Language Comprehension in Old Age

GILLIAN COHEN
University of Oxford

Three experiments examined the effects of aging on comprehension of spoken language. Integrative and constructive aspects of comprehension showed much more marked age-related deficits than registration of surface meaning. Experiment 1 showed that old subjects had difficulty in making inferences based on presented facts. Experiment 2 revealed a similar deficit in old people's ability to detect anomalies in newly presented information by reference to prior everyday knowledge. And Experiment 3, which tested story recall, showed that old subjects were less well able to extract and retain gist information than younger subjects. These difficulties are interpreted as reflecting a limitation in processing capacity such that the demands of concurrently registering surface meaning and simultaneously carrying out integrative and constructive processes exceed capacity in old age.

INTRODUCTION

It is not unusual in everyday life to notice that old people appear to have some difficulty in grasping what is said to them. Commonly, such difficulties are attributed to hearing loss or to lapses of attention. However, the present investigation has set out to test experimentally the hypothesis that there are age-related deficits at the higher level stages of language comprehension which can account for some of the difficulties apparent in informal conversation.

This hypothesis represents a challenge to the established view that language abilities are "crystallized" and show little or no effect of normal aging. This view is based mainly on the results of standard tests of verbal ability. While nonverbal performance tests show a fairly sharp decline with age, verbal ability test scores are maintained, or may even improve (Horn & Cattell, 1967). However, it is arguable that such tests, which are mostly concerned with definitions of single words or word pairs, yield very little insight into the vastly more complex processes involved in ordinary language functions such as the comprehension of discourse and written texts. Scarcely any research has so far been directed toward examining the effects of age on natural language processing of this kind. Geriatric psycholinguistics is virtually an unexplored territory. The

Address reprint requests to author at Department of Experimental Psychology, South Parks Road, Oxford, England. This study was supported by a research grant from the Social Science Research Council, U.K. Grateful thanks are due to Roger H. Freeman for much valuable assistance in preparing materials and analyzing data.

0010-0285/79/040412-18$05.00/0
Copyright © 1979 by Academic Press, Inc.
few studies that have been reported suggest that more careful investigation will reveal that language ability does deteriorate significantly in old age. Botwinnick and Storandt (1974) found that when vocabulary tests were scored qualitatively, taking into account the aptness, economy, and generality of the definitions offered, the quality of the responses did decline with age. Riegel (1968) detected age-related changes in word association responses indicating a weakened or unstable lexical organization, and Wetherick (1965) reported that old people had difficulty in working out the implications of a set of statements and, in a proverb interpretation task, tended to offer concrete illustrations rather than more abstract general explanations. More recently, Bromley (Note 1) studied the descriptive writing of older subjects and found that, compared with young subjects, they used shorter sentences at lower levels of syntactic complexity. And in speech perception, Bergman, Blumfield, Cascado, Dash, Levitt, and Margulies (1976) found marked deficits in the 60–69 age group when the speech input was distorted by increased rate, interruption, or reverberation. The age-related deterioration shown in these various tests strongly suggests that the popular belief that language ability is well preserved in old age is illfounded and overdue for revision.

Moreover, several aspects of mental performance which are indirectly related to the understanding of speech are well known to decline with age. One of the most marked changes in old age is a general slowing down of the rate of information processing (Birren, 1970). Continuous inputs of signals at fast rates are difficult for old people to process efficiently. Deficits in memory are especially noticeable when inputs have to be transformed, or reorganized, and in older people short-term memory is more vulnerable to interference than in younger subjects (Talland, 1968). Craik, Simon, and Moscovitch (Note 2) suggest that lack of processing resources limit depth of processing and so result in impoverished memory traces. Difficulties in the allocation and maintenance of attention have also been reported (Broadbent & Heron, 1962; Rabbitt & Birren, 1967). Understanding speech requires rapid processing of a continuous input, and memory for the earlier portions of the message must resist interference caused by intake of the later parts of the message. Attention must be sustained and divided between continuous monitoring of the input, and simultaneous analysis and integration of the successive elements, and full comprehension depends on depth of processing. Thus it is reasonable to hypothesize that these deficits in memory, attention, and speed of processing will be reflected in defective comprehension of spoken language.

Current models of comprehension (Clark & Clark, 1977; Bransford, Barclay, & Franks, 1972) stress the importance of constructive processes in comprehension. In addition to analysis and registration of the surface meaning of input sentences, new information must be integrated with
previously acquired knowledge and inferences generated. It seems probable that age-related deficits are most likely to be evident at these further stages of comprehension, just as observed deficits are more marked in short-term memory tasks when the incoming items have to be transformed or reorganized rather than simply registered and regurgitated.

Accordingly, the experimental tasks that were designed to test speech comprehension in this study focused on:

(a) the ability to relate together different facts presented within the same message and draw the correct inferences (Experiment 1);

(b) the ability to relate newly presented facts in a message to prior knowledge stored in memory (Experiment 2);

(c) the ability to preserve gist information and structure in the recall of a longer message or story (Experiment 3).

Obviously other tests of comprehension could have been selected but these were chosen as being representative of the kinds of ability required in everyday interchanges.

GENERAL METHOD

The method of cross-sectional comparison was employed, with two groups of old subjects being compared with two groups of young subjects. The composition and nature of the groups is described in detail below.

Group OHE: A group of 20 highly educated old people (OHE) had a mean age of 68 years (range 65–79). The group consisted of 11 females and 9 males. The subjects were retired professional people who had received university education or professional training and were in good health and leading active lives. Two subtests of the WAIS were administered to all subjects, the Vocabulary test and the auditory Digit Span test. Scores on these tests were used for both screening and matching purposes. The auditory Digit Span test served to establish that hearing was sufficiently unimpaired for the requirements of the experimental tasks and that the digit span was within the normal range (by reference to Wechsler's age norms). Scores on the Vocabulary test are known to be highly correlated with educational level and so serve as an additional index for matching groups by education. The mean scores of the OHE group were 69 on the Vocabulary test (maximum possible = 80), and 13 on the Digit Span test (combined forward and backward spans, maximum possible = 18).

Group YHE: A group of 20 highly educated young people (YHE) had a mean age of 24 (range 20–29), and this group contained 14 females and 6 males. These subjects were all students or graduate students. Their mean score on the Vocabulary test was 69, and on the Digit Span test 14.

The OHE and YHE groups are therefore well matched. The similarity of educational background ensures that cohort differences between the groups are minimal. As a result, differences between these groups on the experimental tasks can be attributed to age since the group comparison is relatively uncontaminated by other confounding factors. Since no such claim can be made for the comparison between the other two groups (described below) the main emphasis of this study rests on the comparison between the OHE and YHE groups.

Group OLE: A group of 20 old people of low educational level (OLE) was recruited from the day centre of a local geriatric hospital. The mean age was 79 years (range 70–95). There were 13 females and 7 males. These subjects had left school, on average at the age of 12 and followed manual occupations. Some were suffering from physical ill health and were receiving medication, but none of the group was, in the opinion of the clinicians, showing
evidence of senile deterioration or confusion, and had no detectable neurological damage. All were sensible and alert in conversation, and cooperated well in the experimental tasks. The mean score on the Digit Span test was 9.5, and on the Vocabulary test, 45. Both these scores are slightly above the average for this age group according to the published norms.

Group YLE: This group consisted of 20 young subjects of low educational level (YLE). The mean age was 24 (range 18–29). There were 11 females and 9 males. The subjects had left school at the minimal leaving age of 16 without further education and were either unemployed or in manual occupations. Mean Digit Span was 12, and Vocabulary score was 57.

Clearly the OLE and YLE groups are not well matched and differences between them could be attributed to health, education, or intelligence as well as age. Moreover, the mean age of the OLE group is considerably older than the OHE group. In consequence factorial comparisons of the four groups cannot be made. The results of the OLE group are nevertheless included here because the nature of the cognitive deficits exhibited by these subjects is of interest and of practical importance despite the fact that the causal origin of these deficits is obscure. To restrict the study of age-related effects to highly educated healthy old people yields much clearer experimental results, but gives an unrepresentative picture of old age.

**EXPERIMENT 1. MAKING INFERENCES**

The aims of this experiment were (1) to test the hypothesis that old people would show a greater deficit at the deeper level inference-making stage of comprehension, than at the stage of registration of the surface meaning; (2) to discover whether this deficit is a consequence of memory loss, that is, failure to remember the facts on which the inferences depend; and (3) to test whether the deficit might stem from a reduction in the rate of processing such that inference making is more defective at faster rates of input.

**Method**

**Materials.** Subjects listened to 16 short messages which were prerecorded in either a female voice or a male voice, and played on a Sony TC 95A cassette recorder. The messages recounted everyday events (going to work, posting letters, walking the dog, etc.). Eight of the messages were classified as Simple and eight as Complex. Complex messages contained more facts and were slightly longer (mean length of Simple messages was 60 words, and of Complex messages 75 words). The messages were spoken at two different rates, the Slow rate averaged 120 wpm and the Fast rate averaged 200 wpm.

**Design.** Each subject heard four messages in each of the following conditions, Slow Simple, Slow Complex, Fast Simple, and Fast Complex. Order of conditions and the allocation of messages within conditions was balanced across subjects. Successive messages alternated from male speaker to female speaker, and the allocation of messages to speakers was also balanced across subjects.

**Procedure.** Subjects were tested individually. They were instructed to listen carefully to each message so that they would be able to answer questions about it. At the beginning of the session a sample message was presented with the questions following and the correct answers supplied, so that the subject understood the task and the kind of questions which would be asked.

Following each message the subject was asked two questions relating to the content of the message. One of the questions was classified as a Verbatim question (V), which required reproduction of the presented facts (though not necessarily in the original wording). The
other was an Inference Question (I), which required an inference to be drawn from the presented facts. The order of I and V questions was varied. Sample messages and a sample question of each kind are given below.

Simple:

Mrs. Brown goes to sit in the park every afternoon if the weather is fine. She likes to watch the children playing, and she feeds the ducks with bread crusts. She enjoys the walk there and back. For the last three days it has been raining all the time although it's the middle of summer and the town is still full of people on holiday.

I question: Did Mrs. Brown go to the park yesterday? (no).
V question: What does Mrs. Brown give the ducks to eat? (bread crusts).

Complex:

Downstairs there are three rooms; the kitchen, the dining-room and the sitting-room. The sitting-room is in the front of the house and the kitchen and dining-room face onto the vegetable garden at the back of the house. The noise of the traffic is very disturbing in the front rooms. Mother is in the kitchen cooking and Grandfather is reading the paper in the sitting-room. The children are at school and won't be home till tea-time.

I question: Who is being disturbed by the traffic? (Grandfather).
V question: What is Mother doing? (cooking).

Responses were oral and the subjects were allowed as long as they wished to give their answers. Each answer was scored as 1 (correct), 0 (error), or 1/2 (partly right). In the complex example above, in answer to the I question "Grandfather" scores 1, "Mother" or "Don't know" scores 0, and "the person in the sitting-room" scores 1/2.

Results

The results are set out in Table 1, and Fig. 1 shows the mean percentage of I and V errors for all four groups.

An analysis of variance with arcsine transformation was carried out to compare the percentage of errors made by the OHE and YHE groups. The factors were Groups (OHE and YHE), Question type (I and V), Complexity (Simple and Complex), and Rate (Fast and Slow). Main effects of Groups \((F = (1,38) 13.6)\), Question type \((F = (1,38) 28.3)\), and Rate \((F = (1,38) 8.1)\) were all significant at \(p < .01\). Complexity was not significant. The interactions of Groups \(\times\) Question type \((F = (1,38) 5.2)\) and of Groups \(\times\) Rate \((F = (1,38) 4.8)\) were also significant at \(p < .05\). The triple interaction of Groups \(\times\) Question type \(\times\) Rate was also just significant \((F = 4.8, p < .05)\).

Post hoc comparisons of the means by the Scheffe test showed that old and young groups did not differ for V question errors (OHE = 9.5%, YHE = 6.5%) but did differ at \(p < .01\) for I question errors (OHE = 22.3%, YHE = 11.4%). The old group was significantly worse on I questions at the Fast Rate than at the Slow Rate (Fast I = 27.8%, Slow I = 17.5%), but
the young group did not show any significant rate effect (Fast I = 12.5%, Slow I = 10.6%). There was no effect of Rate on V question errors for either old or young groups.

A similar analysis of variance for the groups OLE and YLE yielded significant main effects of Groups ($F = (1,38) 84.8, p < .001$), and of Question Type ($F = (1,38) 37.3, p < .001$). The interaction of Groups $\times$ Question Type was also significant ($F = (1,38) 5.1, p < .05$). The factors of Rate and Complexity were not significant.

The effect of order of questions was not significant except in the OLE group where Last V questions were significantly worse than First V questions (by Wilcoxon: $T = 52, p < .05$). I questions were unaffected by position in this group, and in the other groups no position effects were evident.

<table>
<thead>
<tr>
<th>Groups</th>
<th>I questions (%)</th>
<th>V questions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>OHE</td>
<td>27.8</td>
<td>17.5</td>
</tr>
<tr>
<td>YHE</td>
<td>12.5</td>
<td>10.6</td>
</tr>
<tr>
<td>OLE</td>
<td>60.3</td>
<td>55.9</td>
</tr>
<tr>
<td>YLE</td>
<td>22.2</td>
<td>20.3</td>
</tr>
</tbody>
</table>
Similar comparisons were carried out for each group to examine the effect of speaker voice, but by the Wilcoxon test no group showed a significant difference between the male voice and the female voice.

Within both the old groups, OHE and OLE, Spearman's rank correlation tests were carried out to find whether the age of the subjects correlated with performance on I questions, or on V questions. In the OLE group the Age by I question errors correlation was significant ($r_s = .39, p < .05$). The correlation of age with V errors did not reach significance, and in the OHE group neither correlation was significant.

For each group the number of I errors made on each individual message was informally examined to see whether the order of difficulty was the same for each group. It was noted that the I questions which caused most difficulty were different for the old groups than for the young groups. Negative premises and exclusion clauses (not when, only if, unless) seemed to cause more problems for the older subjects, and messages concerning relative speed and time seemed especially difficult, but it was not possible to disentangle the effects of logical structure and semantic domain, so that these observations remain to be investigated more carefully in a future study. In line with theoretical models assuming that negative premises suggests that a processing overload is causing the fact that older subjects find more difficulty in drawing inferences from negative premises suggests that a processing overload is causing the deficit.

Discussion

The results of Experiment 1 confirm that there is an age-related deficit in the inferential stage of comprehension. In the OHE vs YHE comparison, where the difference can reliably be attributed to age, no deficit was found for the verbatim questions which required only reproduction of the surface meaning, but inference making was clearly affected. The absence of any age effect on V questions, together with the absence of any effect of message complexity and of position of question, all suggest that the inference deficit is not due to memory loss. On the other hand, the fact that the OHE groups showed a larger inference deficit at the fast rate of input indicates that it is related to a reduction in the speed of processing. The absence of a significant rate effect in the OLE group is puzzling, but is probably due to the fact that scores are near floor level for I questions.

In the comparison of the low education groups, OLE and YLE, the age difference is, as already stated, contaminated by other factors, but the significant correlation of age and I question errors in the OLE group does suggest that age is playing a part in the deterioration of inferential comprehension.

Comparison of the two young groups shows a clear effect of educational
level on the task. However, Fig. 2 shows how the size of the I-V difference increases with age to a very similar extent in both high education and low education groups (7.9 and 9% increases, respectively). That is, the effects of aging appear to operate quite uniformly on different levels of initial ability.

The difficulty that old people evidently experience in making inferences in this task does not necessarily indicate a loss of inferential reasoning ability per se, but could be taken as showing that it is a vulnerable stage of comprehension which tends to drop out when total processing demands exceed capacity. It is possible that with written messages and unlimited study time old subjects would prove to be able to make inferences successfully, and a wider range of tests of inference making ability are currently being planned to clarify this. Nevertheless the present study carries with it the practical implication that in conveying spoken information to old people relevant facts should be explicitly stated and not left for the listener to infer.

**EXPERIMENT 1B**

Experiment 1B was devised as a follow-up to Experiment 1 with the aim of testing the possibility that the I question deficit exhibited by older subjects might originate at a perceptual level. That is, old people might suffer a minor reduction in auditory acuity, or an increase in neural noise making the message more difficult to perceive, and so diverting processing capacity from the deeper level inferential stage of comprehension to the surface level. To test this a group of six young subjects from the YHE group was recalled, and listened to a new set of 16 spoken messages.

![Fig. 2. The I-V difference for each group (difference in mean percentage errors for V questions and for I questions).](image-url)
presented over headphones mixed with white noise. Two different levels of noise were employed, a low level (approximately 75 db) and a higher level (approximately 85 db). All the subjects received half the messages accompanied by Low Noise and half accompanied by High Noise. At the High Noise level the subjects reported that the noise was distracting, and that it was difficult, though still possible, to recognize the speech.

The results are shown in Table 2. The predicted increase in I question errors was not found. All the subjects showed an improvement over their performance in the No Noise conditions of Experiment 1 for V and I questions, and the I-V difference was reduced. Although this paradigm provides only a somewhat crude attempt to simulate any hearing difficulties that might be present in old age, it does suggest that the poor performance on I questions shown by the old groups is unlikely to be due to perceptual difficulties, and can more probably be attributed to cognitive overload.

**EXPERIMENT 2. DETECTING ANOMALIES**

The experiment tested aspects of comprehension which require that newly presented information be related to prior knowledge, that is, to information already stored in memory. An anomaly detection task was used. The subjects listened to short spoken messages and were asked to judge whether each of these contained a mistake, or a statement which could not be true. In one example the statement that the Jones family was very close to an airport was followed later by the statement that it was very quiet and peaceful. In further examples a blind man read a newspaper; a girl with a broken arm in plaster went swimming; a housewife who had run out of bread made sandwiches.

To detect these anomalies the listener must access simple everyday facts stored in memory (that airports are noisy, blind men can’t read newspapers, etc.) and note the discrepancy between this prior knowledge and the new information presented in the message. Haviland and Clark (1974) studied the integration of Given (old) information and New information, and noted that sentences requiring the listener to generate bridg-
ing assumptions, that is to retrieve Given information in order to make sense of the New information, are harder to understand. Analogously, it was hypothesized that anomaly detection which requires input to be concurrently registered and matched against information in memory would be liable to show age-related deficits.

**Method**

The subjects tested were the same as in Experiment 1.

**Materials.** 16 messages were constructed, the mean message length being 75 words. There were two different versions of each message, one containing an anomaly, and the other making good sense. Messages were spoken at a medium rate (approximately 150 wpm.). Half the messages were recorded by a male speaker and half by a female speaker. Two taped sequences were constructed. Each consisted of 16 messages, of which 8 were anomalous and 8 were normal (i.e., good sense), and these were randomly mixed in the sequence. Messages which occurred in an anomalous form in one taped series occurred in a normal version in the other series.

**Procedure.** Subjects were tested individually. They were instructed to listen carefully to each message and to say whether they judged it “Right” (i.e., they thought it made good sense) or “Wrong” (i.e., they thought it contained a mistake or something which couldn’t be true). A sample of an anomalous message was presented before testing began, and the nature of the anomaly was explained. Half the subjects in each group listened to one taped series and half listened to the other. After each message the subject’s response was noted, and if the message was judged anomalous the subject was asked to explain why it was wrong, what mistake it contained.

**Scoring.** Answers were scored as correct when normal messages were judged “Right,” and when anomalous messages were judged “Wrong” and the anomaly was correctly identified. Errors were divided into three categories.

1. **Misses:** The anomaly was not detected and the anomalous message was judged Right.
2. **False Alarms:** Normal messages were judged Wrong and nonexistent anomalies reported.
3. **False Hits:** Anomalous messages were judged Wrong but the real anomaly was missed and some other aspect of the message was reported as anomalous.

Errors of Type 2 (False Alarms) and Type 3 (False Hits) both yield mistaken explanations and it was possible to identify three different kinds of mistaken explanations.

1. **Mistakes of Fact:** False Alarms resulted when the facts in the message were registered incorrectly. In one example where the normal version stated “Mary had lost weight—her dress was too big” the subject judged the message Wrong, saying “Mary had gained weight so the dress would be too small.”
2. **Mistakes of Interpretation:** False Hits resulted when prior knowledge was incorrect. For example, subjects missed the anomaly of describing a house near an airport as very quiet and peaceful, and judged the message Wrong on the (erroneous) grounds that “nobody lives near an airport.”
3. **Value Judgments:** Messages were judged Wrong on moral grounds (which are strictly, irrelevant to the task) rather than on semantic ones. For example, the message about the housewife making sandwiches without bread was judged wrong because “she should have a hot lunch not sandwiches,” and the girl who went swimming with a broken arm in plaster “should have been more careful not to break her arm.”

The mistaken explanations were divided into these three categories.
Table 3 shows the mean numbers of errors for each group and the means for each type of error. Mann-Whitney tests were used to compare the mean numbers of errors. The OHE and YHE groups differed significantly ($U = 123, p < .025$), and for the OLE vs YLE comparison $U = 19, p < .001$. There was no difference between groups in the distribution of errors into the three different categories. Each group made more Misses than False Alarms, and more False Alarms than False Hits.

Table 4 shows the numbers of mistaken explanations falling into the three different categories for each group. Group differences in the frequency of Mistakes of Fact, Mistakes of Interpretation, and Value Judgments were tested by $X^2$. In the comparison of OHE vs YHE groups there was no significant difference ($X^2 = 1.16, ns$), although the OHE group did make slightly more Mistakes of Interpretation. Comparison of the OLE and YLE groups yielded $X^2 = 12.45, df = 2, p < .01$.
Discussion

Both the old groups made more errors on this task than did the young groups with which they were compared. The finding that the proportion of errors classed as Misses is similar for each group indicated that the old subjects were not employing a laxer criterion for what makes sense. There are various ways in which subjects may fail to detect anomalies in this task. The new information in the message may not be registered, or may be forgotten; the prior knowledge may be incorrect or inaccessible; the operation of matching new information to prior knowledge may not be carried out. The mistaken explanations classed as factual indicate that the new information has not been registered correctly. OHE and YHE groups did not differ in the number of factual mistakes.

Mistaken explanations classed as interpretive reveal that incorrect prior knowledge has been accessed resulting in detection of a false anomaly. In the OHE group this type of error occurred slightly, but not significantly, more often. In the OLE group both factual and interpretive errors were more frequent than in the YLE group, but the most striking aspect of OLE performance was the large number of value judgments which were based on irrelevant moral grounds rather than on semantic or logical considerations. It was not the case that OLE subjects failed to understand the task, or were totally unable to make semantic judgments since all of them detected some semantic anomalies correctly. Rather it seems as if personal values are more salient than semantic coherence for this group. This tendency may reflect the growing egocentricity of the aged personality increasingly focused on personal concerns rather than on the external world as described by Welford (Note 4). Factors such as a low level of education and a mentally undemanding way of life probably contribute to this effect.

EXPERIMENT 3. STORY RECALL

Story recall has been extensively studied by, among others, Rumelhart (1977) and Van Dijk and Kintsch (1977). Although recall of stories is obviously rather seldomly required in everyday life, it does test abilities which are important aspects of comprehension. In particular, it tests the ability to perceive the logical and structural relationships between different parts of the story and the ability to select and retain the gist information, the most important elements in the story. In the recall of fairly large amounts of information, such as a story, memory is overloaded, and it is arguable that the best strategy is to maximize recall of the gist at the expense of the less important details. It is therefore of interest to see how far subjects will spontaneously adopt this strategy without being specifically instructed to summarize. If old people have difficulty in language
comprehension tasks which require simultaneous registration and reorganization, they might be expected to show deficits in story recall.

**Method**

The subjects were the same as in Experiments 1 and 2.

**Materials.** A story was constructed which was just over 300 words in length. The structure and content were rather similar to the Circle Island story used by Dawes (1966). The first part (the setting) described the characteristics and goals of two tribes living on an island. The middle section related a quarrel which broke out between them (the events), and the final section described the outcome of the quarrel (the reactions). Both the language and the content of the story were designed to be clear and simple, so that there was unlikely to be any difficulty in understanding it. The story was prerecorded at a fairly slow speech rate (approximately 120 wpm) and presented on the Sony Cassette Recorder.

Subjects were tested individually. They were instructed to listen carefully and try to remember as much as possible as they would be asked to retell the story as fully as possible. Immediately following the presentation of the story, oral reproduction began. Reproductions were taped and later transcribed. When a subject completed his reproduction he was encouraged to try to recall anything more. Testing terminated when he was satisfied he could recall no more.

**Scoring.** The story was analyzed into 48 constituent facts or propositions. Although, of course, more refined techniques of propositional analysis could have been employed, this simple decomposition into facts was considered an adequate basis for discriminating between the performance of the different groups. Transcripts were scored for the total number of propositions correctly reproduced. Changes of wording which preserved the sense were scored as correct (e.g., "the island was governed by a council" reproduced as "a committee ruled the island" was accepted). A consensus of the opinions of three judges was used to isolate the six facts, called summary propositions, which represented the gist of the story, and reproductions were also scored for the number of summary propositions they contained. The summary propositions were: (1) Two tribes lived on an island. (2) The tribes had different ways of life. (3) The Dooni tribe had most power on the governing council. (4) The tribes disagreed about a plan to build a harbor. (5) The Dooni got their own way. (6) Fighting between the tribes broke out. Finally, a class of elements in the story termed modifiers were identified. There were 24 modifiers consisting of comparatives and quantifiers (more, a few), temporal modifiers (recently, soon, no longer), locatives (there, in the north), and logical connectives (because, therefore, in order to). The number of these preserved in the reproductions was also scored. Modifiers were scored separately because reproductions often preserved propositions but omitted the modifiers which convey the relations between propositions. So, for example, a subject would recall that Event X occurred and that Event Y occurred, but omit the connective "because" so that the causal relationship between the events was lost.

**Results**

The results are shown in Table 5. OHE and YHE groups were compared by the Mann-Whitney test on all three measures. OHE was significantly poorer for Total Propositions ($U = 85, p < .01$); for Summary Propositions ($U = 38, p < .001$); and for Modifiers ($U = 54, p < .01$). Also by the Mann-Whitney test, the OLE group was significantly poorer than YLE on all three scores (Total Propositions, $U = 0$, Summary Propositions, $U = 3$, Modifiers, $U = 0$, all with $p < .01$).
TABLE 5
Percentage Correct Recall for Total Propositions, Summary Propositions, and Modifiers in the Story Recall Task (Experiment 3)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Total Propositions (%)</th>
<th>Summary Propositions (%)</th>
<th>Modifiers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHE</td>
<td>44.3</td>
<td>76.25</td>
<td>24</td>
</tr>
<tr>
<td>YHE</td>
<td>60.4</td>
<td>95</td>
<td>41</td>
</tr>
<tr>
<td>OLE</td>
<td>7.6</td>
<td>27.3</td>
<td>2.3</td>
</tr>
<tr>
<td>YLE</td>
<td>49.2</td>
<td>83.3</td>
<td>27.2</td>
</tr>
</tbody>
</table>

$X^2$ tests were used to see whether the groups differed in the proportion of total propositions to summary propositions which were recalled. The comparison of OHE and YHE yielded $X^2 = 3.7, df 1$, which is just significant ($p < .05$, one-tailed). For OLE vs YLE, $X^2$ was 17.1, $df 1, p < .001$.

Transcripts were also examined for errors, that is misrepresentations of the facts of the story. These errors of commission were few compared with the large numbers of errors of omission. Some of the errors were classified as Errors of Anaphoric Reference. These consisted of either attributing characteristics or actions to the wrong protagonists, or failing to make any clear reference, as in the use of pronouns without any preceding proper names or descriptions. The class labeled Other Errors included confabulations and substitutions (e.g., "gold mines" for "copper mines"). The numbers of these types of error made by each group are shown in Table 6. By $X^2$ the distribution of error types was compared for OHE and YHE ($X^2 = 3.8, p < .05$), and for OLE and YLE $X^2$ was 17.05, $p < .001$.

**Discussion**

Both old groups performed less well than the young groups in the story recall task. The conclusion that this is an age-related deficit emerges

TABLE 6
Numbers of Errors of Different Kinds in the Story Reproductions of Each Group (Experiment 3)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Errors of anaphoric reference</th>
<th>Other errors</th>
<th>Total errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHE</td>
<td>15</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>YHE</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>OLE</td>
<td>32</td>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>YLE</td>
<td>3</td>
<td>11</td>
<td>14</td>
</tr>
</tbody>
</table>
clearly since (a) the OHE group is worse than the matched YHE group, and (b) the OHE group is also somewhat poorer than the less well-educated YLE group. It appears that when memory is heavily overloaded, as in story recall, age-related deficits appear more markedly than they do in tasks where the memory load is less severe. This is evident by comparison with Experiment 1, where the OHE group showed no deficit in verbatim recall of facts from the much shorter messages presented in that task, although a more systematic investigation of the relationships between length of text and the extent of the age-related deficit would be necessary to show precisely how overload affects memory in old age. It is possible that old subjects are less practiced than young people in remembering large amounts of information and might improve with further practice.

Old subjects were also less well able to concentrate their recall on the gist information in the story. In the YHE group 10 out of 20 subjects recalled all 6 of the summary propositions, and in the YLE group 7 subjects did so. In the OHE group only 3 subjects achieved this. In the recall of the OLE group the gist information was almost totally absent, and the fact that their score for summary propositions exceeds the score for total propositions is due to serial position effects, only a few of the summary propositions occurring at the beginning and end of the story being retained. It might be expected that if old people suffer from more severe limitations of memory capacity that they would focus on gist information at the expense of non-gist information, but clearly this is not the case. The loss of modifiers tends to weaken the logical and temporal structure of the story, and typically these were replaced by simple conjunctions (e.g., "therefore," "after that" being replaced by "and").

The preponderance of errors of anaphoric reference made by old subjects was an unanticipated but striking aspect of the results and can probably be attributed to limitations of running memory span. Where names and corresponding pronouns, or referents and predicates, are separated in the text by intervening words correct linkage is dependent on running memory which is known to be impaired in old age.

**GENERAL DISCUSSION**

Taken together the results of these three experiments indicate that in old age comprehension of spoken language is handicapped by diminished ability to perform simultaneously the task of registering the surface meaning and also carrying out further processes involving integration, construction, or reorganization of different elements of the meaning. In general, surface comprehension is maintained at the expense of the integrative processes. In Experiment 1 this produced a deficit for inference questions which require integration of different facts in the message. The
process of constructing inferences based on the information presented which normally occurs spontaneously during comprehension (Bransford et al., 1972) is omitted or inefficiently performed. In Experiment 2 where the task required integration of newly presented material and prior knowledge, a similar age-related deficit was observed, and the nature of the errors suggested that the difficulty lay in retrieving the relevant prior knowledge for matching against the current input. Older subjects were more prone to access irrelevant prior knowledge. In story recall, older subjects were less likely to extract and preserve those parts of the study carrying gist information and showed no ability to compensate for overall lowered memory capacity by concentrating resources on the core structure of the study. It is, of course, not possible to determine from these results whether old subjects were unable to identify gist information, or whether they did not choose to give it priority for recall.

It is possible to discern some common factor in all these deficits if they are attributed to diminished processing resources. All the tasks demand concurrent processing of surface meaning and deeper level processing of underlying meaning and this dual demand appears to exceed the capacity of older subjects analogously to the way in which, while normal memory span may be unaffected by aging, running span and backward recall prove too demanding.

While the comparisons of the well-matched OHE and YHE groups give reasonably clear evidence of age-related differences, the much poorer performance of the OLE group is likely to have multiple causes of which age is only one. Clearly the tasks were sensitive to educational level since the YLE group performed less well than the YHE group. It has been suggested that effects of aging on mental function may be offset to some extent by a high level of education and continued mental activity, and that they are accelerated by low education and mental inactivity. The physical ill health of the OLE group was no doubt also a major factor in lowering their performance. Nevertheless, despite the difficulty in identifying the causal origin of deficits in this group, it is of interest to compare informally the pattern of deficits in the two old groups. For both groups the inferential stages of comprehension are the most vulnerable; both show difficulties with anaphoric reference. The decline in gist recall and the tendency to focus on moral anomalies rather than semantic anomalies is more marked in the older and more deteriorated OLE group.

This study goes some way toward indicating which components of the process of comprehension are most liable to be affected by mental deterioration, and more speculatively, it might be suggested that those aspects of comprehension which, ontogenetically, are acquired later are the first to be lost. Language may be one of those cognitive abilities for which life span comparisons reveal a U-shaped curvilinear function with chil-
dren and old people showing a similar deficiency compared with young and middle-aged adults (Comalli, 1970). In practical terms, the findings indicate that information should be presented to old people in such a way that relevant facts are made completely explicit and not left for the listener to infer; key facts or gist need to be stressed and not left for the listener to extract himself; and speech should be slow and messages short.

REFERENCES


REFERENCE NOTES

1. Bromley, D. R. Age differences in the quantity and quality of natural language in written


(Accepted March 12, 1979)