
AFTER A HARD FIGHT, SHE WAS NOMINATED BY HER PARTY TO RUN FOR THE OFFICE OF CHIEF OF STATE	(True)
HER SALARY MORE THAN TRIPLED WHEN SHE BECAME SECOND-IN-COMMAND OF HER NATION	(Invented)
Foil pairings	
ALTHOUGH SHE HELD THE PREMIER POSITION IN THE GOVERNMENT FOR SEVERAL YEARS, SHE NEVER WON A NATIONAL ELECTION	(Closely related)
SHE RECEIVED GRADUATE TRAINING AT A PRESTIGIOUS UNIVERSITY	(Distantly related)
HER EARLY MILITARY TRAINING PREPARED HER FOR SOME OF THE TACTICAL DECISIONS SHE HAD TO MAKE AS COMMANDER IN CHIEF	(Unrelated invented)
SHE WAS ASSASSINATED DURING A CAMPAIGN IN WHICH SHE WAS DOING RATHER WELL	(Unrelated true)

affected the material's usefulness in familiarizing the subjects with the tasks and with the kind of materials with which they would be working.

Procedure

The experiment consisted of five tasks, performed in fixed order: a Study phase for the practice materials, a recognition Test phase for the practice materials, a Study phase for the main set of materials, a recognition Test phase for the main set of materials, and a final recall task. Written instructions read in advance informed the subjects about the Study and Test tasks. Before each task, the subject saw detailed instructions specific to that task. Instructions for the Study phase urged subjects to make use of any given information about parallels to famous people in the real world, and to think about the ways in which the fictional character resembled the real-world model. Instructions for the Test phase described the three possible responses to a Test pairing: "S" for a pairing *Seen* in the Study phase (i.e., the sentence was presented as part of the biography of the character); "P" for a pairing not seen in the Study phase where the fact was highly *Plausible* based on the other facts known about the character; and "I" for a pairing not seen in the Study phase where the fact was *Irrelevant* to the other facts known about the character. The instructions stressed that there were two categories of response, Seen and Not Seen, with pairings Not Seen requiring an additional judgment of plausibility. This set of three response choices was included to emphasize to subjects that "recognizing" a pairing is not necessarily just finding it plausible, and to assist subjects in keeping these judgments distinct.

There was a short break between the practice set and the main set of materials (with a chance to ask questions), and two short breaks during the Test phase for the main set of materials. The entire experiment took just under 2 hours on the average, although there was some variability in times since all tasks except the Study phases were self-paced.

Every subject studied all 42 stories and judged all 336 pairings. There were four conditions of model identification for the 28 modeled biographies: the model could be explicitly identified both at Study and at Test (the *Always* condition), at Study but not at Test (the *Study Only* condition), at Test but not at Study (the *Test Only* condition), or neither at Study nor at Test (the *Never* condition). For the 14 stories not modeled on any real person, no model was ever identified. Of the 28 real-based characters studied by each subject, 7 were in each condition of model identification, with the particular 7 assigned to each condition determined randomly for each subject. The order in which stories were presented in the Study phase and the order in which pairings were presented in the recognition Test phase were randomized for each subject.

Study phase. In the Study phase, each story was shown on the CRT screen for a total of 30 seconds, with a warning buzzer 10 seconds before the screen was erased. The display for each story consisted of the name of the fictional character and a summary description (of nationality, occupation, and era) as a heading, and the story in continuous text a few lines beneath (see Table 1). For stories in the Always or Study Only condition for a subject, the heading had an additional line of information—the name of the famous person whom the fictional character resembled (in the form "Parallels to Jimmy Carter").

Immediately after each story was erased, the subject rated (on a scale from 1 to 7) the extent to which the character was appropriate to the given time and place. This rating was included to encourage subjects to encode the stories meaningfully, in terms of the attributes of the fictional characters. For stories where a model was identified, subjects also rated the strength of the parallel between the identified famous person and the fictional character (on a 7-point scale). For stories where no model was identified, the screen remained blank for 5 seconds.

Test phase. After all the stories had been presented, subjects went into the Test phase. Pairings were presented on the screen individually; each included the name and summary description of a fictional character, and a sentence. For characters in the Always and Test Only conditions, the heading also contained the name of the real-world model, in the format used in the Study phase. As soon as the subject pressed a response key, the screen was erased.

Final recall. In the final recall task, subjects saw the name and summary description of each of the 42 fictional characters (in a random order), and were asked to provide the name of any famous person who either had been identified as a model, or whom the subjects had been reminded of by the fictional character. This task was not mentioned in the general instructions; instructions were presented just before the task began. This was done to decrease the likelihood that in trying to prepare themselves for this final task, subjects would form associations between the names of the fictional characters and the names of the models. Such associations would have eliminated the differences between the Study Only and Always conditions by equalizing access to the models' names at Test for the two conditions.

Results and Discussion

Prior to any detailed analysis, the percentage of correct responses was computed for each subject. The range of this index was from 57.4 to 89.6%, with a mean of 74.2% (SD = 8.5%). Analysis of the recognition data (with exceptions to be noted later) was performed on data collapsed for each subject in the following fashion: For each real-based character, responses to each of the eight pairings in which that character appeared were categorized into the six pairing types described in the first six rows of Table 3. The data for the seven real-based characters in each model-identification condition were averaged, yielding a single proportion correct for each model-identification condition for each pairing type

TABLE 3

EXPERIMENT 1: PROPORTION CORRECT RESPONSES BY IDENTIFICATION CONDITION AND PAIRING TYPE

	Identification condition of real-based character			
	Never	Always	Test Only	Study Only
Major domain true (.021)	.637	.840	.607	.554
Major domain invented (.039)	.727	.831	.595	.587
Closely related foil (.030)	.730	.727	.635	.738
Distantly related foil (.033)	.883	.775	.664	.823
Unrelated invented foil (.007)	.942	.979	.963	.926
Unrelated true foil (.012)	.923	.966	.958	.873
All seen (.019)	.659	.837	.604	.562
All related foils (.019)	.807	.751	.649	.780
All unrelated foils (.005)	.933	.972	.960	.899
Corrected recognition (.032)	.566	.766	.387	.433

Pairings with composite characters

Pairings studied	.525
Pairings not studied	.908

Note. Numbers in parentheses are MS error associated with each row.

No distinction was made between the two types of "Not Seen" responses. This procedure provided a 4 × 6 array (4 identification conditions by 6 pairing-types) of proportions for each subject, which appears, averaged over subjects, in Table 3.

Data are also shown in Table 3 for pairings with composite characters, although these data are less interpretable. The extent to which performance is worse on composites than on real-based stories never identified is attributable to either (or both) of two effects. First, the composite biographies themselves may be less cohesive, given the constraints on their construction. Second, any extent to which subjects falsely recognize a sentence when tested as a related foil may decrease their willingness to give a second "Seen" response to that sentence.

All analyses (except as noted) used the arcsine transformation of the proportion-correct scores. Scores for pairings with real-based characters were submitted to a four (identification condition) by six (pairing type) repeated-measures analysis of variance, with subjects completely crossed with both factors. This analysis revealed significant main effects of model-identification condition, $F(3,159) = 32.1$, $p < .001$; of test-pairing type, $F(5,265) = 115.1$, $p < .001$; and of their interaction, $F(15,795) = 10.3$, $p < .001$.

Although the four model-identification conditions can also be treated as a 2 × 2 (study by test identification) design, only one preliminary analysis was performed this way. An overall 2 (study identification) by 2 (test identification) by 6 (pairing type) repeated-measures analysis of variance showed significant effects of all three variables and all four interactions (all at $p \leq .01$). For ease in making the required paired comparisons, model identification was treated as a single factor with four levels.

Comparison of Always versus Never

It was expected that the Always and Never conditions would show certain differences similar to those found in the existing literature for subjects informed versus uninformed about a relevant schema. Thus, planned comparisons of Always versus Never conditions were made for pairings Seen in the study phase and for related foils. For the contrasts described below (and in subsequent portions of the paper), the error term used in computing an F ratio was the interaction between that contrast and subjects. For Seen pairings, the predicted advantage of the Always condition over the Never condition was found; $F(1,53) = 67.7$, $p < .001$. For related foils, the predicted disadvantage of the Always subjects was marginally significant; $F(1,53) = 3.55$, $p = .06$. Thus, the effects that have been reported for schema identification are found for analogies to individuals as well.

Comparisons of All Conditions

The same breakdown of the data (pairing type by model identification) was used in looking at other effects of model-identification condition in addition to the preliminary Always-versus-Never contrasts above. Within each pairing type there was a significant main effect of model-identification condition (all $p \leq .01$); tests of differences among the means of the four model-identification conditions used Tukey's "honestly significant difference" procedure (Winer, 1971). All reported differences between means are significant at $p < .01$ unless otherwise noted.

Seen pairings. For Seen pairings, the Always condition showed better recognition than all other conditions, within both true and invented sentences. The Never condition was better than the Study Only and Test Only conditions, although this effect was significant ($p < .05$) only for invented sentences. Within Seen pairings, then, the superiority of the Always condition is not accounted for by either knowing the model at study alone, or by knowing the model at test alone, but rather requires knowledge of the model at both points for an advantage to be found. Moreover, facts read about characters who were identified only at study were recognized somewhat less than facts read about characters never identified.

The fact that the advantage of the Always condition extended to the invented sentences implies that superior performance in the Always condition cannot be due solely to having access to matching facts, stored in memory, about the famous person. This finding supports the notion that the benefit of having an analogy must be due to the analogy's allowing better elaborations rather than to the analogy's allowing tagging of relevant facts.

Related foils. The Test Only condition had the greatest difficulty in rejecting related foils (65% correct rejection), followed by Always (75%), Study Only (78%), and Never (81%). All pairwise comparisons show significant differences, except Always versus Study Only, and Study Only versus Never. Thus, the presence of the model's name at Test made these foils more difficult to reject, particularly in the Test Only condition. The closely related foils were harder to reject in all conditions than the distantly related foils (71% correct rejection vs 79%), showing that thematic relatedness made foils harder to distinguish from targets. The difference in relative performance on closely versus distantly related foils seems to depend on the availability of the analogy at Test. In the Always and Test Only conditions, the difference between these two types of foils is rather small (5 and 3%, respectively), while in the Never and Study Only conditions, the difference is larger (15 and 9%, respectively). This interaction between thematic relatedness of foil and presence of the model's name at Test is significant, $F(1,53)=8.52$, $p<.01$, leaving no significant residual interaction between model identification and type of related foil, $F(2,106)=1.74$, $p>.15$. Thus, distantly related foils appear much more related to the studied facts when the analogy is given at Test.

Unrelated foils. The Study Only condition was significantly worse at rejecting unrelated foils than all the other conditions (at $p<.05$ or better). In addition, Always was significantly better than Never. Subjects may be worst off in the Study Only condition because they are not being well cued to their Study encodings, nor do they have available (at Test) the analogy which would allow them to reject the foil as implausible. The superiority of Always over Never could be a consequence either of the model's name at Test allowing rejection of the foil as implausible, or of access to a better memory representation of the studied biography in the Always condition. For the Test Only condition, although availability of the studied information is poor (as shown by performance on Seen pairings), the implausibility of the foil pairing (with the model's name present) gives grounds for correct rejection.

Corrected recognition. The question remains whether better recognition of Seen pairings for Always than for Never characters is truly a memory effect, or whether it could be explained as a bias in the Always condition to respond "Seen" to pairings where the fact is true of the identified real-world model. If such a response bias were the sole basis for the Always superiority, then the extent to which related foils are falsely recognized more in the Always condition than in the Never condition should account for the extent to which Seen pairings are correctly recognized more as well.

A corrected-recognition score was computed for each subject for each model-identification condition as a single measure of recognition memory. A corrected measure of memory should adjust for all nonmemorial bases for correct response, adjusting for the effects of guessing on the basis both of thematic similarity to the studied facts and of plausibility of the fact's being true of the model. Using a standard correction for guessing, corrected recognition was computed as

$$\frac{\text{hit rate} - \text{false alarm rate}}{1 - \text{false alarm rate}}$$

where the hit rate was the proportion correct on Seen pairings, and the false alarm rate was the proportion incorrect on related foil pairings. The means of this index are given as the 10th row of Table 3.

On corrected recognition, the differences between condition means were all significant, except the difference between Study Only and Test Only. The Always condition showed the best memory for the studied information, followed by Never. The other two conditions, where identification at Study and Test were mismatched, lagged far behind. The corrected recognition scores indicate that the superiority of Always over Never cannot be due only to the presentation of more information at Test, but must be attributed to Always subjects' having better encodings from the Study phase or better access to these Study encodings. Thus, if this corrected recognition score can be accepted as an accurate measure of recognition memory, it would seem that knowing the analogy at both Study and Test allows the best memory for a biography, but that if the model is identified at only one of these points, subjects are actually worse off than if the model is never identified.

Final Recall Task

Subjects did not tend to report many of the model's names in the final recall task. On the average, subjects produced 8.28 correct names (.37 Never, 2.76 Always, 2.54 Test Only, and 2.61 Study Only models).

Realization of the Study Only Condition

We successfully manipulated the availability of a model's name at Test, as shown by the significant recognition advantage of Always over Study Only—these conditions differed only in the re-presentation of the models' names at Test. If, in the Study Only condition, subjects could have recalled the models' names at Test unassisted, they would in effect have put themselves in the Always condition.

There is further evidence that recognition is improved by having the model's name available at Test, if it was available at Study. The final recall task provides some index of subjects' ability to retrieve the models' names in the recognition task; a regression analysis of Study Only recognition shows the importance of this ability. A stepwise multiple regression on corrected Study Only recognition was performed, with two independent variables: final recall of Study Only models and the mean corrected recognition in the other three conditions (a measure of overall memory). This regression showed that 54% of the variance in Study Only recognition was accounted for by the number of Study Only models generated, and showed that performance in the other three conditions accounted for only an additional 7% of the variance. Thus, the ability to generate the models' names was a better predictor of Study Only recognition than was general recognition ability.

Conclusions

At least three results indicate that in the Always condition, subjects do not just replace the character's name at Test with the model's name and then judge whether the fact is true of the model: (a) the superiority of Always over Test Only, (b) the persistence of the superiority of Always over Never even after correcting for false alarms to related foils, and (c) the similar recognition of true and invented sentences in the Always condition. These results are consistent with the interpretation that analogies affect memory through facilitation of elaborative processing, and not through tagging.

Of course, ruling out a simple tagging account does not preclude some use of a tagging mechanism.[1] The most plausible alternative tagging account for this experiment assumes that subjects tag previously known facts (just as in a pure tagging theory) and devote disproportionate study time to the invented fact so that it too will be remembered. This alternative account, though, is not supported by our data. The partial tagging

[1] This was pointed out to us by Gordon Bower

theory would predict that in the Study Only condition, with the model's name unavailable at Test, there would be an advantage for the much-studied invented sentence over the true sentences, which were merely tagged. The Never condition, which involves no assumption of disproportionate study, is the control for the test of this partial tagging theory. The 9% advantage of the invented over true sentences in the Never condition decreases to a 3% advantage in the Study Only condition, which is the opposite of the prediction of the partial tagging theory.

In this experiment, best memory was found when the context at Test match the context at Study (with respect to the model's name), a result similar to the context effect commonly found for single-word stimuli (e.g., Light & Carter-Sobell, 1970; Tulving & Thomson, 1971, 1973; Watkins & Tulving, 1975). In these earlier experiments, recognition memory for a word was facilitated by a match between the verbal contexts at Study and Test. One explanation for this facilitation that can be applied to the current results as well is an uninteresting context-match explanation. Under a context-match account, facilitation is due to having the identical verbal item simply accompany the stimulus at Study and Test; under this account, the notions of invoking prior knowledge and using analogies are unnecessary. A context-match explanation for the current results entails only the same set of processes used to explain the results of a word-recognition experiment (where no analogy can be involved). The difference between Always and Never (in both of which there is context matching) is explained as an effect of a more distinctive context in the Always condition.

An alternative explanation for the current results (applicable to the single-word studies as well) relies on the notion of elaboration (see Anderson, 1976, Sect. 10.1): Under this account, the context in which an item appears determines the elaborations that are made, and the greater the match between the elaborations made at Study and those made at Test, the higher recognition will be. In this experiment, elaborations would be influenced by the analogous prior knowledge evoked by the model's name. Similar theories have been offered by schema theorists going back to Bartlett (1932), who proposed that memory for an item depended on the schema evoked at study being re-evoked at test; the same schema was needed at test in order to interpret the information set down in the memory trace at study by that schema. Bartlett's proposal differs from the elaboration proposal only in using the notion of "schema" instead of "elaborations."

One effect that would be consistent with the elaboration account but not with context match is an effect of the goodness of the analogy. One prediction of the elaboration explanation is that the nature and number of elaborations the subjects form should depend on the quality of the analogy between the study material and the subjects' prior knowledge. To test whether such an effect had occurred, the analysis described below was performed.[2]

For each subject, responses to Test pairings with the two Always characters he or she judged most strongly resembled the model were collapsed, as were responses to the pairings with the two Always characters judged least similar to their models. Performance measures for these breakdowns, as well as the mean resemblance ratings for the selected characters, are shown in Table 4. Subjects were somewhat better at recognizing pairings involving characters who strongly resembled their models than those involving characters who less resembled their models, despite the very restricted range of resemblance. This weak superiority, tested by a sign test over subjects, was significant: Of 39 differences in recognition of Seen pairings (ties excluded), 27 were in favor of strong resemblance, $p = .025$ (two-tailed). Subjects had

[2] We thank Lynne Reder for suggesting this type of analysis.

TABLE 4

Proportion Correct for Always Characters Judged Most and Least Similar to Models

	Strong resemblance	Weak resemblance
Seen	.85	.80
Related foils	.75	.76
Corrected recognition	.78	.74
Mean rating of resemblance (1 to 7 scale)	6.7	4.0

better memory for characters who had good models than for those who had poorer models, a result that a simple context explanation would not predict.

What this post hoc breakdown suggests is that the quality of the encoded representations for biographies studied with models identified depends on the extent to which the biography as a whole can be integrated with prior knowledge about the real-world model.[3] A better test of this hypothesis, though, would require experimentally manipulating the goodness of the analogies between the fictional characters and their identified real-world models.

Experiment 2

The purpose of this experiment was to obtain strong evidence that the superiority of the Always condition in the first experiment depended on subjects' forming analogies between the new information and previously known facts about the real-world model. It is conceivable that the effect found was due merely to the existence of an association between a distinctive (or a distinctive and familiar) proper name and the fictional character, as a simple context-match explanation would claim. To contrast these possible sources of the Always superiority, four conditions were included in a second experiment. The treatments of the Always and Never conditions were repeated here in unchanged form, and two new conditions were included: one in which distinctive but previously unfamiliar names were given as the "real-world models," and one in which the original famous names were arbitrarily (and thus inappropriately) assigned as "models" to the fictional characters.

Method

The method of Experiment 2 was similar to that of Experiment 1, with differences noted below.

Subjects

Subjects were 41 Yale University undergraduates who were paid $2.50 per hour for approximately 2 hours of participation.

Materials

The stories and fictional characters from Experiment 1 were used again here; the difference between the methods occurs in the set of models' names associated with the fictional biographies. In addition to the 28 famous names from Experiment 1, 28 novel names were constructed to agree as much as possible with the famous names (in terms of ethnicity, superficial form, and length). For example, the name corresponding to Robert F. Kennedy was Michael H. Donovan, and that corresponding to Golda Meir was Hannah Enkol. The goal was to create a set of names that had all the features and properties of the famous names, differing only in prior familiarity. Informal ratings confirmed that no well-known name had inadvertently been included among the counterfeit names.

[3] The facilitating effect of a good model might appear to support a tagging explanation, under the assumption that the "better" models are ones where the to-be-remembered facts more closely match known facts, and thus where tagging is easier. The data, though, indicate that the "better" models are not particularly those where the individual facts are better known. There was little correlation between a story's strength-of-parallel rating in this experiment, and the average rating given the three true facts of that story in the rating study, $r(28) = .29$, $p > .10$.

Design

All subjects saw all 42 stories and all 336 test pairings as in the first experiment; however, the four conditions of model identification within each subject were different. In this experiment, the identification of the model at Study was always matched by the identification of the model at Test. What varied between conditions here was the type of name given as the analogous real-world figure, and the relation of that name to the facts of the biography.

Of the 28 real-based biographies, each subject saw 5 with no model ever identified (the *No Model* condition, equivalent to the previous experiment's Never condition); another 13 were seen with the correct real-world model identified (the *Appropriate* condition, equivalent to the Always condition of the first experiment); another 5 were seen with the name of an incorrect famous model, that is, with the name of one of the 28 original famous models from a different biography (the *Inappropriate* condition); and the remaining 5 were seen with the unfamiliar manufactured names described earlier (the *Counterfeit* condition). The preponderance of stories in the Appropriate condition (13 versus only 5 in the Inappropriate and 5 in the Counterfeit condition) was designed to induce the subjects to consider the models' names as useful information, and to assist the illusion that the Counterfeit names must just be less familiar rather than manufactured.

Stories were assigned to each of the four conditions randomly for each subject. Five Counterfeit names were randomly selected (from the set of 28) for use with the 5 stories in the Counterfeit condition. To select "models" for stories in the Inappropriate condition, a pool was created from the names of the 15 real-world models whose stories were not in the Appropriate condition (that is, of the models for the 5 No Model, 5 Counterfeit, and 5 Inappropriate condition stories). From this pool, five names were randomly selected and randomly assigned to the five Inappropriate-condition stories, with the constraint that a story could not be assigned the name of its correct model.

Procedures

The procedures were the same as those of Experiment 1. The practice phase used the original practice materials, with the modification that three biographies were studied and tested with the correct model and three (one real-based and two composite) with no model. The instructions for this experiment were slightly modified from the previous instructions to include a warning that there might be "some variability both in how much you know about the historical figure, and in how similar the fictional character is to that historical figure."

Results and Discussion

The data were analyzed according to the procedures used earlier. Each subject's responses were collapsed over stories to produce an array representing the proportion of correct responses to each of the 24 types of test items (4 model-identification conditions by 6 pairing types). The overall proportion of correct responses was slightly higher than in the first experiment: Here, mean performance was 80.9% correct, ranging from 64.9 to 93.7% (*SD* = 6.4%). The proportion of correct responses, averaged over subjects, for each sentence type for each model-identification condition is presented in Table 5.

Comparison between Table 3 and Table 5 shows that the portion of this experiment which replicates a portion of the first experiment shows very similar results: The pattern of data over the six pairing types in the No Model and Appropriate conditions parallels that in the old Never and Always conditions (with a correlation of .972), although overall performance is somewhat better in the second experiment.

Analysis of variance of the arcsine-transformed proportion-correct scores showed a main effect of model identification,

TABLE 5

EXPERIMENT 2: PROPORTION CORRECT RESPONSES BY IDENTIFICATION CONDITION AND PAIRING TYPE

	Identification condition of real-based character			
	No model	Appropriate	Inappropriate	Counterfeit
Major domain true (.020)	657	885	709	571
Major domain invented (.034)	.722	861	737	615
Closely related foil (.027)	.810	790	815	766
Distantly related foil (.018)	.917	852	.932	844
Unrelated invented foil (.004)	980	989	.980	961
Unrelated true foil (.005)	971	.981	971	.961
All seen (.017)	673	879	.716	.582
All related foils (.013)	864	.821	874	804
All unrelated foils (.002)	975	.985	977	962
Corrected recognition (.025)	622	853	675	481

Pairings with composite characters

Pairings studied	.565
Pairings not studied	.957

Note Numbers in parentheses are *MS* error associated with each row

$F(3,120)=11.71$, $p<.001$; of pairing type, $F(5,200)=92.60$, $p<.001$; and of their interaction, $F(15,600)=7.15$, $p<.001$. Differences between conditions were assessed using the same means-comparison procedure as before (Tukey's HSD procedure), with differences reported significant at $p<.01$ unless otherwise noted.

Seen Pairings

For all the Seen pairings (that is, for the true and invented Seen pairings combined), the Appropriate condition showed the best performance, and the Counterfeit condition the worst. All differences were significant except for the difference between the Inappropriate and No Model conditions. There was a small interaction between model identification condition and pairing type for the two types of Seen pairings, $F(3,120)=2.13$, $p<.10$. The contrast between the Appropriate condition and all other conditions accounts for this interaction, $F(1,40)=5.70$, $p=.02$, leaving no significant residual, $F<1$. The invented sentences (as in the earlier experiment) seem to be more memorable than the true sentences; they are better recognized as targets and better rejected as unrelated foils in both experiments. Their advantage as targets does not extend to the Appropriate condition, presumably because the Appropriate analogy especially benefits the major-domain true facts.

Related Foils

Combined performance on the two types of related foils showed rather small differences among conditions. These differences were not quite significant: $F(3,120)=2.59$, $p=.06$. Within the individual types of related foils, however, this marginal nonsignificance decomposes into a negligible effect of condition for the closely related foils, ($F<1$), and a strong effect for the distantly related foils, $F(3,120)=6.44$, $p<.001$. This effect for the distantly related foils comes from poorer rejection (lower proportion correct) in the Appropriate condition than in the No Model and Inappropriate conditions (the latter at

$p < .05$), and poorer rejection in the Counterfeit than in the Inappropriate condition ($p < .05$).

Performance on the related foils can be explained in the same terms as the related foil results in Experiment 1. For the closely related foils, subjects have difficulty (compared to the unrelated foils) in all conditions, attributable to the foils' strong thematic relationship to the studied sentences. For the distantly related foils, however, the thematic relatedness is much lower; here, the misleading effect of the sentence's being true of the model lowers performance for the Appropriate condition relative to the other conditions. Poor performance in the Counterfeit condition on distantly related foils probably reflects that memory is generally poorer in that condition; even within the closely related foils, the Counterfeit condition shows the worst performance (although the differences are not significant).

Unrelated Foils

There were no significant differences between conditions for the combined unrelated foils, or for either of the two subtypes of unrelated foils (all at $F < 1$); subjects rejected these foils very consistently (97.5% mean correct rejection) in all conditions. One characteristic that might account for subjects' success in rejecting these pairings is that the pairings are often quite implausible, even considering only the information on the screen at Test; since these foils are random pairings between fictional characters and facts from other characters' biographies, it is often the case that the event or characteristic described in the sentence is obviously about a character of a different profession or era.

Corrected Recognition

Corrected recognition scores were calculated according to the formula given for Experiment 1. On this measure, the Appropriate condition performs best (significantly better than all others), the Counterfeit condition performs worst (significantly worse than all others), and the No Model and Inappropriate conditions fall in between and are not different from each other ($p > .10$).

Conclusions

The pattern of results above supports the claim that the benefit for the Always condition in Experiment 1 resulted from the analogy between the fictional biographies and prior knowledge about the historical figures. The poor performance in the Counterfeit condition excludes a simple context effect: In that condition, there was a cue uniquely associated with each fictional character, but performance was worse than in the No Model condition, where there was no uniquely associated cue at all. The direction of the difference implies that a poor cue can in fact be interfering (relative to no cue). This could have occurred here for a number of reasons—perhaps subjects spent a large portion of the study time in the Counterfeit condition trying to figure out who the model was (with, of course, no success). The lack of difference between the Inappropriate and No Model conditions indicates that the benefit cannot lie in just having a familiar name associated with each fictional character.

GENERAL CONCLUSIONS

These two experiments constitute strong evidence that providing subjects with a good analogy to prior knowledge will both improve their memory for studied material and bias them toward false recognition of statements consistent with that prior knowledge. One important feature of the study materials in these experiments is that the information to be learned had only abstract parallels to previous knowledge. Previously known facts about a real-world figure were not directly applicable to the new character, but had to be connected to the abstracted fact the subject read, and then reinterpreted in terms of the new character. For example, the sentence "Soon after he came to power, his country's powerful near neighbor broke diplomatic ties" has to be connected to prior knowledge in order to be interpreted in terms of Fidel Castro (i.e., that the sentence refers to the United States), and then reinterpreted in terms of the fictional character's time and place. The need for this indirect, reinterpretive processing did not interfere, however, with the effects found in previous experiments where more direct interpretation in terms of prior knowledge was possible. The mechanisms underlying the effects appear to be independent of this abstract–direct distinction.

In terms of the locus of the advantage for the informed condition, in these experiments the superior performance of subjects when they knew of the analogy was not a result of differences occurring only at Study, nor of differences occurring only at Test, but rather was dependent on having the appropriate cue to prior knowledge at both points. The advantage of the always-informed conditions over the never-informed conditions cannot be explained as a result of the availability of a secondary source of information at Test (that is, by the use of prior knowledge of the model to evaluate the plausibility of the test pairings): Recognition of Seen pairings, corrected for guessing on both the basis of pre-experimental relatedness of the fact to the model and on the basis of similarity to the studied material, still shows that the Always/Appropriate conditions resulted in better memory for the studied facts than the Never/No Model conditions.

The current results indicate that the advantage of informed subjects cannot operate through a mechanism of tagging information already in memory, but rather is a consequence of having enough information in memory about the identified model to enable elaborations on and interrelations of the information to be learned. For the Always and Appropriate conditions, facts known beforehand as true of the real-world model were not better recognized than facts not known beforehand. On the other hand, if the two sets of facts are too disparate, as in the Inappropriate condition, the "model" provides no advantage.

The results also indicate that the effects found in prior research were not dependent on the fact that, in those experiments, the comprehensibility and coherence of the study materials were very poor without the cue. Our experiments demonstrate that even with comprehensibility fairly high (as it was for all our material), there is still an advantage of having a cue to relevant prior knowledge.

These results imply that the effects of prior knowledge in both improving memory and introducing bias are not limited to situations where a cue can invoke only general or schematic knowledge, but also occur when a cue can induce the information of analogies between study material and stored information about particular individuals. The processes involved in the effect, then, must not distinguish between knowledge based on a single instance and knowledge based on multiple instances. One possible implication of this finding of no difference is that effects usually attributed to the operation of schematic knowledge are actually produced by analogies to the specific instances from which a schema is assumed to derive. Of course, these results alone do not provide a definitive answer on this issue.

As to the question of what mechanisms underlie the benefit of analogy, our results are quite consistent with the notion of elaborative processing outlined earlier. The elaboration theory would claim that the name of the real-world model provides access to knowledge that can be useful in elaborating the new material; the elaborations that are made with this relevant prior knowledge available are more redundant and fuller, and thus result in better memory performance, than those made without the prior knowledge. These elaborations can be beneficial for both previously known and previously unknown facts. For instance, consider the following statement about Yoshida Ichiro, a fictional Japanese politician of the 20th century: "He was made aware of rapidly worsening problems with his nation's energy supply, but he ignored them out of fear of political repercussions." The information that this character is analogous to Lyndon Johnson makes it easy to elaborate upon the "fact," even though it was not previously known to be true of Johnson. With prior knowledge available, the following interpretation is possible:

> He was straining his support in the legislature with his guns-and-butter policy and did not want to jeopardize these efforts with an unpopular energy program. Perhaps, he intended to turn to the energy problem after bringing the war to a successful conclusion. With a military success, his prestige might rise to the point where he could call on the nation's patriotism to make the sacrifices necessary to conserve energy. Also, having his political base in a region with a major investment in existing energy policy, he was particularly vulnerable to negative political repercussions of either conservation efforts or a shift to new energy sources.

While the elaboration explanation fits nicely with our data, it is probable that the data could also be explained by extensions of other current theoretical analyses of depth of processing (see Cermak & Craik, 1979). Under a depth-of-processing explanation, processing done with the relevant prior knowledge available would be considered deeper than processing done without that knowledge. The advantage of Always over Never would thus be attributed to differential depth of processing in the two conditions. However, an explanation relying solely on a measure of depth of processing at Study cannot account for the advantage of Always over Study Only (where study encodings are equivalent), and certainly cannot account for the advantage of Never over Study Only (where the Study Only encoding is deeper than the Never encoding). To account for our results, then, an explanation relying on depth of processing would have to be expanded to include the notion that similarity between presentation and test encodings interacts with the effects of depth of processing (cf. Lockhart, Craik, & Jacoby, 1975; Morris, Bransford, & Franks, 1977).

What do these results imply, on a general level, about the process of learning by analogy? Our results show that the effort spent on application of prior knowledge can be well repaid by improved memory, but only when some cue to the old knowledge is available at time of retrieval (either generated by the learner or provided by the task). Learning by analogy to prior knowledge helps only when there is a close relationship between the old and new material. This relationship need not be identity at the level of detail, but there must be parallels, at least at some abstract level.

References

ANDERSON, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Erlbaum, 1976.

ANDERSON, J. R., & BOWER, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–123.

ANDERSON, J. R., & BOWER, G. H. A propositional theory of recognition memory. *Memory and Cognition*, 1974, **2**, 406–412.

ANDERSON, J. R., & REDER, L. M. An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N. J.: Erlbaum, 1979.

BARTLETT, F. C. *Remembering*. London: Cambridge Univ. Press, 1932.

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

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

CERMAK, L. S., & CRAIK, F. I. M. (Eds.) *Levels of processing in human memory*. Hillsdale, N.J.: Erlbaum, 1979.

CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

DOOLING, D. J., & CHRISTIAANSEN, R. E. Episodic and semantic aspects of memory for prose. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 428–436.

DOOLING, D. J., & MULLET, R. Locus of thematic effects in retention of prose. *Journal of Experimental Psychology*, 1973, **97**, 404–406.

LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

LOCKHART, R. S., CRAIK, F. I. M., & JACOBY, L. L. Depth of processing, recognition and recall. In J. Brown (Ed.), *Recognition and recall*. London: Wiley, 1975.

MEDIN, D. L., & SCHAFFER, M. M. A context theory of classification learning. *Psychological Review*, 1971, **78**, 314–332.

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

MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 519–533.

POSNER, M. I., & KEELE, S. W. Retention of abstract ideas. *Journal of Experimental Psychology*, 1970, **83**, 304–308.

REDER, L. M. The role of elaborations in memory for prose. *Cognitive Psychology*, 1979, **11**, 221–234.

REED, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, **3**, 382–407.

RUMELHART, D. E., & ORTONY, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum, 1977.

SCHANK, R., & ABELSON, R. *Scripts, plans, goals and understanding*. Hillsdale, N.J.: Erlbaum, 1977.

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

STERNBERG, R. J. *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, N.J.: Erlbaum, 1977.

SULIN, R. A., & DOOLING, D. J. Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 1974, **103**, 255–262.

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

TULVING, E., & THOMSON, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–373.

VERBRUGGE, R. R., & MCCARRELL, N. S. Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, 1977, **9**, 494–553.

WATKINS, M. J., & TULVING, E. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, 1975, **104**, 5–29.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971. 2nd ed.

YNTEMA, D. B., & TRASK, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

(Received December 29, 1978)