"MEMORY FUNCTION" IN PSYCHIATRIC PATIENTS OVER SIXTY, SOME METHODOLOGICAL AND DIAGNOSTIC IMPLICATIONS*

By

M. B. SHAPIRO, M.A., Ph.D.

F. POST, M.B., M.R.C.P., D.P.M.

B. LÖFVING, F.K.(Stockholm)

and

J. INGLIS, M.A., Dip.Psych.

Institute of Psychiatry, Maudsley Hospital, London

THE aim of the research of which this study forms part, is to examine the relationship of certain aspects of mental functioning to the psychiatric illnesses of old age.

One of the apparent deficits of function to which psychiatrists commonly attach importance is "memory impairment". This is considered part of various clinical syndromes, especially of the "organic" disorders of senility. The assessment of memory is, therefore, of considerable practical importance and most psychiatrists working with elderly patients use some kind of "memory tests". Such tests are usually unstandardized and lack objective scoring criteria.

There are, in relation to memory assessment, some points of conflict between clinical usage and the evidence of objective psychological investigations. For example, the existence of a "memory function" which could be considered as relatively independent of general mental functioning or intelligence, has not been well substantiated when some such clinical tests have been put to critical examination. Eysenck and Halstead (2) after a brief review of the literature and an experimental attack on this problem, conclude that ". . . the ability involved in the clinical tests of memory studied in this research was identical with that involved in the intelligence test used, and that, therefore, it was misleading to accept scores on these various tests as estimates of a person's 'memory' ability."

Some doubts have also been expressed about the kind of "memory disorders" which may be observed. The consensus of both lay and psychiatric opinion concerning the disturbances noted in the senile organic disorders, for example, accepts "Ribot's Law of Regression" as applied to memory; that is to say, that memory for recent events is in these patients poorer than memory for events in the remote past. Mayer-Gross, Slater and Roth (8) say, of senile psychosis, that "The onset is gradual, and the disease may seem at first to be an acceleration of the normal processes of ageing. The patient's memory usually becomes affected first; he is unable to retain or recall events, though his *memory* for remote experience seems intact." On the other hand, Jones and Kaplan (7) in their review of some psychological studies which have attempted to examine the adequacy of this "Law", conclude that it oversimplifies the nature of the apparent memory loss in the disorders of old age.

* The work reported in this paper was made possible by a grant from the Research Fund, made available from the endowment by the Board of Governors of The Bethlem Royal Hospital and The Maudsley Hospital.

In addition, little or no attention is paid clinically to the fact that the concept of "memory" may subsume several distinct processes. Hull (6) for example, has noted that ". . . it is at once apparent that under the expression 'poor memory for recent events' . . . are grouped without distinction two things which are psychologically very different and which may quite probably be expected to vary more or less independently of one another. One is the rate of formation of associations, and the other is the degree of retention of these associations, once they are formed."

This paper is concerned with the first stage of the investigation of some of these problems. This stage comprises the examination of the discriminating power of certain objective test results as between "functional", "doubtful" and "organic" groups of patients, these groups being defined respectively as suffering from "Psychiatric illnesses not associated with cerebral pathology", "Illnesses associated with very mild or dubious organic involvement", and "Psychiatric illnesses definitely associated with cerebral pathology". The tests were designed to examine systematically some of the notions commonly held about the nature of mental dysfunctioning in elderly psychiatric patients. Five groups of tests were decided upon and they are briefly described below:

(*a*) Measures which attempt to systematize the formal unstandardized procedures which are part of the psychiatric investigation given routinely to patients in this hospital and also in common use elsewhere: Mayer-Gross, Slater and Roth (8). These measures are together labelled the "Clinical Sensorium Test" and include such tasks as learning and remembering a name and address and reciting the months of the year backwards.

(*b*) Measures which are an objectification of the observations the clinician uses when he makes a judgment of "memory impairment" in a patient. These tests are subsumed under the title "Clinical Questionnaire" and are especially concerned with finding out what differences, if any, there are between memory for recent events and memory for remote events. This questionnaire contains, for example, such items as "Where were you born?", "How far did you go in school?", and, "How did you get here?", "When were you admitted?"

(*c*) Measures which attempt to objectify the kind of symptoms of "memory disorder" which would result in a relative taking the patient to see a doctor. This group of items is called the "Memory Questionnaire" and includes observations as to whether the patient can find his way about the ward, knows where the lavatory is, and the like.

(*d*) The kinds of test a psychologist might use or invent to examine those various aspects of mental function which might be subsumed in the concept of "memory", e.g.:

1. "Memory for first testing session", which consists of simple questions of the form "How many days ago did I see you?"

2. The New Word Learning and Retention Test: Nelson (10) which is concerned with the learning and retention of the definitions of words presented orally.

3. The Bender Gestalt Test: Bender (1), which is a simple test of the ability to perceive visual patterns and reproduce them accurately. This test is included in view of the fact that apparent differences in "memory" may be due to differences at the stage of impression in the learning process. Clinical use of the test has suggested that a tendency to poor recall of the designs may be characteristic of brain damaged patients: Hanvik and Andersen (4), a simple recall version of the test is, therefore, included.

(*e*) A test of general intelligence is included which consists of four subtests of the Wechsler Bellevue Adult Intelligence Scale: Wechsler (13). The purpose of this test is to see whether differences appearing on "memory" items can be accounted for in terms of differences in intelligence.

METHOD

Subjects

The sample of patients tested numbered 102. These patients were both men and women, who were admitted consecutively to the geriatric unit of the Bethlem Royal Hospital over a period of about one year. Their ages ranged from 60 to 92 years, with a mean age of 68·64.

An attempt was made to examine every patient admitted during the period the investigation was being carried out. Five patients, however, had to be omitted from the investigation because they were too demented or too disturbed or out of contact to respond to any questions. Eight of the 102 patients could not be tested on all items because of defective vision. No patient had been given E.C.T. before testing in the admission period during which the investigation took place.

The procedures used in dividing this sample into various diagnostic groups will be described later.

Tests

A standard method of administration and objective scoring criteria are used for each of the following tests:

Clinical Sensorium Test. This consists of six short subtests as follows:

1. This item comprises seven simple questions concerned with the patient's orientation for time and place: e.g. regarding the day of the week and the location of the hospital.

2. This item is intended to measure immediate and delayed recall of a name and address. The name and address is read aloud by the examiner and repeated until the patient is able to reproduce it correctly. After five minutes the patient's memory for the address is tested.

3. This is a "Logical Memory" test adapted from the "Memory" item of Year X (Form L) of the Terman-Merrill Binet (12). The examiner reads a short paragraph, which gives an account of a city fire, to the patient who then has to repeat as much of this as he can remember.

4. This item is called a "Concentration Test". It consists of two tasks: naming the months of the year backwards and repeating digits forward and backward.

5. This is an "Arithmetic Test" which includes three simple problems and the "Serial Sevens" test in which the subject is required to count backward from 100 by sevens.

6. This item is the so-called "Cowboy Story": Henderson and Gillespie (5). It consists in a short story about a cowboy and his dog, the patient being required to grasp the point of the story when it is read to him.

Clinical Questionnaire. This consists of three main types of question intended to measure the patient's ability to answer questions about his past life and recent personal and general events. It comprises the following:

1. "Memory for remote personal events", which consists of twelve questions concerned with the patient's early life, such as place of birth, name of

school, age at and year of marriage, children, occupational history and the like.

2. "Memory for recent personal events", which consists of ten questions concerned with admission to hospital (time, form of transport, etc.), and immediate memory for time and person (such as last interview with doctor, name of doctor, etc.). This test is given on the 6th–10th day after admission to hospital.

3. "Memory for recent general events", which consists of ten questions concerned with well-known public events, persons and dates. The patient is also asked about some recent event within his own sphere of interest: e.g. sport, films, politics, etc.

Memory Questionnaire. This includes three sets of questions on memory for immediate events or orientation in hospital:

1. "Memory for time and hospital routine", includes questions dealing with ward activities, social activities, visiting days, times of meals, etc.

2. "Memory for people on ward", in this test the patient is asked to tell who is in charge of the ward, who sleeps in the same room, and the like.

3. "Orientation on ward": in this test the patient is shown a plan of the ward and is asked to indicate where various places are situated (lounge, bedroom, dining room, etc.). The original intention had been to ask each patient to draw a plan of the ward but this procedure was discarded, since it was found to be too difficult for most patients. In addition, each patient is tested on the ward, being asked where he sleeps and so on.

Psychological Tests. These are of three main types:

1. "Memory for first testing session." This test is given two days after the administration of the Wechsler and the New Word Learning and Retention tests and consists of six questions, such as: "How many days ago did I see you?" "Where was it?" etc.

2. New Word Learning and Retention Test: This test is intended to assess the patient's ability to learn and his capacity to retain simple word definitions. Its administration in this investigation is slightly modified from the method described by Nelson (10). It consists in getting the patient to learn, by successive auditory presentations, the first five successive words of the Wechsler Vocabulary list which he fails to define correctly; that is to say, the learning test is introduced at the point where the Vocabulary test is discontinued according to Wechsler's criterion.

(*a*) "Learning Measure"—The words are presented and defined orally. After this first presentation the examiner asks for the correct definition of each in turn. If the patient fails to respond correctly the examiner repeats the definition. The examiner changes the wording somewhat (in an unstandardized manner) on each presentation, in order to prevent rote learning of a verbal pattern and to make sure that the patient has grasped the meaning of the word. When all the five words have been correctly defined twice by the patient, this part of the test is stopped. In order to prevent over-learning of the first-learned words, each word correctly recalled on two occasions is dropped from later presentations of the list. If necessary each word is presented ten times. In this way the test is brought within the capabilities of most patients; nevertheless, in this study, some still failed to learn all the words. In that case, if a word is still not learnt after ten presentations it is dropped and counted as a failure.

The learning score is the number of definitions necessary to learn all the words. As some of the patients in this study failed to complete the test, and

additional score of "partial learning" was calculated equal to the number of words which were learnt to the criterion.

(*b*) "Retention Measure"—This consists in a measure of the patient's ability to retain the five words over varying periods of time. The patient is not informed, at any time, that he is going to be asked about the words again. Word retention is tested twice; first, after an interval of 20 to 25 minutes; this period being occupied by the administration of two subtests of the Wechsler intelligence scale. It has frequently been pointed out that one of the most important conditions governing retention is the amount and sort of activity in the intervening time interval between learning and reproduction. The first interval in this version of the test is, therefore, fixed at 25 minutes so as to be under the control of the examiner and contained within one testing session. This enables a measure of retention to be secured which is relatively uninfluenced by uncontrolled extraneous factors. The sum of correctly defined words after this period is the first score of retention. Since, in this study, some patients failed to learn all the words an alternative score of retention was given in which the number of words retained was considered as a proportion of the number of words actually learned. The second test of retention is given two days after learning has taken place. It is administered and scored in the same way as the first test of retention.

3. Bender Gestalt Test: This is a relatively simple test of perceptual function, performance on which also involves reproduction by drawing. The test consists of six designs of the Bender Gestalt somewhat simplified and enlarged in order to enable patients with slightly defective vision to do the test. Each patient is asked to copy all the designs and, after an interval of one minute, to draw them from memory. A quantitative adaptation of Bender's scoring was developed by Miss Löfving.

Intelligence Tests. These comprise four subtests of the Wechsler Adult Intelligence Scale. They are the Vocabulary, Similarities, Block Design and Picture Completion, subtests.* A few minor alterations were made in the administration procedure described by Wechsler. In the first item of the Similarities subtest the patient is prompted not once but twice, if he gives an inferior answer and is told that a likeness is required and, if necessary, the correct answer is also given twice. No time limit is applied in this test or in the Picture Completion test. Apart from these alterations, the material and procedure are exactly the same as Wechsler's. The subtest weighted scores are used in all calculations. No I.Q.'s were computed as no data are available which would permit these to be reliably calculated for this age group.

The patients in the present study were tested by one of the psychologists (Miss Löfving) on most tests; they were administered in three sessions. Each session required from thirty to sixty minutes. The interval between the first and second session was fixed at two to three days. In addition, each patient was tested once by the junior doctor in charge of the case. The doctor administered the clinical tests (i.e. the Clinical Sensorium and the Clinical Questionnaire). This session required from 45 to 60 minutes. Most of the patients were co-operative; occasional irritability and agitation were fairly easily smoothed over. On the whole the patients appreciated the testing as part of the full general

* A product moment r of ·956 has been obtained between a combination of these subtests and the full-scale Wechsler test. For this validation study a sample of 65 patients was drawn from a senile population of the Bethlem Royal Hospital, identical to that of the present experimental population in terms of age, intelligence and clinical diagnosis. A detailed report of this study by Miss Löfving is to appear.

examination to which they were submitted during the early part of their stay in hospital.

RESULTS

The analysis of the results will be presented in two stages; these stages being the outcome of different methods of classifying the patients into the "functional", "doubtful" and "organic" categories.

First Investigation

In planning the study, much consideration was given to the problem of contamination of data. In its "direct" form this contamination consists in the clinician having some knowledge of the psychological test results, and perhaps unwittingly, using such knowledge as part of his criteria for classifying the patients into the various diagnostic groups. When the groups are arranged on this basis and these test results are then further examined, it is of course found that they discriminate very well between groups. The error inherent in such a method is obvious, it lies in the fact that the clinician has pre-judged the validity of the test items whose validity is, in fact, under scrutiny.

Care was taken to avoid this form of contamination in this first investigation. The psychiatrist (Dr. Post) had no knowledge of any of the test results. The first classification was done on the basis of his "global" clinical impression of each patient's illness gained from a study of the history and thorough examination of the patient. The criteria which he used in sorting the patients into groups can be outlined as follows:

Functional. This title was used in the sense in which the term is employed in describing conditions such as obsessional neurosis, manic-depressive psychosis, schizophrenia, etc. In most of the patients the illness was of recent onset and insufficient time had passed to allow all problems of diagnosis to be clarified. It was, however, possible for the psychiatrist to state in a high proportion of cases that there was nothing to suggest the presence of cerebral pathology from the way in which the illness had developed or from the state of the patient as seen during several weeks of observation.

Organic. This label was applied to patients with histories highly suggestive of cerebrovascular incidents or of progressive memory and intellectual impairment; these signs were also evinced on routine clinical examination. Physical examination may, in some cases of this group, also have revealed abnormal neurological signs confirming the suggestions in the history of cerebral pathology.

Doubtful. This term was used to describe cases in which neither the history not the clinical examination allowed the psychiatrist to place the individual as lying clearly in either the functional or in the organic group. The psychiatrist was doubtful whether the illness was purely functional when any of the following features were present; a history suggesting a character change occurring over a number of years (e.g. increasing hypochondriasis, pre-occupation with money, deterioration of personal standards); an account of fainting turns without clear cut loss of consciousness or epileptiform features; very mild difficulties of memory; doubtful impairment of some special function (such as spatial orientation); or dubious neurological findings. In addition, poor organization of delusional pre-occupations or the occurrence of occasional visual hallucinations may have made the diagnosis of a functional condition doubtful.

On this first classification the distribution according to clinical grouping was as follows:

Group	Numbers
Functionals	52
Doubtfuls	24
Organics	26
Total	102

The means and standard deviations of each of these three groups were computed on 29 variables. The significance of the differences of the means of the groups for each variable was estimated by analysis of variance. The results are shown in Table I.

TABLE I

	Functionals		Doubtfuls		Organics			
	M	SD	M	SD	M	SD	F	P
Age	66·35	4·89	69·40	6·10	69·44	7·93	3·00	—
(a) *Clinical Sensorium:*								
Orientation	6·31	1·58	5·75	1·20	4·46	2·30	10·03	XXX
Name and address test	4·75	2·08	4·04	2·42	2·54	2·29	8·65	XXX
Logical Memory	8·00	4·06	6·73	3·70	6·15	4·29	2·05	—
Concentration	12·84	2·99	11·00	2·95	9·65	3·18	10·23	XXX
Arithmetic	3·08	1·39	2·50	1·53	1·88	1·66	5·66	XX
Cowboy Story	3·79	1·93	2·92	2·23	2·69	2·52	2·73	—
Total: Clinical Sensorium	38·77	10·72	32·81	10·45	27·58	13·32	8·74	XXX
(b) *Clinical Questionnaire:*								
Memory for remote personal events	9·62	1·99	8·63	2·46	6·92	2·50	12·50	XXX
Memory for recent personal events	7·60	2·18	7·54	1·99	5·58	2·96	6·96	XX
Memory for recent general events	6·50	1·94	5·00	2·32	3·88	2·99	11·50	XXX
Total: Clinical Questionnaire	23·71	5·08	21·17	5·62	16·42	7·59	13·08	XXX
(c) *Memory Questionnaire:*								
Memory for time and routine	15·42	3·36	14·30	3·04	12·04	4·77	6·78	XX
Memory for people on ward	11·48	6·67	9·41	7·37	6·32	6·56	3·34	X
Orientation on ward	12·68	2·68	11·64	2·90	9·48	3·61	7·92	XXX
Total: Memory Questionnaire	39·58	10·66	35·41	12·27	27·92	13·09	9·05	XXX
(d) *Psychological Tests:*								
Memory for 1st testing session	9·27	2·69	7·70	8·29	6·38	3·09	3·31	X
Learning (NWLRT)	17·80	10·47	23·20	15·18	29·40	15·68	6·38	XX
Retention Ia	8·35	2·02	7·54	2·87	8·62	2·89	1·26	—
Retention Ib	3·87	1·19	3·37	1·50	3·04	1·37	3·60	X
Retention IIa	7·34	4·73	6·89	3·68	5·61	3·58	1·08	—
Retention IIb	3·23	1·38	2·70	1·63	2·61	1·58	4·42	X
Bender Copy	20·28	3·29	18·55	3·64	15·68	3·82	13·95	XXX
Bender Recall	11·12	5·23	9·30	4·83	6·48	4·18	7·46	XX
(e) *Intelligence Tests:*								
Vocabulary	10·87	3·14	9·21	3·18	8·23	3·34	6·12	XX
Similarities	7·54	3·67	6·08	2·84	4·50	3·11	7·26	XX
Picture Completion	8·90	3·62	7·05	3·87	5·27	4·16	7·88	XXX
Block Design	6·06	2·61	4·86	2·32	4·04	2·76	5·61	XX
Total Subtests (Pro-rated)	83·31	28·79	69·25	25·91	55·19	26·92	9·21	XXX

Key—Levels of significance: X = ·05; XX = ·01; XXX = ·001.

As may be seen from Table I, when the groups are thus compared most of the variables are found to discriminate between the group means at high levels of confidence. The only variables which fail to differentiate significantly between the means are two items of the retention test, i.e. of the New Word Learning and Retention Test, and two subtests of the Clinical Sensorium, i.e. The Logical Memory test and the Cowboy Story. The mean ages of the groups are not significantly different.

Despite the care which had been taken to avoid "direct" contamination of the data in this first investigation, a close examination of the results shown in Table I and the method by which they had been derived, suggested that a kind of "indirect" contamination might have seriously affected the findings.

Thus, while the psychiatrist who classified the patients had no knowledge of the actual results of the psychological tests or of the clinical tests administered by the junior doctor, and while he was in any case critical of the validity of the tests in the Clinical Sensorium group, Norris and Post (11), nevertheless, he could not fail to form impressions of general cognitive efficiency and memory functioning from the patient's history and during his clinical examination. These impressions may then have entered into his criteria for sorting the patients into the groups. The purpose of giving the test items, however, was to objectify such impressions so that their actual diagnostic value could be empirically determined. If these impressions formed any part of the criteria for classification then agreement between the psychiatrist's sorting and the test results might only mean that his impressions of cognitive status could be confirmed by objective tests. "Indirect" contamination might, therefore, have taken place, since the aim of the study was to investigate the possible nature of the relation of cognitive and memory function to illness and not merely to measure a relationship already assumed to exist. Two factors mainly might have contributed to this indirect contamination, these being:

1. That the objective measures used in fact gave good estimates of the sort of cognitive and memory functioning that the psychiatrist may have used implicitly in his sorting criteria.

2. The psychiatrist was a good clinical judge of such functions.

These facts taken together might result in significant differences between the group means without the underlying variables having any relation to brain damage.

It was, therefore, necessary, in order to eliminate the possibility of such contamination, for the psychiatrist to re-classify the same patients, again into three categories: "functional", "doubtful" and "organic", *but this time using only the most objective criteria unrelated to cognitive or memory function.*

Second Investigation

In order that the groups might be so reformed, the psychiatrist provided a check list of signs (so far as possible avoiding any related to cognitive function) which he considered would help him to decide whether a patient could be classified as organic or not. He provided eight signs, four from the history and four from the examination of the patient. The relevant information was taken from the case notes by one of the psychologists (Inglis). This was done about one year after the first classification. The signs were:

History. (*a*) Cerebral incidents or attacks, that is to say, a history strongly suggestive of focal cerebral disorder, including fits, transitional paresis and the like. (*b*) Confusional episodes, that is to say, episodes of abnormal behaviour and psychotic experiencing in a setting of disorientation. (*c*) Incontinence, except if occurring only at the height of the illness. (*d*) Character change, such as deterioration of personal habits and/or caricaturing of previous personality traits: e.g. increased pre-occupation with money, health and so on.

Examination. (*a*) Confirmed neurological signs or symptoms, including aphasia. (*b*) Perplexity, that is to say, an impression or complaint of bewilderment or puzzlement. (*c*) Grossly abnormal EEG. (*d*) Presentation of a clinical picture not clearly belonging to one of the recognized "functional" psychiatric syndromes, e.g. depression with schizophrenic symptoms and visual hallucinations.

The only additional information provided to the psychiatrist was the sex

and age of each patient. The pages on which the information was recorded were coded by random numbers so that the physician could not recognize the patients by name. The three groups were then reformed. This regrouping was itself done twice, the second time as a check on the first; they were done as follows:

The psychiatrist read the sheets on which the information was recorded and then formulated a diagnostic impression from the items on the check list. This impression was not derived by giving explicit "weights" to each item recorded but was made in terms of the psychiatrist's clinical judgment of the diagnostic implications of the individual's record as a whole.

A check on the objectivity of the above method was then done to make sure, for example, that the psychiatrist had not merely remembered the patients (and so his previous diagnosis) on the basis of distinctive patterns of symptoms. This check consisted in numerical weights being given to the various signs. The highest weightings were given to what seemed to be the most severe and also the most objectively founded evidence of cerebral pathology. The main recommendation concerning what weight was to be given to each sign was made by Shapiro, who had not seen any of the actual raw data. The patients were again reclassified, this time on the basis of their weighted scores alone, this reclassification was done by another psychologist (Yates) who was not otherwise connected with the research and who had not previously seen any of the data.

The agreement between these two latter methods of reclassification, the psychiatrist's based on a clinical assessment of the signs and the other done by numerical weighting, was almost complete. Of the 67 cases reclassified by the psychiatrist as "functional", 66 were also classified as "functional" by the check method. Of the 18 classified as "organic" on this occasion by the psychiatrist, 15 were also similarly classified by the check method. Because of this good agreement, it could be inferred that previous knowledge had not significantly distorted the physician's judgment in reclassifying the patients and, therefore, his regrouping could confidently be used as a basis for the second investigation.

Three main questions may now be answered through analysis of the results of the second investigation:

1. It can be shown whether there is any other than a chance relationship between the psychiatrist's first grouping and his second.

2. Given that the two groupings are not independent then it is possible to examine whether any significant change has taken place in the classification.

3. It can also be shown what effect any change in grouping has on the discriminating power of the tests and if such effect is consistent with the hypothesis of "indirect contamination".

Of the 102 cases in the sample, 75 individuals remain in the same category on the second occasion as they occupied on the first. The obtained relationships are clearly shown in Table II.

TABLE II

		Those Diagnosed on the First Occasion as:			
		Functional	Doubtful	Organic	Total
Those Diagnosed on the Second Occasion as:	Functional ..	47	14	6	67
	Doubtful ..	5	10	2	17
	Organic ..	0	0	18	18
	Total	52	24	26	102

The relatedness of the two groupings was tested in the manner described by Garrett (3): "χ^2 test of independence in contingency tables." The χ^2 value is 76·45 with df equal to 4, a value significant at beyond the ·01 level. The two

sets of data must, therefore, be regarded as related. These results serve to confirm the reliability of the psychiatrist's diagnosis as between these three groups after a period of one year, this reliability being of the order of 75 per cent.

As it has been shown that the two sets of data are related, it is then necessary to examine if any systematic change has taken place in the grouping on the second occasion as compared with the first sorting. Unsystematic variation alone would be expected to cause approximately as many individuals to be displaced in any one direction as in another. That is to say the number of individuals whose diagnosis changes from "doubtful" to "functional" should be the same as the number changing from "doubtful" to "organic". An alternative to this null hypothesis, however, was formulated by Inglis as a result of reading the case notes. This alternative hypothesis predicted that the flow of change would be significantly greater in the direction of "functional" diagnosis, away from both "doubtful" and "organic" groups. The reason for such an expectation is as follows. If the patient had, on the first occasion, presented enough "objective" evidence of brain damage (such as abnormal neurological findings strongly indicative of pathology) he was placed in the "organic" group. Any impression of apparent preservation of cognitive efficiency did not then seem to influence the diagnosis towards the "doubtful" or "functional" categories. If, however, an individual showed few or none of these objective signs, but at the same time, his efficiency of mental functioning seemed impaired, then this fact *did* appear to influence the diagnosis, making it more likely that he would be put in the "doubtful" or "organic" groups. It follows, if the suggested hypothesis is correct, that the reclassification of the patients, when impressions of cognitive and memory function are excluded, would cause the flow of change to take place mainly in one direction, away from "organic" towards "doubtful" and "functional", and from "doubtful" towards "functional". This hypothesis was tested on the data in the manner described by McNemar (9): "Test for change in correlated proportions." This technique gives a C.R. of 2·46 which, on a one-tailed test of significance (the direction of change having been predicted) is significant at beyond the ·01 level. This result permits confident rejection of the null hypothesis and confirms the alternative hypothesis that the influence of the clinician's impression of cognitive and memory function has served to bias the classification in favour of "organic" diagnosis. This result is also consistent with the hypothesis that the first investigation had produced a classification which had resulted in "contamination" of the data.

Analysis of variance of the group results on each of the tests was again undertaken in order to see what effect the changes in classification have on the discriminating efficiency of the objective tests.

It can be seen from Table III above that the group means for age are not significantly different. Some comments may also be made on other results.

(a) *Clinical Sensorium Test.* Of these measures which attempted to systematize the formal unstandardized procedures which were part of the routine psychiatric investigation only one of the six subtests, that of "Orientation" (i.e. for time and place) discriminated significantly between the group means: ·01 level. The test as a whole fails to differentiate between the group.

(b) *Clinical Questionnaire.* Of the measures which attempted to objectify the observations which lead clinicians to diagnose "memory impairment" two of the three subtests provide significant differences, i.e. "Memory for remote personal events" (at the ·01 level) and "Memory for recent general events" (·05 level), as does also the total score (·05 level). "Memory for recent personal events" fails to reach significance level.

TABLE III

	Functionals		Doubtfuls		Organics			
	M	SD	M	SD	M	SD	F	P
Age	66·90	5·60	69·25	4·88	69·77	8·65	1·97	—
(a) *Clinical Sensorium:*								
Orientation	5·95	1·60	5·69	1·78	4·35	2·52	5·13	XX
Name and address test	4·36	2·36	3·44	2·48	2·88	2·21	3·06	—
Logical Memory	7·25	3·82	7·34	4·37	6·15	4·78	0·53	—
Concentration	11·85	3·04	11·50	3·56	9·88	3·37	2·54	—
Arithmetic	2·72	1·45	2·75	1·44	2·12	1·93	1·09	—
Cowboy Story	3·39	2·28	3·09	2·33	2·91	2·71	0·31	—
Total: Clinical Sensorium	35·52	10·91	33·81	13·39	28·29	14·22	2·42	—
(b) *Clinical Questionnaire:*								
Memory for remote personal events	9·00	2·39	8·81	2·40	6·94	2·61	4·85	XX
Memory for recent personal events	7·31	2·16	7·31	2·58	5·77	3·25	2·77	—
Memory for recent general events	5·89	2·38	4·69	2·50	4·24	3·11	3·52	X
Total: Clinical Questionnaire	22·20	5·83	20·81	6·80	16·94	8·29	4·37	X
(c) *Memory Questionnaire:*								
Memory for time and routine	15·07	3·45	13·75	3·86	12·00	5·30	4·28	X
Memory for people on ward	10·92	6·85	8·88	7·01	6·59	7·58	2·70	—
Orientation on ward	12·21	2·71	11·93	3·15	9·24	4·24	6·22	XX
Total: Memory Questionnaire	38·20	11·03	34·56	12·46	27·82	15·15	4·98	XX
(d) *Psychological Tests:*								
Memory for 1st test session	8·68	2·88	8·16	3·35	6·50	3·21	3·47	X
Learning (NWLRT)	19·44	12·12	25·66	15·35	27·47	16·43	3·06	—
Retention Ia	8·43	2·20	8·63	3·92	7·40	1·74	1·33	—
Retention Ib	3·80	1·18	3·25	1·61	3·00	1·51	3·01	—
Retention IIa	6·67	2·62	6·38	4·29	4·82	3·11	2·47	—
Retention IIb	3·13	1·31	2·66	1·97	2·12	1·66	3·26	X
Bender Copy	19·43	3·40	19·38	4·02	15·24	4·47	8·82	XXX
Bender Recall	10·30	5·08	9·38	6·00	6·35	4·09	3·99	X
(e) *Intelligence Tests:*								
Vocabulary	9·62	3·26	10·19	3·82	8·71	3·69	0·80	—
Similarities	6·54	3·57	6·13	2·71	4·88	3·41	1·57	—
Picture Completion	7·61	4·15	8·44	3·48	5·88	4·81	1·69	—
Block Design	5·33	2·59	5·94	2·27	3·94	3·11	2·62	—
Total: Subtests (Pro-rated)	72·75	30·82	76·72	23·50	58·53	30·40	1·90	—

Key—Levels of significance: X = ·05; XX = ·01; XXX = ·001.

(c) *Memory Questionnaire.* Of the measures which attempted to make objective estimates of the kind of symptoms of "memory disorder" which would result in a relative taking the patient to see a doctor, two of the three subtests have significantly different means, i.e. "Memory for time and hospital routine" (at the ·05 level) and "Orientation on Ward" (·01 level), as does the total score (·01 level). The group means for "Memory for people on ward" are not significantly different.

(d) *Psychological Tests.* The kinds of tests the psychologist might use to investigate various aspects of "memory" show the following results.

1. "Memory for first testing session." On this test the difference between the group means are significant at the ·05 level.

2. "New Word Learning and Retention Test." The group means on the learning measure of this test are not significantly different. Only one item of the retention measures differentiates significantly between the means and that item at only the ·05 level.

3. "Bender Gestalt." The group means on this test are significantly different at a high level of confidence. In this respect the Copying test is the best item in the whole battery (·001 level). The recall test also differentiates (·05 level).

(e) *Intelligence Tests.* The means of the diagnostic groups on these tests are not significantly different.

The fact that there are fewer significant differences between the group means and the fact that the levels of significance which are obtained are lower, is consistent with the hypothesis that the results of the first investigation were due in part to the effects of "indirect contamination".

DISCUSSION

1. The dual form of the investigation serves to show that the danger of "contamination of data" may arise in an "indirect" as well as in a "direct" way. That is to say the "direct" form of contamination may result when the individual who selects the groups whose characteristics are being studied has knowledge of the test results and uses this knowledge as part of the criteria for classifying the groups. "Indirect" contamination, however, may occur when part of the criteria for forming the groups consists in judgments or clinical impressions of the very variables which the tests given to these groups are attempting to objectify and whose relevance to such grouping the tests are being used to examine. This form of contamination can also produce good discrimination between the various groups according to test results without the functions assessed by the tests having any real relevance to the actual group differences. For example, suppose that it was desired to examine whether the presence of characteristic X had any relation to the existence of condition Y. To study this problem it would first be required to secure two groups, a group having condition Y (group Y) and a group not having condition Y (group not-Y). It can be seen that if the person sorting individuals into group Y and group not-Y used his "judgment" of the presence of X as one of his criteria for putting an individual into group Y and his "judgment" of the absence of X as one of his criteria for putting an individual into group not-Y, and, further, if his "judgment" concerning the presence or absence of characteristic X were good, then it would follow that any objective measure of X, given to these two groups, would show that characteristic X was possessed more frequently by the members of group Y than by the members of group not-Y. But, it can also be seen that this result would not answer the initial question as to whether characteristic X had any real relevance to condition Y. This question can only be answered when groups Y and not-Y are sorted initially without any reference to X and *then* the groups are tested for the presence or absence of this characteristic.

2. The data in this study, when freed, so far as possible, from the influence of contamination, fail to confirm, for patients of this group, the diagnostic value of certain clinical tests in common use, notably those procedures which were part of the routine psychiatric investigation, i.e. the Clinical Sensorium, including some which have long been accepted instruments of diagnosis, such as the "Name and Address Test" and the "Cowboy Story".

3. The results obtained on the tests included in the Clinical Questionnaire agree with the suggestion made by Jones and Kaplan (7) that "Ribot's Law" as applied to memory in the organic disorders of old age cannot be uncritically accepted, since, in the present investigation "Memory for remote events" in fact shows more significant differences between the group means than "Memory for recent events". It may be suggested, therefore, that the apparent preservation of memory for past events which is often described as characteristic of such elderly organic patients, does not stand up to objective examination. The possibility exists that the seemingly better memory for past events com-

monly mentioned, may be due to the fact that old people are uninterested in the present and tend to talk much more about the past.

4. Examination of the relationship of certain aspects of mental functioning to the psychiatric illnesses of old age confirms that performance on some tasks which appear to involve "memory" is relatively more impaired in the organic cases. The exact nature of the psychological functions involved in performance on such tasks remains to be investigated.

5. It can be inferred that the differences between the groups on these "memory" tests cannot be accounted for in terms of differences in intelligence alone, since the so-called memory tests have significantly different group means while the intelligence tests have not. This cannot, however, be interpreted to mean that these tasks necessarily have a "memory factor" in common. What causes the discrimination between the groups may be something quite specific to the individual tests. This problem remains to be studied.

6. It must be noted that the test which discriminates between the group means at the highest level of significance is apparently a test of perceptual function, i.e. Bender Gestalt. The nature of the psychological mechanisms underlying performance on such a task remains to be explored.

7. The amount of overlap between the group distributions on all the tests, even those showing significant differences between group means, is large. It is unlikely that any of the measures could be usefully employed as instruments of clinical diagnosis before they have been subjected to further investigation and refinement, except perhaps the Bender Gestalt test.

Summary

1. The aim of the research is to examine the relationship of certain aspects of mental functioning to the psychiatric illnesses of old age. The present paper has attempted to attack the problem of differential "memory function" by examining the discriminating power of certain test results as between "functional", "doubtful" and "organic" groups of elderly patients.

2. The design of the present study shows that spuriously positive results may arise from inadequate control over factors contributing to the selection of the criterion groups.

3. The investigation fails to confirm, for the patients tested, the diagnostic usefulness of certain tests in common use.

4. The results fail to confirm the predictions of "Ribot's Law" as applied to memory in these elderly patients.

5. The study shows significant differences between the means of the results of the psychiatric groups on some tasks which appear to involve "memory function". These differences cannot be accounted for merely in terms of group differences in general intelligence. The nature of the possible underlying "memory function" remains to be further examined.

6. The investigation so far shows that the best discrimination between the groups is achieved by a simple perceptual task (Bender Gestalt).

7. The study fails to show differences between groups on any of the tests which are sufficiently great to enable these tests to be immediately employed as useful diagnostic instruments.

References

1. Bender, L., *Instructions for the use of the visual-motor Gestalt test*, 1946. New York: Am. Orthopsychiat. Assoc.
2. Eysenck, H. J., and Halstead, H., "The Memory Function. I. A factorial study of fifteen clinical tests", *Amer. J. Psychiat.*, 1945, **102**, 174–180.
3. Garrett, H. E., *Statistics in Psychology and Education*, 1948. 3rd ed. New York: Longmans, Green and Co. Inc.
4. Hanvik, L. V., and Andersen, A. L., "The Effect of Focal Brain Lesions on Recall and on the Production of Rotations in the Bender Gestalt test", *J. Consult. Psychol.*, 1950, **14**, 197–198.
5. Henderson, D., and Gillespie, R. D., *A Textbook of Psychiatry for Students and Practitioners*, 1952. 7th ed. London: Oxford University Press.
6. Hull, C. L., "The formation and retention of associations among the insane", *Am. J. Psychol.*, 1917, **28**, 419–435.

7. JONES, H. E., and KAPLAN, O. J., in Kaplan, O. J. (Ed.), *Mental Disorders in Later Life*, 1945. Stanford University Press.
8. MAYER-GROSS, W., SLATER, E., and ROTH, M., *Clinical Psychiatry*, 1954. London: Cassell and Company Ltd.
9. MCNEMAR, Q., *Psychological Statistics*, 1949. London: Chapman and Hall, Ltd.
10. NELSON, E. H., *Experimental Investigation of Intellectual Speed and Power in Mental Disorders*, 1953. Unpublished Ph.D. Thesis, University of London.
11. NORRIS, V., and POST, F., "Treatment of Elderly Psychiatric Patients. Use of a Diagnostic Classification", *Brit. Med. J.*, 1954, 675–679.
12. TERMAN, L. M., and MERRILL, M. A., *Measuring Intelligence*, 1949. London: George G. Harrap and Co. Ltd.
13. WECHSLER, D., *The Measurement of Adult Intelligence*, 1944. 3rd ed. Baltimore: The Williams and Wilkins Co.