Recall and recognition memory deficits in depression

A.N. Brand a, J. Jolles b and C. Gispen-de Wied c

a Department of Clinical and Health Psychology, State University of Utrecht, Utrecht, The Netherlands. b Department of Neuropsychology and Psychobiology, State University of Limburg, Maastricht, The Netherlands and c Department of Biological Psychiatry, Academic Hospital, State University of Utrecht, Utrecht, The Netherlands

(Received 11 October 1991)
(Revision received 29 January 1992)
(Accepted 4 February 1992)

Summary

The aim of the present study was to establish the nature of memory deficits of depressive subjects in word learning tests. A word learning test consisting of 1, 3 or 5 learning trials was used. We found that patients were characterized by inferior memory recall compared to controls when 5 learning trials were given. Patients performed significantly slower than controls on a recognition test but both patients and controls recognized the same number of words. This suggests that the memory deficits that are present in many depressive subjects may be restricted to impaired active retrieval from memory. A second experiment revealed that recognition memory and delayed recall as well as immediate recall were impaired in depressive patients after 1 learning trial. These shortcomings vanished after 3 trials, except for immediate recall. These data suggest that not only retrieval but also encoding of information into memory may be impaired in depression, especially in the beginning of a task when demands on cognitive effort are high. The results are discussed in terms of resource allocation and demands on effort that may change in the course of a task.

Key words: Memory; Depression; Recall; Recognition

Introduction

Much research is devoted to the assessment of memory deficits in depression. Many studies have found that depressive patients have a decreased performance in memory and learning tests (Henry et al., 1973; Miller, 1975; Sternberg and Jarvik, 1976; Cutting, 1979; Glass et al., 1981; Cohen et al., 1982; Koh and Wolpert, 1983; Roy-Byrne et al., 1986; Brand and Jolles, 1987), whereas others have not (Miller and Lewis, 1977; Davis and Unruh, 1980; Donnelly et al., 1982; Dunbar and Lishman, 1984; Coughlan and Hollows, 1984). In addition, some studies question the relative importance of the memory deficits compared to the
affective and motivational symptoms, or attribute the memory deficits to methodological artefacts (Friedman, 1964; Miller, 1975; Kopelman, 1986).

The disagreement about the question whether or not depression is characterized by memory deficits may be a consequence of differences in patient selection, definitions of memory, and sensitivity of research methods (Glass et al., 1981; Cohen et al., 1982). Many of the methods currently used in clinical memory testing are based on weak or outdated theoretical assumptions and concepts (see Erickson and Scott, 1977; Russell, 1981; Mayes, 1986; Kopelman, 1986). The need for more reliable and sensitive memory tests to differentiate between different memory processes has frequently been expressed (Erickson and Scott, 1977; Jolles, 1985; Hayes, 1986; Brand, 1987).

Methods based on information processing concepts have relevance in this respect (e.g. Byrne, 1976; Koh and Wolpert, 1983; Brand and Jolles 1987). Brand and Jolles (1987) used two clinically applicable versions of Sternberg’s Memory Scanning Task (Sternberg, 1975; Brand, 1987) with depressives and other psychiatric patients, and found that depressive patients had impaired memory scanning and inefficient search strategies. As this task focuses on aspects of short-term memory, using letters as stimuli, it is not yet clear how these impairments are related to other memory processes such as retrieval from short-term memory (STM) and long-term memory (LTM) and encoding of relevant material into memory.

The present study was designed to investigate the latter aspects of memory. Earlier findings have shown that not only accuracy of recall and recognition are important measures of these aspects, but also the speed of these measures, as this gives more insight into the efficiency of these processes (Brand and Jolles, 1985, 1987; Brand, 1987). We therefore evaluated the contribution of both accuracy and speed to the study of memory deficits in depression.

The paradigm used was a multi-trial free recall test involving a list of 15 words. This serial learning test is adapted from Rey’s Auditory Verbal Learning Test (RAVLT; Lezak, 1983; Mayes, 1986) in that the stimuli are presented visually. Visual presentation has some theoretical and practical advantages over auditory presentation while equivalence of outcomes is maintained (Brand and Jolles, 1985). Interest in the test is growing in clinical research settings because it allows for differential assessment of different patient groups (e.g. Damasio et al., 1985; Butters et al., 1986; Blau and Ober, 1988), and shows promising results in studies into the relationship of health risk factors and aging (Houx et al., 1989). Our adaptation of the test measures a number of relevant memory aspects that are not covered by similar memory tests such as immediate recall which is in particular a measure of retrieval from STM; delayed recall as a measure of retrieval from LTM; delayed recognition (encoding into memory); recognition speed (efficiency of retrieval); omissions and repetition errors (efficiency of memory and of self-monitoring or evaluation). In depression, recall memory seems to be more impaired than recognition memory (Calev and Erwin, 1985; Blau and Ober, 1988). This could indicate that depressive patients have retrieval problems but no encoding deficit. One theoretical framework that could account for the deficits in depressive patients in different tasks is the resource allocation model (Ellis and Asbrook, 1988; 1989). This model predicts that cognitive impairments in depression are most evident in tasks that require cognitive effort. The empirical evidence for this model has largely come from studies in which mood induction procedures were used with normal, healthy subjects. Nevertheless, the results of these studies parallel clinical findings, including patients who suffer from affective disorders (Roy-Byrne et al., 1986; Cohen et al., 1982).

It has also been suggested that free recall and recognition make different demands on a subject's resources, i.e., in the amount of effort needed. Free recall might demand more effortful, controlled processing, whereas recognition could rely more on passive, automatic detection.

The present study started out with the hypothesis that the inferior memory performance of depressive patients is restricted to (effortful) retrieval processes, whereas encoding of information into memory is unimpaired. This hypothesis was tested by administering the visual verbal learning test to a sample of depressive patients...
and matched controls and measuring immediate and delayed free recall and delayed recognition.

**Experiment 1**

In Experiment 1 we assessed whether depressive patients are characterized by deficient recall and normal recognition. A verbal learning test was used, with 15 words presented visually for free recall and measurement of accuracy and speed in a final recognition test. Our expectation was that depressive patients would be impaired with respect to free recall and speed measures, but that recognition accuracy would not differ from healthy performance.

**Method**

**Subjects**

Twenty-four depressive patients participated in the experiment. They were diagnosed by using DSM-III-R criteria (APA, 1987). Five of the subjects were out-patients. All patients had a primary depression, i.e., a depression without an organic or physical origin (see Table 1 for further subject information). No patients were psychotic. The education level, as judged on a 7-point scale (Werhage, 1964), was 4.8 ± 1.2. Three patients were males and only one was left-handed. The patients were tested within 5 days after hospital entrance. In these 5 days none of the patients received any relevant drug or other treatment.

**Control subjects**

Five male and 21 female local citizens served as a control group. They had responded to a local newspaper advertisement and received a small payment (Dfl. 10) for their participation. The control subjects were matched to the patients with regard to age and level of education; these variables did not differ significantly between the groups (t-test P-values of 0.67 and 0.39, respectively).

**Materials**

The word-learning task used included a learning list consisting of 15 meaningful monosyllabic words. The words have a frequency of 20 to 400 per million in Dutch and refer to concrete objects such as rock, bike, roof, etc. The recognition list consisted of the 15 (target) words from the learning list and 15 new but similar (distractor) words drawn from the same population. The 30 words had a semi-random distribution in that no more than 3 target or distractor words occurred in succession. The learning and recognition lists were presented visually rather than auditorily (which is the more usual procedure) because this allows better experimental control of stimulus/response events and better perception of the stimuli. (Band and Jolles, 1985). An Apple IIe microcomputer was used for presentation. The words were presented in the same order for all subjects. The words were displayed in capital letters (10 mm high and 7 mm wide), in white against a dark background in the center of the computer screen.

**Procedure**

The subject faced the screen at a distance of 80 cm. The task involved five immediate recall trials, followed by a 20-min delay period (filled with non-verbal tests, see below), a delayed recall trial, and a recognition test.

The instructions were read aloud by the experimenter. The subject watched the words that were presented one at a time. Stimulus presentation and interstimulus interval were both 1 s. The subject was asked to recall as many words as possible at the end of each presentation of the list. There was no restriction as to the order of recall. The first trial was followed by four trials
with words presented in the same order. The subject was requested not to ask the experimenter whether he/she had reported a word previously, although no instruction was given as to whether it was allowed to repeat words already reported.

After the fifth trial, three nonverbal tasks were given (Rey's Complex Figure Test; Road Map Test; Trail Making Test; see Lezak, 1983, for references). After 20 min the subject was asked to recall the words. This delayed recall trial was followed immediately by the recognition test involving yes/no recognition of the 15 words presented in the learning phase together with 15 different distracting words. This time the words were presented in a self-paced fashion, with a 1-s response-stimulus interval (RSI). The responses were given by way of thumb key paddles ('yes' by the preferred hand and 'no' by the other).

--- depr  contr

![Graph of percentage correct over trials](image1)

**Results**

Fig. 1 presents the mean percentage of words recalled correctly by the patients and the controls in the five immediate recall trials and in the delayed recall trial. The percentage of words recognized in the recognition test is also given. The number of words recalled in the five immediate recall trials was analyzed with a 2 (Group) × 5 (Trial) ANOVA, with Trial as a repeated measures factor. There was a significant Group main effect (F(1,48) = 7.55, P = 0.008), and Trial main effect (F(4,192) = 117.2, P < 0.0001). The Group × Trial interaction was also significant (F(4,192) = 4.05, P = 0.003). Post hoc t-tests showed that the depressive patients recalled significantly fewer words than the controls in trials 1, 2 and 4. The P values for trial 1 to trial 5 were < 0.001; 0.003; 0.16; 0.013 and 0.23, respectively.

Analysis of the delayed recall scores did not yield a significant Group effect (F < 1). Likewise, the groups did not differ in the mean number of...

![Graph of recognition time in msec](image2)
wrong or new words summed across the five immediate recall trials (0.6 for the depressive patients and 1.0 for the controls), or in the number of repetition errors (4.3 for the patients and 4.9 for the controls).

The number of hits and correct rejections in the recognition test was analyzed in a 2 (Group) × 2 (Response Type) ANOVA. There were no significant main effects nor interaction effects. Fig. 2 shows the mean recognition times for the hits and correct rejections in the two groups. These data were analyzed in a similar design. Data from the first 2 of the 30 recognition trials were excluded, as these trials were considered as warm-up trials. The depressive patients were significantly slower than the controls (F(1,48) = 6.44, \( P = 0.013 \)). There was no Response Type effect nor was there an interaction. Frequency analysis of all latencies of the two groups revealed that the distribution for the patient group was more skewed to the right suggesting that patients took longer to recognize the words.

**Discussion**

The patients recalled fewer words than the control subjects. This was observed in the immediate recall trials, but not in delayed recall. More specifically, the performance of the patients in trials 1, 2 and 4 was inferior to that of the controls. Their learning performance, as defined as the percent increment from trial 1 to trial 5, was the same as that of the controls, indicating that the performance of the patients was lower than that of the controls at the start and reached an asymptotic level that was also lower. The recognition scores of the groups were the same, although there were individual patients with more false positive and false negative errors than is seen normally. This was particularly true for patients with a diagnosis of major depression.

Free recall is primarily regarded as a measure of retrieval of encoded material (Butters et al., 1986; Calev and Erwin, 1985). However, low recall scores do not answer the question of whether material is actually encoded into memory. This can be assessed with a subsequent recognition test. In view of the patients' inferior performance on immediate recall but not on recognition, it seems that many patients are characterized by a primary deficit in retrieval, at least in the initial phases of learning, whereas encoding into memory, as measured by recognition scores, is normal. These findings are in line with our working hypothesis and with findings in the literature (Calev and Erwin, 1985; Roy-Byrne et al., 1986; Blau and Ober, 1988). The significantly slower recognition speed of the depressive patients compared to that of the controls might provide additional evidence for the existence of retrieval deficits in depression, although other mechanisms such as motor processes add to this speed measure.

A second experiment was needed to determine whether the failure to obtain differences between the recognition scores of the groups could have been attributed to a ceiling effect. Was the yes-no test of 15 words and 15 distracters after five learning trials too easy for many of the subjects? It is possible that the demands made on cognitive effort in a recognition task after five learning trials are too small to differentiate between subjects. Resource allocation models suggest that depressed subjects show greater performance decrements when processing more demanding tasks (Ellis and Ashbrook, 1988). Depressed subjects may lack (cognitive) effort, and this may have its greatest impact on the active retrieval processes that rely on the allocation of effort. A recognition test after five learning trials may not require much effort so that it can be accomplished by even seriously ill patients (see also Ellis and Ashbrook (1988) for the notion that different tasks may differ in the demands placed on the subject). We therefore made the recognition task more difficult.

**Experiment 2**

A second experiment was performed with the aim of assessing whether recognition memory is impaired in depressive patients when different demands, in terms of effort, are made on them. As it is possible that the recognition response reaches a near maximum after five presentation trials, we administered the recognition test earlier in the learning sequence, i.e., after one learning trial and after three learning trials, in a between subjects design. In this way we hoped to assess
TABLE 2
Subject characteristics in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>1-trial condition (A) controls</th>
<th>1-trial condition (A) patients</th>
<th>3-trial condition (B) controls</th>
<th>3-trial condition (B) patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Sex</td>
<td>2M, 6F</td>
<td>2M, 6F</td>
<td>4M, 6F</td>
<td>3M, 4F</td>
</tr>
<tr>
<td>Mean age</td>
<td>42.1 ± 15.0</td>
<td>47.6 ± 14.3</td>
<td>45.8 ± 11.9</td>
<td>46.0 ± 13.2</td>
</tr>
<tr>
<td>Age range</td>
<td>29 - 75</td>
<td>28 - 69</td>
<td>26 - 70</td>
<td>27 - 62</td>
</tr>
<tr>
<td>Educ level</td>
<td>M ± 1 SD</td>
<td>M ± 1 SD</td>
<td>M ± 1 SD</td>
<td>M ± 1 SD</td>
</tr>
<tr>
<td>HDRS</td>
<td>4.1 ± 1.1</td>
<td>4.3 ± 1.2</td>
<td>5.0 ± 0.5</td>
<td>5.1 ± 1.1</td>
</tr>
<tr>
<td>DSM-III-R classification:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>296.32 (major dep without melanch)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>296.33 (major dep with melanch)</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>300.40 (dysthmic disorder)</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>311.00 (depression not otherwise specified)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

whether encoding processes (expressed in recognition accuracy) in the earlier phases of learning are affected in addition to retrieval. We expected that both encoding and retrieval would be hampered when the opportunity for learning is restricted, especially in depressive patients.

Method

New groups of subjects consisting of 15 depressive patients and 18 control subjects participated in the present experiment. Information about the subjects is presented in Table 2. The patients were tested within one week after admission to the clinic and were drug-free at the time of testing. Four of the patients were out-patients. The patients were randomly assigned to one of two experimental conditions (A: one-trial condition: N = 8; B: three-trial condition: N = 7). These patient subgroups did not differ in age (t(13) = 0.23), education level (t(13) = 1.31, P = 0.21), or on the Hamilton Depression score (t(7) = 0.39).

The control subjects were healthy volunteers who were recruited and paid in the same way as the controls in Experiment 1. They were similarly assigned to the conditions (A: N = 8; B: N = 10). These subgroups did not differ in age and educa-

![Fig. 3. Delayed recall and delayed recognition scores with standard deviations of two groups of depressive patients (white bars) and control groups (darkened bars) in a one-trial and three-trial condition (hits and correct rejections in the recognition test are combined).](image-url)
Fig. 4. Mean recognition times for positive (hits) and negative responses (correct rejections) made by depressive patients (open circles) and controls (filled circles) after a one-trial (solid lines) or a three-trials (dashed lines) learning condition.

In order to obtain a better understanding of group differences in the two conditions, separate analyses were run for each condition.

**One-trial condition**

Separate analyses of the free recall data in the one-trial condition yielded comparable figures as mentioned above: both immediate recall and delayed recall were impaired in the patients.

The recognition data in this condition were analyzed in a 2 (Group) × 2 (Response Type) ANOVA with repeated measures on the latter factor. The patients gave fewer correct responses than the controls (F(1,14) = 23.1, P < 0.001). This was true for both hits and correct rejections (no Response Type effect or interaction). The speed of recognition of the patients was also slightly lower than that of the controls (F(1,14) = 4.7, P = 0.049), and the latencies for the correct rejections were longer than those for the hits (F(1,14) = 18.6, P < 0.001). The interaction was just not significant (F(1,14) = 4.34, P = 0.056).

**Three-trial condition**

In the three-trial condition the recall data were analyzed in a 2 (Group, patients vs controls) × 3 (Trials) ANOVA with repeated measures on the last factor. There was a small but significant Group effect (F(1,15) = 3.9, P = 0.034; one-tailed test). As hypothesized the patients recalled fewer words than the controls. There was also a large Trial effect. These main effects were also found in Experiment 1. This time however there was no significant interaction between Group and Trial.

The difference in delayed recall between patients and controls was not significant (P = 0.11) in this analysis, and no significant differences were found for other score parameters.

The difference in recognition scores (hits and correct rejections) between patients and controls in a separate analysis of this condition were not significant, nor were there significant interactions with Group. The patients gave correct responses in 89% of the trials, the controls in 90%.

Separate analysis of the recognition times again showed that the responses of the patients were slower than those of the controls for both hits and correct rejections (F(1,15) = 9.14, P = 0.009). This time there was no Response Type effect, nor was there a significant interaction.

---

**vs controls) as between subject factors. An interaction between Condition and Group on the first trial measures would be suspect in the sense that the patient groups would not be comparable in the two conditions. No interaction of this kind was significant. The largest interaction was for recognition accuracy (F(1,29) = 4.1, P = 0.052). That is, the patients in the one-trial condition had a tendency to recognize fewer words than the subjects in the other three groups. Better performance in the three-trial condition was the case for delayed recall (F(1,29) = 40.7, P < 0.001) and recognition accuracy (F(1,29) = 7.8, P = 0.009) but not for speed. Better performance by the control subjects emerged in the first recall trial (F(1,29) = 12.9, P = 0.001), delayed recall (F(1,29) = 7.5, P = 0.01), number of wrong words (F(1,29) = 7.9, P = 0.009), recognition accuracy (F(1,29) = 4.4, P = 0.043) and recognition speed (F(1,29) = 11.8, P = 0.002).**
Discussion

The data of the second experiment show that when depressive patients have little familiarity with the words (one-trial group), delayed recall, recognition memory, as well as immediate recall are impaired. Immediate recall was also impaired in the three-trial patient group. The difference with respect to the number of words recalled was small, but significant (1-tailed test). However, delayed recall and recognition accuracy did not show significant differences in the three-trial condition, which was also found in Experiment I after 5 trials. The data suggest that when the depressives have greater familiarity with the learning material (i.e., when demands on effort have become smaller), the material is encoded into memory better, which results in better retrieval (delayed recall) and recognition memory.

General Discussion

The outcome of the two experiments is generally in agreement with the hypothesis that retrieval from memory is impaired in depression. That is, depressives generally recall fewer words than control subjects. This conclusion is consistent with the subjective complaints expressed by these patients: many patients frequently complain of forgetfulness, of difficulty finding words, and the like. The retrieval deficits are particularly manifest in the beginning of a learning task.

The encoding of information into memory (shown by recognition accuracy) is also impaired, but only when demands on effort are high, that is, in the beginning of a novel task. It might be that a recognition test after one trial demands the allocation of more cognitive effort and can be performed less automatically than when recognition is tested after five trials. It is this effort that is lacking in many patients suffering from depression (Cohen et al., 1982; Roy-Byrne et al., 1986).

There is agreement about the notion that motivational aspects and energetical demands of performance tasks vary in the course of a task. Normally, there is a shift from controlled, effortful processing in the beginning of a task toward more automatic, superficial processing later in the task. Consequently, a test of recognition memory with material that is only processed once may be more effort-demanding than when the material is processed more often. Thus, even the encoding of material, as reflected by recognition accuracy, may be impaired.

Memory retrieval deficits in depressive patients have also been found with Sternberg's memory comparison task (Brand and Jolles, 1987; Brand, 1987). Both short-term memory and sensorimotor parameters were found to be inferior to those of controls. Earlier work with healthy volunteers, using the memory comparison task and the RAVLT, showed that the short-term memory speed parameter and recognition speed were closely related, suggesting common features of retrieval measures in different tasks.

Depressive patients showed a slow recognition speed. This may be ascribed to less efficient encoding and to deficits in decision and motor output processes. Although effects on these different processes may not be discerned in the present data, the evidence from the memory comparison study mentioned earlier pointed at impaired (sensori)motor processes that were found in addition to impaired memory processes (Brand and Jolles, 1987; Brand, 1987). Likewise, motor dysfunctions in depression have been reported frequently (Byrne, 1976; Cornell et al., 1984; Brand et al., 1990), and Miller and Lewis (1977) and Cutting (1979) also mentioned differences in decision making factors between depressive patients and controls.

It may be suggested that the impaired recall and recognition of depressive patients is related to motivational differences. However, Sigmon (1987) in a study with a coding task in depressed college women did not find any support for motivational deficits in depression. Motivational differences, if existent in the present samples, may be caused by the fact that the control subjects (drawn from the normal local population) were paid for their services, and the patients were not. However, the payment was not substantial and probably not of influence. Since impairments in encoding and retrieval in the patients were most marked in the beginning of the task, when demands on effort are high, the conclusion that these differences were due to differences in effort, is more warranted, and more in line with
current notions (Cohen et al., 1982; Roy-Byrne et al., 1986).

The outcome of this study is relevant in both a fundamental and clinical sense. The fundamental aspect is that it is becoming more and more clear that different tasks make different demands on effort. This notion is incorporated into the resource allocation model (Ellis and Ashbrook, 1988). Moreover, demands on effort may vary in the course of a task. These two notions are important when comparing results from different research studies: are measurements made at the same level of subject/task demands in any two studies, so that the same amount of effort is called for?

The clinical relevance of this study is that memory performance may adjust to normal if patients are stimulated to active rehearsal and repetition. The differentiation between deficits in retrieval and encoding of information is of considerable clinical importance as patients in an early stage of dementia (who are difficult to distinguish from elderly depressive patients) are characterized by global deficits in memory (Jolles and Hijman, 1983; Jolles, 1985, 1986), as exemplified by a deficiency in both recall and recognition, even when demands on effort are very low.

Acknowledgements

The research presented here was partly supported by Grant 560-264-010 from the Netherlands Organization for the Advancement of Pure Scientific Research. Special thanks are addressed to A. van der Wijk who performed the second experiment presented in this article.

References


