

18

An Elaborative Processing Explanation of Depth of Processing

published in: L.S. Cermak & F.I.M. Craik (Eds.), Levels of processing in human memory. Hillsdale, NJ: Lawrence Erlbaum Associates. 1979.

John R. Anderson
Lynne M. Reder
Carnegie-Mellon University

The purpose of this paper is to discuss a theoretical view that we think accounts for the results that have been organized under the rubric of "depth of processing" (to be called DOP). We argue that the variation in memory with DOP is a result of the number of elaborations subjects produce while studying the material, that these elaborations establish more redundant encodings of the to-be-remembered information, and that elaboration is what is critical, especially for long-term retention. Because extent of elaboration is the critical variable, a better spatial metaphor for the DOP phenomena might be "breadth of processing." We argue that depth of processing is as important to prose material as it is to the verbal learning material with which DOP is most commonly studied. With prose, elaborations take on another dimension of importance: They prove to be critical to the comprehension of the material. We make these points about elaboration, DOP, and prose processing with linguistic examples and interpretations of empirical results.

What Mechanism Underlies DOP?

At one level, the term *depth of processing* just summarizes an intuitive viewpoint about what makes for good memory: One can ask subjects to perform various orienting tasks while processing material. One can consult one's intuitions as to which orienting tasks demand "deeper processing." The prediction is that subjects engaged in what seem to be deeper processing tasks will perform better. There exist no explicit rules, however, for measuring the "depth" of a task. It is not clear how well subjective intuitions about depth will hold over a randomly selected set of orienting tasks in terms of predicting memory performance.

Nonetheless, this intuitive rule has worked thus far in predicting which tasks will produce superior performance.¹ The typical contrast (e.g., Hyde & Jenkins, 1973) is one that pits a phonemic judgment against a semantic judgment where the latter clearly seems to require deeper processing and seems to produce better memory. Of course, this intuitive rule of thumb, even if it is a good predictor, is inherently unsatisfactory. It is a matter of considerable frustration that where intuition seems to do so well, it is so hard to provide a succinct theoretical analysis of the mechanisms underlying the phenomenon.

It would be inaccurate to imply that the depth of processing metaphor has not been translated into theoretical proposals. A number of theorists (e.g., Craik & Lockhart, 1972; Craik & Tulving, 1975; Kintsch, 1975; Wickelgren, 1973) have advocated that processing information at various depths results in different types of traces being left in memory. For instance, Kintsch proposes that propositional traces have much slower decay rates than lexical traces. A similar notion has been offered by Wickelgren (1973), who suggests that propositional traces decay less rapidly because they suffer less interference from other memory traces.

Although these assumptions can explain a number of phenomena associated with depth of processing, there are a number of phenomena that cannot be so explained. For instance, the advantage of elaborative instructions for sentential material (e.g., Bobrow & Bower, 1969) cannot be explained by different traces, nor can the advantage of an elaborative context on larger units of prose (e.g., Bransford & Johnson, 1973). Presumably, the subjects in these experiments, under all conditions, are leaving propositional traces in memory. Yet, elaborative instructions leave "more deeply processed" propositional traces, and these seem more memorable.

The Elaboration Hypothesis

We hypothesize that manipulations designed to affect what has been referred to as *depth* of processing are having their effect by changing the *number* and *type* of elaborations stored. We believe that subjects typically store much more than the input presented in the memory situation. Consider the following transcript collected by Anderson (1972) from a subject as she went through 10 words of a larger free-recall list. This is her second pass through the list. The word following each number indicates the word being studied. Following it is a transcription of the subject's remarks. Each underlined word in the transcription is a word that was in the set of 40 to-be-recalled words:

¹T. O. Nelson (1977) has recently written a paper criticizing the concept of depth of processing. One of the things he questioned was the degree to which there was interrater reliability as to the depth of a task. He proposed 13 tasks to be rated. As a curiosity, we took to independently rating these tasks, excluding 2 tasks whose descriptions were ambiguous. The correlation between our rank order ratings was .974.

1. garrison—*garrison, lieutenant, dignitary*
2. dignitary—*crown queen, oh . . . dignitary*
3. vulture—*vulture . . . bird, there was a bird present . . . vulture, bird . . . garrison*
4. disk—*disk . . . disk, record, disk, can't remember statue*
5. crown—*crown queen, the dignitary visits the crown queen, the lieutenant is in the garrison*
6. bowl—*bowl . . . bowl of flowers . . . the dignitary visits the crown queen, and gives her a bowl of flowers*
7. present—I am *present*, I also am a *student*, I think . . . *student*
8. student—*student*, I am *present*, I also am a *student* . . . the *dignitary* is also a *student* philosopher
9. dragon—oh, I forgot all the fairy tales . . . *goose, dragon, mother goose* fairy tales
10. kitchen—*kitchen*, still the mother, the *widowed* mother [p. 374]

This subject was given no special instructions; she was just asked to say out loud what she was thinking about during study. She is not an atypical subject. Moreover, we believe that the protocols just reflect the tip of the elaborative iceberg. From our introspections, it became apparent to us that the ideas and elaborations that occur during a memory experiment are generated at a rate that is too rapid to report verbally.

There is probably a purpose for the large amount of elaboration that is generated. We speculate that the rich elaboration affects memory performance. That is, we take the depth-of-processing results as an indication of a function and a consequence of the elaborative process—improved memory for material elaborated.

There are two critical questions that need to be addressed in making this theoretical proposal concrete. First, why should amount of elaboration affect memory performance? Second, why should amount of elaboration vary with depth of processing?

To answer these questions, it is necessary to articulate our view of the nature of the representation of information in a memory task. We assume that long-term memory is a network of interconnected propositions and that when a subject goes through an experiment, the subject adds propositions to this memory network. At a minimum, the subject adds propositions encoding the memory items. So, given the paired-associate "dog—chair" the subject would encode the proposition: "In the context of the experiment, I learned that *chair* was the response for the stimulus *dog*." Typically, of course, the subject encodes much more than this minimum. Frequently, this additional elaboration will be about the meaning of the words rather than the words per se.

Any particular encoded proposition is fragile. There is a significant chance that the subject will not be able to activate that proposition at test. So, if a person's memory for the item rested on the minimum proposition, poor memory

would be the result. However, if the subject encoded multiple propositions that were partially redundant with the to-be-remembered information, he or she would have a much better chance of recalling it at time of test. For instance, suppose the subject gave the following semantic elaboration of the pair, "dog-chair":

Elaboration (a). The dog loved his masters. He also loved to sit on the chairs. His masters had a beautiful black velvet chair. One day he climbed on it. He left his white hairs all over the chair. His masters were upset by this. They scolded him.

Suppose the subject would not recall at test the original pair but could recall some elaboration from memory such as, "The dog climbed on the chair." Then, cued with "dog," the subject would have a good chance of guessing "chair" as a response. The probability of selecting "chair" over other items in the proposition would depend on such things as a knowledge of the statistics of list construction (e.g., noting that all responses were nouns), having the word tagged as a response, and the amount of elaboration of this concept versus alternatives like "climb."

It is interesting to note that we are making a commitment to a reconstructive² interpretation of memory like that advocated by Bartlett (1932) and Neisser (1967). The basic idea is that a memory episode is encoded as a set of propositions. This set can vary in its richness and redundancy. At time of recall, only a subset of these propositions will be activated. The richer the original set, the richer will be the subset. Memory for any particular proposition will depend on the subjects' ability to reconstruct it from those propositions that are active. This ability will in turn depend on the richness of the original set and hence on the amount of elaboration made at study.

Having now dealt with the question of why elaborations help recall, we turn to the question of why richness of elaboration should vary with depth of processing. Why should subjects be unable to form as rich an elaboration of peripheral information as of semantic information? For instance, consider our "dog-chair" example. Why could subjects not form to themselves the following elaboration about the pair as typographic objects?

Elaboration (b). The word *dog* is in the book. The word *dog* is also known to be above the word *chair*. The book has the word *chair* printed in large red letters. On one page, the word *dog* is larger than the word *chair*. The word *dog* has its green letters printed beside the word *chair*. The book tells about this. The book illustrates the word *dog*.

²The reconstructive position was one that has been criticized by Anderson and Bower (1973). The Anderson and Bower criticisms were largely directed at interpretations of certain empirical data by reconstructionists. Although these criticisms are valid, we feel that the phenomena of depth of processing and comprehension of connected discourse demand elaborative and reconstructive processes. So, though the first author may quibble about the data and reasoning used to argue for the reconstructive position, we do agree with the general conclusion

This elaboration has an interesting property vis-à-vis the semantic elaboration illustrated earlier. They are basically isomorphic with respect to the graph structure of the propositions interconnecting them. This is illustrated in Figure 18.1, which sets forth schematically the structure of interconnections among the propositions. So, with respect to redundancy of connections, there is no necessary difference between orthographic and semantic processing. However, there are three observations to be made about Elaboration (a) versus Elaboration (b):

1. There is an enormous difference in the ease of generating the two types of elaborations. Semantic elaborations seem to come to mind without problem, whereas generating orthographic elaborations is like pulling teeth.
2. Even if differences in ease of generating the elaborations were removed by giving subjects (a) or (b) to study, subjects would still be likely to do better on (a) than (b). The reason for this is that it is easier to further elaborate on (a) than (b).
3. Even if the experimenter could manage to have the subject encode *only* elaboration (a) or *only* elaboration (b), there would still be an advantage for (a). This is because one's reconstructive processes are better able to interpret the "semantic" remnants at delay.

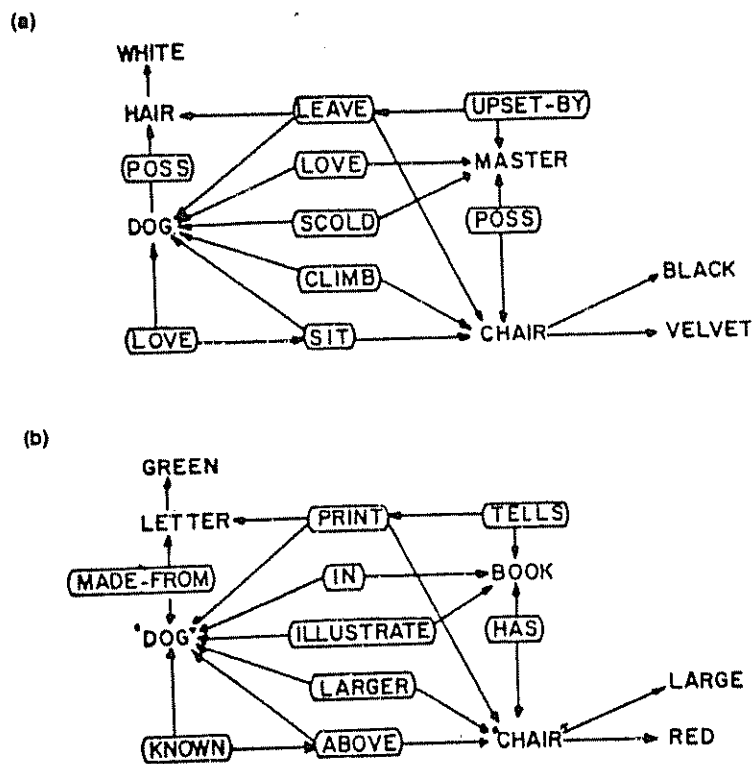


FIG. 18.1. A graph structure illustrating the connectivity among concepts in (a) the semantic elaboration and (b) the typographical elaboration. From Anderson (1976).

We realize that the foregoing remarks presuppose some empirical evidence about the superiority of elaborations like (a) to elaborations like (b), and in fact, we do not have such evidence. However, we regard that result as so obvious that we have not been motivated to perform the experiment.

A critical assumption in our theory is that subjects are better able to *make* certain types of elaborations at study and are better able to *interpret* certain types of elaborations at test. A person is better at elaborating that type of information with which she or he has had the most experience. This is because one's elaborative abilities are a function of the availability of mental procedures to make elaborations, and these procedures do not come into existence except through practice. For two points of view as to the detailed nature of these elaborative procedures, see Anderson (1976) and Reder (1976; in press). We do not have to go into such technical detail here to make our points. It should be noted, however, that the type and extent of elaborations are the product of our real-world experience with concrete objects such as pets and furniture. A member of a non-Western culture would produce elaborations very different from those in (a) given earlier and might be no better able to elaborate the statement: "The dog is on the chair" than the statement: "The word *dog* is next to the word *chair*." It is also conceivable that by suitable experience and practice, a person could become facile at generating typographical elaborations as well as semantic elaborations of pets.

It is not the depth of processing per se that is important, but one's prior practice at making elaborations about various types of information and practice at interpreting the previously stored elaborations. The "better" processing is that which generates *more* elaborations of the input that can be interpreted at retrieval. For most people, semantic elaborations are easy to generate and facilitate recall. For a mnemonist, however, other kinds of elaborations might be more useful. An astronomer can extensively elaborate star patterns, whereas most people cannot. The instructions that can produce rich elaboration and the materials that can be richly elaborated must be defined with respect to the processor. The most critical determinant of retention is the number of elaborations. Holding "quality" of elaborations (i.e., whether they are deep or shallow) constant, the "quantity" of elaborations will predict recall; manipulating "quality" (to the extent this can be measured) while holding "quantity" constant should not be as good a predictor.

Elaboration and Other Theoretical Viewpoints

To review and expand upon our theoretical position: We assume information is encoded in a network of propositions interconnecting concepts. A subject encodes a to-be-remembered event by adding further propositions to the network. In a memory test, the subject is cued with a probe that gives him or her

direct access to some concepts in memory. From these concepts, the subject must retrieve enough propositions encoding the event to permit recall. Our analysis is focused on the process by which a subject searches out from a concept node to try to retrieve the correct propositions. Connected to such a node will be a set of n irrelevant propositions encoding information unrelated to the memory episode and a set of m relevant propositions encoding the memory episode. We view the relevant propositions as highly redundant. That is, the subject need only retrieve a few of these m propositions to enable recall. There are two important variables affecting recall— n and m . The n irrelevant propositions provide interference—as this number increases, memory performance should deteriorate, because it will be harder to find the relevant information. For a review of evidence relevant to this interference prediction, see Anderson (1976, Chap. 8). We will return to the effect of n later. The m relevant propositions provide redundant encoding of the information—as this number increases, memory performance will improve. We see elaborations as affecting this redundancy factor.

Note that this analysis replaces the qualitative concept—depth of processing—with a quantitative concept—number of elaborations. We claim that people are better able to elaborate certain kinds of information than others and that “deep,” semantic information tends to be more conducive to elaboration than “shallow,” phonemic information. However, the degree of “depth” is not a function of semanticity per se, but rather is a function of the extent of past experience with the information. Thus, we speculate that a phonetician would find the phonetic level very helpful. Also, within the semantic level, we expect to see large variations depending on the amount of experience with that topic.

We agree with Fisher and Craik's (1977) conclusion that certain types of information are inherently easier to remember than others and find ourselves in disagreement with the extremely relativistic positions put forth by Bransford, Franks, Morris, and Stein (Chapter 15, this volume) and Tulving (Chapter 19, this volume). No doubt, if one encodes semantic information, one will be ill-prepared for a phonetic test and vice versa, but it still can be meaningful to inquire, within the context of a theoretical framework, which type of information is more easily encoded. For instance, even though two different input devices are encoding different information into a computer memory, it makes sense to ask which is encoding more information (as measured within an information-theoretic framework) and why.

We would like to consider one class of explanations of depth of processing that seemed especially popular at this conference. This kind of explanation states that a deeply processed item suffers less interference (Wickelgren, 1973), or is more distinctive (Eysenck, Chapter 5; Nelson, Chapter 3, this volume), or is more differentiated from other items (Bransford, et al., Chapter 15, this volume). These various explanations are at different stages of articulation and

are probably not identical, but they all seem to hinge on the idea that the target memory must stand out from an interfering background of other memories.

These "interference—distinctiveness" interpretations have often been applied to the semantic—phonemic difference. It is argued that there are relatively few phonemic features but an abundance of semantic features. Therefore, phonemic features have occurred in more contexts and are subject to greater interference from past knowledge. This explanation seems unsatisfactory, because the assumption of a greater vocabulary of semantic primitives is not well motivated. One could propose a semantic vocabulary of 0's and 1's or a richer phonemic vocabulary. A common problem with these explanations is that a given memory trace is said to be more distinctive than another trace if memory for the first is better; i.e., the explanation tends to be circular.

Earlier we alluded to "the effect of the size of n "—the number of irrelevant propositions emanating from a concept node. The larger n , the greater the interference when looking for a relevant proposition. The need for elaboration of a proposition is less when there is less interference. We believe many of the claims made about distinctiveness can be thought of as differential interference. The results of Eysenck (Chapter 5, this volume) and Nelson (Chapter 3, this volume) are quite consistent with the interference view. However, to show that there is an important effect of interference is not to show that there is not an effect of the other variable in our model, viz., the amount of elaboration. We were forced to this second variable, because it seems clear that there are many phenomena that cannot be accounted for by interference.

Consider the two "dog—chair" paragraphs mentioned earlier. Certainly both are quite distinctive, but the semantic paragraph seems guaranteed to lead to better memory. An even more cogent argument for the insufficiency of the distinctiveness account comes from a memory experiment by Goldstein and Chance (1971), where the materials were pictures of faces and of snowflakes. Both types of stimuli were quite distinctive and discriminable. If anything, the features that define snowflakes are more unique and suffer less interference from other knowledge. However, memory for the faces is much better, presumably because subjects are able to attribute meaning to these stimuli (i.e., elaborate upon the stimuli). The more meaningful the stimuli, the greater the propensity to generate elaborations.

Memory for pictorial material is a particularly good domain to make the case for elaboration. Pictures are so rich in details that distinctiveness per se would seem unimportant. Bower and Karlin (1974) demonstrated better memory for faces processed under a "deep" orienting task (judgments of likability) than when processed under a "shallow" task (judgments of sex). Similarly, Bower, Karlin, and Dueck (1975) demonstrated that subjects' memory for doodles (ambiguous cartoons) was better when a meaningful interpretation was applied to the material. Given the high distinctiveness of the stimuli to be learned, we feel that these depth-of-processing results clearly implicate an explanation in terms of richness, or number, of elaborations in the memory code.

PROCESSING OF PROSE

It might seem odd to claim that subjects are engaging in so much elaboration, since much less would seem necessary for a typical experiment. After all, it is a very simple thing a subject is asked to do in a typical DOP experiment—e.g., remember a word. Moreover, if a subject brings in a powerful mechanism that generates dozens of propositions, an equally powerful reconstructive mechanism is required to interpret them at test. This has the character of using a sledgehammer to swat a fly. The reason we posit these mechanisms for this task is because we view them as general mechanisms not especially designed for a verbal-learning experiment or for any memory experiment at all. Rather, we think they are processes critical to dealing with many everyday situations such as comprehension of a message, making an image of a description, predicting what will happen in the future, day-dreaming, finding contradictions, etc. These processes get recruited in the typical memory experiment simply because they are there. One reason that we think it is useful to consider these elaborative processes in the context of prose comprehension is that it provides an additional perspective on their significance.

At the outset, we would like to disavow a certain interpretation of the foregoing remarks and subsequent remarks. There is a position that has been espoused informally and to some degree in print (e.g., Schank & Abelson, 1977; Thorndyke, 1975), which asserts that the "correct" domain to work in is prose processing or, even more specifically, "story" processing; that working with more "impoverished" material leads to distortions of the basic phenomena. This point of view and the methodological imperialism it leads to is nonsense. The human engages in a wide variety of behavior. There is no reason to believe that story comprehension of the kind currently in vogue is more representative of normal processes than is free recall. In fact, there are probably more adult behaviors (e.g., remembering a grocery list) that come close to a free-recall experiment than adult behaviors that come close to reading one of the very simple stories typically employed. The fundamental feature of human behavior is its extreme diversity, and the hallmark of human intelligence is its ability to manage that diversity. It seems silly to regard any paradigm as prototypical. A theorist may miss important generalities in human behavior if the tasks studied are restricted to a narrow paradigm.

At any one time, of course, a psychologist must study human cognition within the context of a particular paradigm. However, this should not obscure the fact that the studied mechanisms operate over a much larger range of tasks than could be produced in one experiment. In trying to understand a mechanism like elaboration, one should learn from all available paradigms. We think that there is a lot to learn from DOP and its typical verbal-learning methodology. However, we feel, too, that there is much that can be added to this knowledge by considering prose processing.

Comprehension of Prose

An important component to the processing of a text is detecting the connections among the sentences of the text. In normal text, there is as much left unstated—to be read between the lines—as there is directly stated. Reading or listening are slow perceptual processes relative to the speed of mental elaboration. Connected discourse almost always consists of the main points and leaves the rest to be filled in by the reader. Consider the following subpassages from Reder (1976):

1. Tim wanted a new model airplane. He saw the change lying on his father's dresser.
2. The heir to a large hamburger chain worried that his wife did not love him anymore. Perhaps he consumed too much beer and french fries. No, he couldn't give up the fries. Not only were they delicious, he got them for free!

Clearly, in comprehending these texts, one almost automatically makes a large number of connecting inferences. For instance, in 1, the reader infers that the money isn't Tim's; it is his father's money; he wants it so that he can buy a model airplane. In 2, the reader infers that the heir is fat, that his wife may dislike fat men, that he gets his fries from his parents' hamburger chain, etc.

We believe that generation of elaborations is the means by which these connections are made. In addition to the obvious, necessary inferences, it seems that other elaborations spew forth too. These other elaborations are often idiosyncratic. The idiosyncratic elaborations made by the second author (LMR) for (1) include: Tim is about 8 to 12, has a crew cut; the father's dresser is just at Tim's eye level; the model airplane is silver with chevron decals; the father is the absent-minded type who would not notice the change missing but who would be furious if he found out his son took it. For (2) LMR's elaborations include: The hamburger chain is like A & W Rootbeer or Jack-in-the-Box; the wife is pretty and independent, yet had convinced herself before marriage that she loved the heir because she really needed the money; the heir is stupid, lazy, and vaguely miserable about life. Many of these elaborations would only be generated by LMR and are of only moderate plausibility. Some may be useful to connect subsequent lines of the passage with earlier lines. Other elaborations such as Tim's crew cut might seem irrelevant to any comprehension function. However, it is actually difficult to judge in advance what will prove relevant and what will not. Consider the following "trick" story from Reder (1976):

3. Alice went to Jimmy's house for lunch. Jimmy's mom served them tunafish sandwiches. Alice liked her sandwich very much and had almost finished it when, all of a sudden, her dentures fell out of her mouth.

Readers quickly detect the humor in this passage, and the comprehender knows why it is funny; however, it is much harder to pinpoint exactly how the anomaly was detected. It is worth considering the introspections LMR reported as to how she detected the anomaly. Part of her elaboration of this passage involved Alice having the smooth skin of a young girl. When LMR heard about the dentures falling out, she elaborated a wrinkled face with an old mouth and exposed gums. The anomaly was detected by noting the contradiction of these two facial features. Thus, elaborations about personal appearance proved critical to detecting the contradiction.

The idea that elaboration may be an important component in human cognition is old. In perception, a similar notion was advanced a century ago by Helmholtz. He proposed that a person combines information from impoverished stimuli with general world knowledge to make inferences about these stimuli. We believe that the predictive and inferential function of the elaborative process is at least as essential as its function in promoting good memory. This function of elaboration is clear when one considers the need to find connections among sentences in prose material, whereas it is not so clear when one considers the free-recall task.

With respect to prose comprehension, it is also clear that it is not functional for people to be good at elaborating orthographic information, whereas it is functional for them to be good at elaborating the semantic content of a passage. That is, there is seldom any payoff for remembering orthographic information, whereas there is frequent payoff for remembering the semantic content of messages. As a consequence, people have learned how to elaborate on semantic content but not on the orthographic. If the nature of our world changed and orthographic information became critical, we would expect to see a gradual change in one's ability to remember orthographic information.

Experimental Evidence

From the point of view of providing strong empirical support for the elaboration theory, a serious problem arises in that the theory implies that an experimenter has poor ability to manipulate the amount and direction of elaboration. For instance, in the typical verbal-learning DOP study, we propose that the amount of elaboration is a function of past experience. This means one must rely on intuition as to what sorts of information subjects are more adept or practiced at elaborating. Unfortunately, in this respect, we have not avoided the unsatisfactory intuitive explanations that have characterized other approaches to DOP.

We believe, on the other hand, that it is possible to obtain more compelling evidence when studying prose material. This is because the richer stimulus material offers more potential to control a subject's processing. We review some of the experimental literature relevant to elaborative processing of prose and then present a summary of some of Reder's work designed to get at these issues.

There are a number of results concerning selectivity in memory that are

consistent with an elaboration—plus—reconstruction viewpoint. If subjects have more ability to make certain types of elaborations than others or if subjects are directed to make certain elaborations rather than others, one should see better memory for material consistent with the preferred elaboration and more distortion of material in the direction of the preferred elaboration. The classic example, of course, is the research of Bartlett (1932). He had pre-World War I English subjects study a northwest Indian story, "The War of the Ghosts." He obtained what he interpreted as systematic distortion of the material in the direction of the knowledge of his subjects. This distortion took the form of additions to the material that made the story more consistent with the world view of his subjects, deletion of inconsistent information, and transformations of inconsistent information to make it more consistent with prevailing beliefs.

There has been a long history of debates (e.g., Anderson & Bower, 1973; Gould & Stephenson, 1967; Spiro, 1977) over the extent to which Bartlett's subjects were really misremembering and the degree to which they were knowingly confabulating in response to perceived task demands. It seems that at least to some degree, subjects are aware of their distortions and are able to assign lower confidence to these than to veridical recalls. However, we feel that this debate misses an important point: The behavior of subjects in Bartlett's task is typical of prose processing. Normally, the reader does not make distinctions between what was actually read in a passage and what is a plausible inference. With most stories, the inferences made are plausible extensions of the story and are not distortions. It was Bartlett's clever story selection that served to highlight the elaborative behavior of subjects.

A number of recent experiments have shown that subjects can be influenced by information they are given about a passage. For instance, Sulin and Dooling (1974) had subjects study identical passages except that the main character was either named Gerald Martin or Adolf Hitler. Subjects showed much greater confusion to foils true of Hitler when they had studied the passage that used Hitler's name. Subjects showed a somewhat similar pattern of confusion when they only learned that the individual was Hitler after reading the passage (Dooling & Christiaansen, 1977). This suggests that distortion can occur at time of test as a reconstructive process and need not operate as an encoding process.

Bower (1976) reports an interesting experiment looking at the effect of prior information on subjects' memory for a passage. Subjects were given a story that consisted of subpassages. Half of the subjects were given prior information that would suggest an interesting, unusual interpretation; half were given no prior information. The story follows the principal character through five episodes: making a cup of coffee in the morning, visiting a doctor, attending a lecture, going shopping in a grocery store, and attending a cocktail party. The meaning of these episodes can be very different depending on whether or not we view the heroine as pregnant. Subjects given the interesting interpretation recalled many more inferences appropriate to the pregnancy theme. They also recalled more of

those episodes related to the theme. This is just what we would expect if they were using their information about pregnancy to elaborate. These elaborations should make the text information more redundant and introduce additional inferences.

Hayes (1976) has found a similar correlation between number of intruded inferences and overall memory for text. Hayes and his colleagues tried to find out what mechanisms allow some people to remember more than others. They pretested subjects on their memory for various historical facts and then classified them as those who remember a lot of history and those who do not. The subjects were then given a fictitious history passage to read. The same subjects who knew more veridical history performed better on a test of the fantasy history passage. Subjects were also asked to free recall the passage that they had read. Not only did the subjects with better history memory recall more, they also "recalled" many elaborations that were not asserted. These elaborations were not simple paraphrases of the passage, nor were they simple inferences. The subjects classified as having poor memory for history offered almost no elaborations. From this finding, Hayes conjectured that embellishing the input with elaborations promotes better retention.

A recent experiment by Schallert (1976) makes similar points. She looked at ambiguous passages, either biasing a passage's reading by giving prior information or not biasing it. She found that subjects in the biased group remembered more information consistent with the bias. She introduced an important DOP manipulation in which subjects either processed the sentence at a "shallow" level (counting four-letter words) or a "deep" level (rating for ambiguity). She found that the biased subjects were more likely to remember consistent information when they were processing the material at the semantic level. The elaboration hypothesis predicts that subjects should be generating more elaborations under semantic-orienting instructions. It also claims that elaborations are responsible for this biasing in recall. Therefore, it predicts the interaction found by Schallert between DOP and disambiguating information.

Manipulation of Prose Elaborations

Reder (1976; in press) performed experiments that manipulated quite directly the amount of elaboration given to prose material.³ She had subjects study short stories of about 20 sentences. An earlier study had shown that subjects have very good memory for these stories even after a week's delay. So the dependent measure chosen was not how well subjects could answer questions, but rather the

³The most direct manipulation would have been simply to instruct subjects to "elaborate more." This direct manipulation was avoided because of the obvious problems of demand characteristics. It is also uncertain whether subjects could directly relate such a verbal command to the desired underlying operations. Finally, it is a rather bizarre request.

speed with which they could answer questions. The use of reaction-time measures was also important to the additive factors logic that she used. The intent of the study was to encourage subjects to process these stories in the same manner they use outside the experimental setting—which we claim would be a rich elaborative manner. Therefore, subjects were not required to make verbatim memory judgments about the story; rather, they were required to judge the plausibility of the probe with respect to the story. One of the variables of the study was the ease with which subjects could make such plausibility judgments. Three kinds of targets were used. Although these targets were all clearly plausible, they varied in the ease with which they could be judged as plausible. There were verb-based statements that followed directly from the verb of one of the sentences. These are statements that are highly plausible but are seldom spontaneously generated by subjects. (One independent group of subjects generated elaborations to the story, and another rated the plausibility of those generated elaborations.) For instance, if the original text contained: "The heir told his father he wanted no part of his greasy food fortune," the subject might be asked to judge the plausibility of: "The heir communicated with his father."

A second class of statements, called high-plausible statements, were both high in judged plausibility and in frequency of generation. For instance, suppose the original text contained: "The heir decided to join Weight Watchers. Twenty-five pounds later, he realized his wife did love him after all." A high-plausible statement to judge might be: "The heir lost weight."

The third type of statement, the so-called medium-plausible statements, were lower in plausibility than either the high or verb-based statements. They were more frequently generated than the verb-based but less frequently than the high-plausible. For instance, if the subject had studied: "Now he worried that she [his wife] had been after his money all along," the to-be-judged statement was: "The heir had not worried about her motives before marriage."

The average lengths of the three types of probes were equal. For reasons unnecessary to unpack here (but see Reder, 1976), it was expected that both the medium-plausible and verb-based statements would take longer to judge than the high-plausible. Of principal interest was how this variable would interact with the other two variables of the study.

A second variable, called treatment, was a manipulation designed to vary the amount of elaboration relevant to making the judgment. At one extreme, the statement was actually *presented*. This should give the maximal opportunity for elaboration. At the other extreme, the *not-presented* condition, no effort was taken to induce relevant elaborations beyond presenting the story. In the treatment between these two extremes, a question was asked during the story that focused the subject on that part of the story that was relevant to answering the question. This was called the *primed* condition. For instance, the subject might read: "Anyway, real marital strife lay elsewhere. His wife had never revealed before marriage that she was an intellectual, that she read books." Then the

subject would be asked to judge the plausibility of: "The heir was delighted that she joined the Book-of-the-Month Club," which is, in fact, not plausible. This question was a prime for judging the plausibility of: "The heir did not like the fact that she read books."

The final variable was the delay at which the subject was asked to make plausibility judgments. The subject would either be asked immediately after the relevant portion of the story or after the story had been completed. In the immediate-presented condition, the statement as a test query followed immediately after its presentation as a statement in the story. The test also followed immediately after asking the prime in the immediate-primed condition. In the delay conditions, the test was approximately 2 minutes after the reading of the relevant portion of the story.

These three variables—plausibility, treatment, and delay—were combined factorially to create 18 conditions. The delay manipulation should affect the availability of information with which to make the plausibility judgment. Treatment (presented versus primed versus not presented) should also manipulate availability of information by affecting amount of elaboration. Therefore, since treatment and delay affect the same process, one would expect them to interact. The plausibility of the statement should affect the ease of making the reconstructive computations that decide plausibility from the elaborations. If one assumes that the subject first retrieves relevant information and then computes plausibility, there is no reason to expect plausibility to interact with the other two variables. It affects a later stage in the information processing.

Figure 18.2 plots reaction time to judge plausible statements as a function of the three variables. "Foil" implausible statements were also used to keep subjects honest, but they did not vary systematically. The pattern of data corresponds quite closely to the predictions of Reder's (1976) model, which is a formal version of the elaboration model. All the main effects are highly significant. Of critical importance, there is a two-way interaction between treatment and delay: $F(2,86) = 35.8$; $p < 0.001$. The presented condition initially was much faster than the other two, but at a delay, the presented statements took about as long as the primed. The difference between primed and not-presented, on the other hand, grew with delay. The difference between primed and not-presented was significantly bigger in the delay condition than in the immediate, $t(132) = 2.3$, $p < 0.05$. Therefore, the interaction of treatment with delay is not solely attributable to a big slowdown in the presented statements. That possibility would have made the interaction less interesting: In the immediate-presented conditions, subjects may have been faster due to the repetition of an identical phrase, which may have enabled them to bypass normal comprehension mechanisms.

Plausibility was not affected by treatment or delay; i.e., there was no significant interactions with this factor. The lines, connecting levels of inference type over treatment and delay conditions, are essentially parallel. There is one

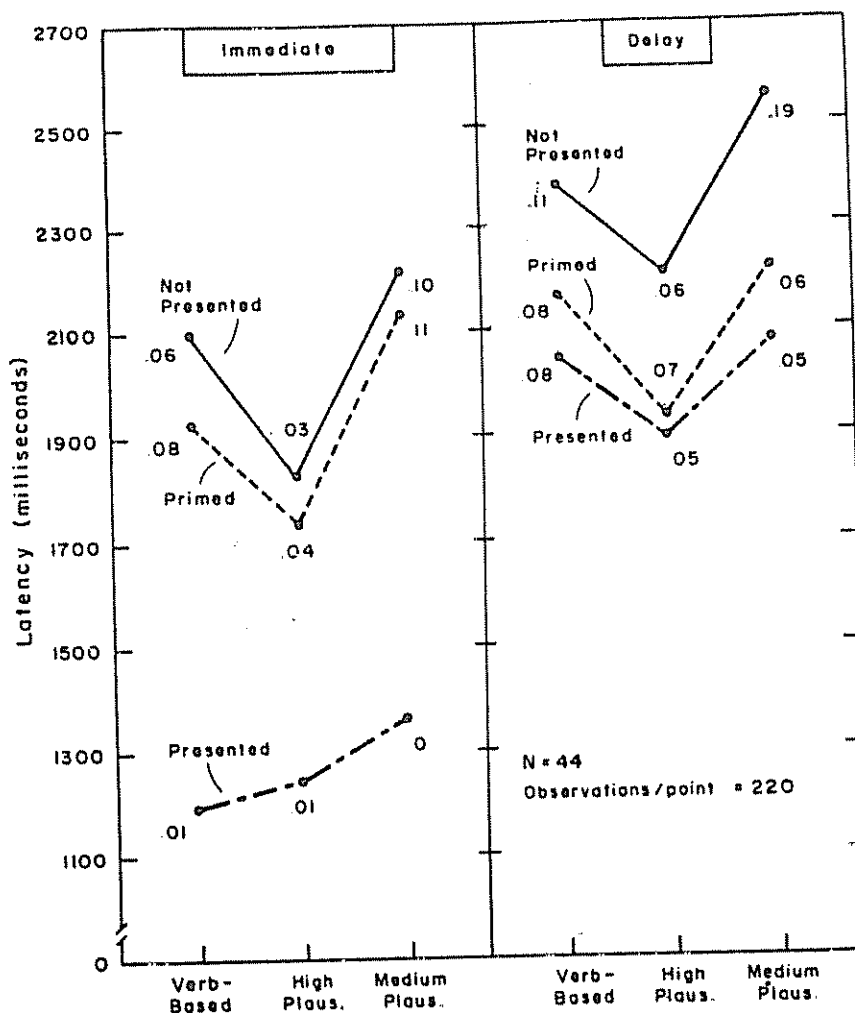


FIG. 18.2. Mean latencies of correct plausibility judgments (and error rates) as a function of plausibility type, treatment, and delay. From Reder (1976).

exception to the rule of parallel lines, the immediately presented condition. Although there was no overall significant interaction with plausibility, a special contrast was constructed that compared the plausibility effects for the immediate-presented condition with the average effects across treatment and delay. The immediate-presented condition had a different effect due to plausibility than did any other condition, $F(2,172) = 3.54, p < 0.05$.

The immediate-presented condition may involve somewhat different processing than the other conditions. For instance, it is possible that subjects were sometimes making their plausibility judgments on the basis of a template match with a verbatim trace of the sentence still active in short-term memory. In none of the other conditions is this likely. It is remarkable how similar the plausibility effect is in these other conditions. The parallel character of these plausibility functions would not have been predicted on one prevalent view of sentence processing (e.g., Anderson & Bower, 1973; Kintsch, 1974; Schank, 1975). This

view claims that one tries to verify sentences by simply retrieving them from memory, only resorting to more elaborate inferential reasoning if the statement cannot be found. That view is not consistent with the obtained plausibility effect in the delayed presented condition, since other memory data of Reder mentioned earlier indicates that the presented statements are stored. The reconstructive view of plausibility judgments, on the other hand, is consistent with the results, even the effect in the delayed presented condition.

Upon introspection, it seems clear from everyday examples that plausibility judgments are typically made by means of reconstructive computations and not by means of direct retrieval. For example, to decide if the three characters in *No Exit* were happy or if the young boys in *Lord of the Flies* were savages, one does not search memory for a specific proposition that asserts that the *No Exit* characters were unhappy, nor for a proposition that asserts that the boys in *Lord of the Flies* were savages.

Directions for Future Research

What most impresses us about elaborative processes is that they seem to provide a mechanism for producing powerful effects in overall level of recall. Although measurements using reaction time—as in Reder's experiment—are often theoretically more sensitive, the striking effects should be seen in percent recall. Our respective research endeavors in this area are aimed at discovering what manipulations can increase the amount of relevant elaborative processing that a student can do for prose material and whether these manipulations have their anticipated effects on percent retention. Many mnemonic devices advocated for learning material (see Bower, 1970, for a review) are not directly related to the content of the material. For example, a Roman orator using the method of loci to remember a speech might imagine a puddle of water in front of a temple to prompt discussion about water projects. On the other hand, elaborative processing that also facilitates retention is naturally associated with the studied material. The act of elaborating text is basically "exercising" the reader in thinking about the content. We are excited about the theoretical notions of elaborative processing, because we feel this theoretical analysis may be setting a firm conceptual base for practical applications to human memory.

ACKNOWLEDGMENTS

The preparation of this manuscript was a collaborative effort. The preparation and the research reported were supported by NSF grant BNS76-00959 and ONR grant N00014-72-C-0242 to the first author and by an NSF Graduate Fellowship, a Rackham Dissertation Grant, and an NIMH postdoctoral fellowship to the second author. We would like to thank Miriam Schustack for her comments on the manuscript.

REFERENCES

- Anderson, J. R. FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5). New York: Academic Press, 1972.
- Anderson, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Anderson, J. R., & Bower, G. H. *Human associative memory*. Washington, D.C.: V. H. Winston, 1973.
- Bartlett, F. C. *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press, 1932.
- Bobrow, S., & Bower, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, 80, 455-461.
- Bower, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, 58, 496-510.
- Bower, G. H. *Comprehending and recalling stories*. American Psychological Association, Division 3, Presidential Address, Washington, D.C., Sept. 6, 1976.
- Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 1974, 103, 751-757.
- Bower, G. H., Karlin, M. B., & Dueck, A. Comprehension and memory for pictures. *Memory & Cognition*, 1975, 3, 216-220.
- Bransford, J. D., & Johnson, M. K. Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.
- 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.
- Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, 104, 268-294.
- 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.
- Fisher, R. P., & Craik, F. I. M. The interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 701-711.
- Goldstein, A. G., & Chance, J. E. Visual recognition memory for complex configurations. *Perception & Psychophysics*, 1971, 9, 237-241.
- Gould, A., & Stephenson, G. M. Some experiments relating to Bartlett's theory of remembering. *British Journal of Psychology*, 1967, 58, 39-49.
- Hayes, I. R. Personal communication, 1976.
- Hyde, T. S., & Jenkins, J. J. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 471-480.
- Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.
- Kintsch, W. Memory representations of text. In R. L. Solso (Ed.), *Information processing and cognition*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.
- Neisser, U. *Cognitive psychology*. Englewood Cliffs, N.J.: Prentice-Hall, 1967.
- Nelson, T. O. Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 151-171.
- Reder, L. M. The role of elaborations in memory for prose. *Cognitive Psychology*, in press.
- Reder, L. M. *The role of elaborations in the processing of prose*. Unpublished doctoral dissertation, University of Michigan, 1976.
- Schallert, D. L. Improving memory for prose: The relationship between depth of processing and context. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 621-632.
- Schank, R. C. *Conceptual information processing*. Amsterdam: North-Holland, 1975.
- Schank, R. C., & Abelson, R. P. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

- Spiro, R. J. Inferential reconstruction in memory for connected discourse. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 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 human story comprehension and memory*. Unpublished PhD dissertation, Stanford University, 1975.
- Wickelgren, W. A. The long and the short of memory. *Psychological Bulletin*, 1973, *80*, 425-438