WORLD HEALTH ORGANIZATION

Regional Office for the Eastern Mediterranean

ORGANISATION MONDIALE DE LA SANTÉ

Bureau régional pour la Méditerranée orientale

SEMINAR ON THE APPLICATION OF PSYCHIATRIC EPIDEMIOLOGY
Khartoum, 17 - 21 February 1975

EM/SEM.APPL.PSY.EPID./6

ENGLISH ONLY

DIAGNOSTIC PROCESSES AND THE CLASSIFICATION OF MENTAL DISORDERS*

by

Dr John E. Cooper**

* The participants should make themselves familiar with the 8th Revision of the International Classification of Diseases, Chapter V "Mental Disorders", and should have studied the World Health Organization publication "Glossary of Mental Disorders and Guide to their classification", WHO, Geneva, 1973: particular attention should be paid to the "Introduction" and "Notes for Users".

The paper will not be read in full at the Seminar but will form the basis for discussion on the topic: it would be helpful if participants made note of points in the paper that in their opinion particularly need discussion.

Additional material can be found in the excerpts of Maudsley Monograph No.20: "Psychiatric Diagnosis in New York and London" by J.E. Cooper, R.E. Kendell, B.J. Gurland, L. Sharpe, J.R.M. Copeland and R. Simon (Appendix 1), and Section 2 of Acta Psychiatrica Scandinavica, Supplement 201, 1968 "An Experimental Approach to Psychiatric Diagnosis" by M. Shepherd, E.M. Brooke, J.E. Cooper and T. Lin (Appendix 2) attached.

** Professor of Psychiatry, University Department of Psychiatry, Mapperley Hospital, Nottingham, United Kingdom.

If several psychiatrists are presented with the same patient or case-history and asked to make a diagnosis, disagreements are likely to arise, particularly if the patient's illness falls within the "functional" psychoses, or within the neuroses and personality disorders. The size and nature of such disagreements is a greater problem in psychiatry than in most medical and surgical disciplines because of the lack in psychiatry of independant and reliable laboratory investigations that might contribute to the diagnostic process It is usually necessary to rely upon what can be learned about the patient by means of interviews and observation only , and it is easy to show that differences are common between clinicians in their ideas of what is normal and abnormal, and how they perceive and interpret the statements and behaviour of patients. It is also easy to demonstrate that psychiatrists use important words like "schizophrenia" to mean different things in different countries, and even psychiatrists who have been trained and educated together often find that they disagree to a surprising degree about, for instance, the most usual way to classify the various types of depressive illnesses.

In epidemiological work, it is particularly important to specify as clearly as possible what is meant by diagnostic terms, so that others in addition to the doctor who made the diagnosis can deduce what was wrong with the patients. This paper deals with a number of ways in which diagnosis can be made as reliable and as communicable as possible. It is arranged in four sections:

a) notes on some general problems of psychiatric diagnosis
b) diagnosis in ordinary clinical work
c) diagnosis in statistical returns from hospitals and clinics
d) diagnosis in special research studies.

a) Notes on some general problems of psychiatric diagnosis

In their general medical training, psychiatrists learn to think of the diagnostic process as a search for the underlying disorder that is causing the patient's signs and symptoms; in other words, the diagnosis is usually regarded as the key to aetiology, from which treatment and prognosis can usually be deduced. There are nowadays a sufficient number of well-understood medical

and surgical conditions for this aetiological search often to be successful and it is all too easy to assume on taking up psychiatry that this way of thinking is still appropriate. But only in the organically caused conditions (such as dementias, acute confusional states, brain damage, epilepsy and mental deficiency) do we know anything specific about mechanisms of causation in psychiatric illness. The concepts that underlie words such as "schizophrenia", "mania" or "obsessional neurosis" are more abstract and less well defined than those underlying medical terms such as "pernicious anaemia", "malaria", "pulmonary tuberculosis". These differences are usefully illustrated by considering the way in which the information about a patient is summarised and condensed through several stages as a diagnosis is made -

1. Complaints The patient presents complaints, some of which may be immediately recognisable as due to morbid processes, but often these are mixed up with social or work problems, or interpersonal relationship difficulties unique to the patient.
2. Symptoms The doctor's first job is to sort out these complaints and identify any symptoms. These are descriptions by the patient of feelings or experiences which the doctor has been taught to recognise as part of a morbid process. Symptoms are recognisably the same from one patient to another.
3. Syndromes Symptoms tend to cluster together in groups, and these recurring groups of symptoms are called syndromes. In general medicine, the word syndrome is often used to imply a recognised group of symptoms, for which the underlying cause is not known. For instance, "Cushing's Syndrome" was recognised and called by that name for years before the present understanding of its causation was acquired.
4. Disease concepts When the underlying physiological, anatomical or biochemical mechanism that is involved in an illness are understood, it is possible to develop an understanding of the whole chain of events by which the causative event or process produces the symptoms and syndromes; this knowledge leads usually to rational treatment and firm expectations of the course of the illness (prognosis). The examples already noted of pernicious anaemia, malaria, and pulmonary tuberculosis show how much information can be implied in one or two words.

Our concepts of Schizophrenia, Manic-depressive psychosis and the Neuroses are still at the level of syndromes, since we do not yet understand the underlying disturbances. However, in spite of the large individual

component in the way patients with these disorders present, there is still enough common to the complaints of patients for the symptoms and syndromes to be described and recognised, (even across cultures in many instances). The fact that schizophrenia and manic-depressive psychosis respond to fairly specific chemical and physical treatments is a strong pointer to the existence of some underlying biochemical or electro-physiological disturbances, as yet too subtle for us to detect directly. Thus our present day concepts of these psychiatric conditions can be thought of as intermediate between the simplest type of syndrome noted above, and the familiar medical disease concept, we infer that at least an important part of the symptoms is due to a disturbance that is the same in all the patients with a particular diagnostic label, but we do not yet know what that disturbance is.

The conditions called "Personality disorders" need yet another set of ideas for their description, since most of these disorders are concerned with excesses in quantity or time of behaviour that by itself is not necessarily abnormal.

These comments should make it clear that when we talk of "psychiatric diagnosis" we are using a very general term which covers several quite different concepts of the type of disorder present in the patient. It is for this reason that attempts to describe the psychiatric patient in one or two words lead to difficulties and disagreements. To do justice to the complexity of the patient's complaints, several statements are often necessary, for instance, many patients need three separate statements - one giving the mental illness present, one giving the personality type, and one giving the intellectual level. This leads to the consideration of multi-axial system of classification, which is a subject of current interest in psychiatric classification.

b. Diagnosis in ordinary clinical work

A systematic and rational method for sorting out the information about a patient is the essential first step for the production of a useful diagnosis, whether this is to be used for ordinary clinical work, for statistical returns, or for research.

A psychiatrist will remain unclear about a patient in his own mind unless he has his ideas sufficiently well organised to be able to present the case briefly and coherently to others. The most important step in this process is the production of what is often called a "formulation", which is different from a "summary". A summary is simply a condensed version of all the relevant facts about a patient, collected under the

usual headings of Family History, Childhood Development, Social, sexual and work history etc. The summary should contain no interpretation or inferences.

A formulation is best thought of as the attempt by its author to make sense of the information in the summary. In writing a formulation the author tries to show that he has some concept of what sort of a person the patient is, what has happened to him, both personally and psychiatrically, in the past, and how this is related (if at all) to the present illness and problems. Reasonable inferences must be made, particularly about relationships in time between events and illnesses, and the cause of the present illness must be discussed If interpretations are made, their purpose and theoretical framework should be specified.

The central part of this type of formulation is a discussion of the possible diagnoses - the differential diagnosis, which leads on naturally to treatment and prognosis. An outline scheme for the writing of a formulation is given in Appendix A.

An important difference between a summary and a formulation is that the summary of a patient should be virtually the same, whoever prepares it. But the formulation of a particular patient may differ in some respects for each psychiatrist who prepares one, since it will contain the psychiatrist's own individual views, interpretations and points of emphasis.

Even in ordinary clinical work, the diagnosis given to the patient often has to be recorded for administrative purposes, and even though the psychiatrist making the diagnosis may not be very concerned about the future, his notes may be used by others in years to come if the patient has further illnesses. Thus there are good reasons why the diagnosis should always be recorded in a manner that can be understood by others, any clinician can record diagnoses in his own private system, but in addition, some accepted system of classification should also be used, either national or international. The World Health Organisation glossary and guide to the ICD-8 is for this purpose, and should be used in addition to any local or national systems that are required.

c. Diagnosis in statistical returns from hospitals and clinics

Medical students are usually taught little or nothing about medical administration and the planning and provision of medical services, but yet their conditions of work are in the end determined by decisions made by administrators and planners. Planning decisions are often made on the basis of information collected about ordinary clinical work, and this information often includes the diagnosis of patients.

It is to everybody's advantage if information used for planning purposes has some meaning, so some elementary rules for the collection of diagnostic information are suggested -

1. The diagnosis collected, usually the one recorded on the patient's case notes or record sheet, should be made by a doctor. All too often, investigation of hospital statistical systems shows that the diagnosis is collected or recorded from the case-notes by a nurse or a clerk, at a time when the doctor in charge of the patient has not yet made a diagnosis, or if he has, he has not yet written it down in a comprehensible form. Doctors should be taught to record diagnoses clearly and within a few days of admission, this may be superceded by a final diagnosis after investigations or at discharge, but an admission diagnosis should still be clearly specified.
2. All diagnoses should be expressed by the doctor in the terms and code numbers of an agreed system of classification. Local areas or individual nations may have their own systems which will have to be used, but an official international system should also be used at the same time. The World Health Organisation glossary to ICD-8 is intended as an international reference system to help communications, and is explained in the "Glossary and Guide". A common fault of many hospital systems is that the doctors record diagnoses on their own personal systems, leaving a records clerk or nurse to try to convert their idiosyncratic and undefined terms into an official system. Doctors must be taught to regard classification and glossaries as a useful way of learning about psychiatric illnesses as other doctors see them. Any doctor can, of course, record diagnoses in his own words or using any system he chooses, but he should also make sure that he records a diagnosis using an accepted system that others can understand.

3 Both the doctors and the hospital administrators should know what type of diagnosis is required by the collecting office. Do they require the immediate admission diagnosis, or the diagnosis of the most serious condition present if more than one, or do they require the final discharge diagnosis? These may all be different in some patients. If the collecting system can cope, if necessary, with more than one diagnosis for each patient, then an agreed set of rules should be used to decide the order in which diagnoses are given. Some examples and comments on these points are to be found in 'Notes for users' in the World Health Organisation Glossary and Guide.

It is clear from these comments that most of the problems under discussion originate in the attitudes of the psychiatrists towards the keeping of case-notes and the recording of diagnoses. This will only improve if they are taught from the earliest stages in their medical training to take a positive interest in medical administration, planning and epidemiology

d. Diagnosis in special research studies

This section is divided into two, the first dealing with studies in which the patients (and possibly their relatives) can be interviewed by a psychiatrist or other trained research worker, thereby making available detailed and reliable information about all relevant aspects of the patient. The second section contains comments upon different types of studies, usually field surveys on a large scale, in which the amount and type of information obtainable about the patients or subjects is very limited by practical problems of distance or shortage of trained staff, and where the psychiatric diagnosis may be less important than closely related but different statements about degree of disability and need for treatment.

So long as the purposes of a diagnostic study or survey are clearly stated and the methods used are described and adhered to, almost any level of diagnostic information, however crude, is useful for some purpose. It is, however, essential to match the objectives of a study to the methods used to collect the information, and to ensure that information and diagnoses collected for one purpose are not used for a quite different purpose, for which they are unsuitable.

d (1) Diagnosis in special research studies with detailed information.

Because of the ease with which inter-observer variation can affect psychiatric diagnoses, there have been a number of studies in recent years aimed at investigating the conditions under which these differences and variations are likely to be large, and how they can be minimised. In order to minimize diagnostic disagreement, the careful use of agreed systems of classification and accompanying glossaries, as discussed above, is an obvious and essential procedure, but the other components of the diagnostic process also need to be standardised as much as possible.

The complicated process of making a diagnosis can be divided into many stages, but for our purposes the following four represent a convenient division -

1. The interviewing technique of the psychiatrist.
2. His perception of the patient's speech and behaviour, and the patient's perception of his.
3. The inferences and decisions made by the psychiatrist on the basis of what he has perceived.
4. The choice by the psychiatrist of a particular diagnostic term for the patient.

The second and third of these components together constitute a complicated middle stage of the diagnostic process during which the psychiatrist perceives, classifies, summarizes and to some degree interprets statements made by the patient. These processes go on in the psychiatrist's mind as the interview is proceeding, and are his guides in the choosing of further questions or lines of enquiry which in turn result in more information. This rapid and almost automatic system of information extraction and processing is essentially sequential in nature. The topic in the mind of the psychiatrist at any moment is determined largely by what immediately preceded it, so divergence between possible alternative lines of enquiry is likely to be cummulative as the interview proceeds. To achieve in the final stage of diagnosis the high degree of reliability and repeatability necessary for research purposes, all the stages of the diagnostic process must be under some sort of conscious control, as far as the patient's condition and expectations of the interview will allow. To obtain the best results, a complete system of standardised interviews and rating procedures is needed, in addition to the use of a glossary

and classification. If well-trained and committed research workers are available, the diagnostic process can be standardised to a surprising degree without exceeding the tolerance of either the patient or the psychiatrist. (A high degree of standardisation has recently been achieved in two large international studies, (i) the World Health Organisation International Pilot Study of Schizophrenia, and (ii) the US-UK Diagnostic Project, which can be consulted for details of procedure).

The maximum agreement between a group of psychiatrists will be achieved if account is taken of the following -

1. The background and training of the psychiatrists
   The psychiatrists in the research team will find their task of agreeing on the definition and criteria of abnormality contained in the standardised interviewing schedules easier if they are fairly similar in general background, training and psychiatric interests than if they are a very heterogenous group They all have to learn the same rules of rating, and be prepared to use standards and criteria which seem reasonable to others but with some of which they may themselves disagree or have been previously unfamiliar.
2. The need for standardised interviewing and rating procedures
   These are necessary to ensure that the same symptoms are enquired about with every patient, in the same manner Any symptoms thus elicited must then be rated according to an agreed set of rules and definitions. The interviewers must be prepared to train together for weeks or even months until their reliability can be shown to have reached an acceptable level. The training process consists partly of becoming familiar with, and therefore confident about, the schedules, and partly in them being flexible enough in its use to keep the interview both standardised and clinically acceptable to the patient. Interview schedules covering both mental state and history must be included in the system.
3. The use of diagnostic terms with accepted and explicit meaning
   This implies the use of a diagnostic classification with an accompanying glossary of terms, the best known example of this is the Int ernational Classification of Diseases (ICD-8), which now has a glossary and guide. Other examples are The American D.S.M. 1 and 11 and the British Glossary of Mental Disorders, which are all very similar in that they are all based on the ICD-8. It is hoped that in the future, more and more countries will begin to use the World Health Organisation version in statistical returns and research work.

For some research purposes, specially detailed or unusual classification systems may need to be devised, but it is usually possible to state how they are related to the ICD-8. If possible, the results of such studies should also be given in ICD terms or the nearest equivalent, so that other psychiatrists in different countries and cultures can understand at least in part what the results mean.

d (ii) Diagnostic objectives in surveys of mental illness

For some purposes, it may not be necessary to have the information about patients in the very reliable and detailed forms as has been discussed above, and shorter and easier methods of classifying patients may be all that is required. For instance, if the task is to plan a psychiatric service for a rural area with scarce resources and few trained personnel, the diagnostic objectives would be very different from those in a study comparing the effects of two or more drugs on specially selected in-patients. In studies where the research workers have severe problems in travelling and contacting patients it may be necessary to have quite modest diagnostic objectives. The classification of patients into very broad groups such as "psychosis" "neuroses" may be all that is possible, and particularly if it is not practical to have a face to face interview with some of the subjects in the survey. Such terms can still have quite useful meanings in terms of the type of care or the number of medical personnel needed for a service, even though they do not give much information about the details of the patient's illness. In the planning of psychiatric services, the classification of patients into various grades of disability and chronicity may be more useful than the formal psychiatric diagnosis. In addition, the concept of "being a case" has been used and studied in some surveys, and is usually regarded as a measure of the need for psychiatric treatment. All these concepts are closely allied to the more conventional diagnostic terms, but they may be surprisingly independant of them in some patients. Much more work is needed on these other ways of classifying patients, particularly with regard to out-patients and hospital services and clinics, where patients with the less severe types of mental disorders can be treated. Whatever the terms chosen for classifying the patients, they should always be defined as strictly as possible before the survey starts, and preliminary testing should be done to check that the definitions agreed upon can be used reliably by the research workers.

In particular, the method of diagnostic surveys in which all the subjects are not interviewed directly, has been used in a number of surveys of rural areas (e.g. Lambo and Leighton). Diagnoses made in this indirect

way may still be quite useful in making rough estimates of the number of different types of disorders in an area, but the different qualities of such information and the lower levels of confidence implied in such efforts must be recognised and described.

Standardised survey interviewing methods are also now being developed which take the form of several stages of assessment, of different degrees of detail A brief interview of, say fifteen minutes duration, can be used to identify those subjects in a population with more than a certain number of symptoms or degree of disability, and these are then given a more thorough interview to determine in more detail exactly what are their symptoms and diagnoses. The combination of a brief screening procedure with a more detailed assessment of those subjects put forward by the screening procedure allows a large population to be dealt with with the maximum utilisation of the time of the survey workers.

Further reading (Optional)

1. The International Pilot Study of Schizophrenia. W.H.O. Geneva 1973

2. Psychiatric Epidemiology· An international symposium. edited by E.H.Hare and J.K.Wing. Oxford University Press. 1970

From: Cooper, J.E., Kendell, R.E., Gurland, B.J., Sharpe, L., Copeland, J.R.M. & Simon, R., "Psychiatric Diagnosis in New York and London", Maudsley Monograph No. 20, Oxford University Press, London, 1972.

APPENDIX 1

THE BACKGROUND TO THE STUDY

THIS study is concerned with the diagnostic meaning of routinely collected mental hospital statistics, and its design allows comparisons to be made between statistics arising from area mental hospitals in London and State hospitals in New York Psychiatric statistics have not been a popular subject for study in the past, although marked differences have been known to exist between the United States and Great Britain in this field for many years One striking example of these differences will serve to indicate the field of interest of this work, before moving on to an introductory discussion of two issues with which this study is concerned, namely, statistics relating to mental illness, and the reliability of diagnosis in psychiatry

Figures prepared by Kramer (Kramer, 1961, 1969*a* and *b*) showed that the mental hospital first-admission rate for England and Wales for manic-depressive psychosis (or reaction) in the age-group 55–64 years was about 20 times the corresponding American rate This difference is so large that it demands an explanation Possible contributors to the disparity are differences between the two parent populations in hospital utilization and in biological and genetic constitution, and the presence of different ecological influences Before an investigation into any of these, or other, possibilities is justified it is necessary to establish what proportion of the differences in reported statistics can be attributed to the patients themselves If the patients on two sides of the Atlantic turned out to be very similar to one another when both were assessed by the same methods, one would be forced to conclude that the reported diagnostic differences were, at least in part, artifacts produced by differences in diagnostic criteria Such a finding could hardly fail to be of interest to those concerned with the collection and use of statistics, and would also raise the question of whether psychiatrists are entitled to accept at face value diagnostic statements in psychiatric investigations carried out in countries other than their own Thus, an examination of this statistical disparity appeared to be worthwhile whatever the results

NATIONAL STATISTICS OF MENTAL ILLNESS

The importance and usefulness of statistics relating to mental illness have been recognized by prominent psychiatric writers for over 100 years, so it is surprising that so little is known about their deficiencies In the introduction to his *Study of the Major Psychoses in an English County*, Shepherd (1957) traces briefly the history of the recognition of the importance of psychiatric statistics, starting with Prichard (1835), Esquirol (1838), and Thurnham

(1845), and culminating in the British nationwide system of collection made possible by the inception of the National Health Service in 1948, upon which his own study was based The most prominent early studies in this field were not carried out in Great Britain, and Shepherd instances workers in other countries, such as Ødegaard (1945, 1946) and Svendsen (1952) in Scandinavia, and Pollock and Malzberg (1937) in the United States More recently, the World Health Organization, with its international responsibilities, has brought together published psychiatric statistics from many different countries (World Health Organization, 1963, 1966, 1968) These and many other studies provide ample evidence for the usefulness of routinely collected psychiatric statistics, but it must not be forgotten that until very recently the type of information used was quite limited, in that the statistics regularly available were concerned only with admissions to mental hospitals or psychiatric out-patient clinics Even in 1969, Kramer felt obliged to begin an extensive review of contemporary statistics of mental disorders in the United States with the statement 'Annual morbidity statistics on the prevalence and incidence of mental disorders are not available for the United States or for any other country' (Kramer, 1969*b*)

In the majority of published studies, there are no more than passing comments on the quality of the psychiatric diagnoses being used, and in particular there has been insufficient emphasis upon the presence of two distinct components in psychiatric statistics as they are usually collected There is first of all a group of 'hard facts' such as age, sex, marital status, and length of stay in hospital, however these are collected they are likely to be comparatively reliable With the second component, a diagnosis, quite different problems are encountered, due mainly to variations in diagnostic habits among psychiatrists Every psychiatrist is familiar, at an anecdotal level, with the diagnostic differences that can occur between himself and his colleagues under ordinary clinical conditions In the context of clinical work or teaching activities, these differences are often stimulating and productive of useful discussion, but for statistical purposes they are merely sources of variation which must be minimized

Part of the explanation for this lack of appreciation of the inherent problems of psychiatric diagnosis may lie in the fact that those responsible for the collection and compilation of psychiatric statistics are usually administrators and statisticians Experienced clinical psychiatrists who might be expected to be only too familiar with the problems have tended to neglect this field Probably the most widespread manifestation of this indifference is the reluctance of many hospital clinicians in Great Britain to follow the repeated requests of the Ministry of Health to use the nomenclature of the *International Classification of Diseases* when making official hospital admission diagnoses for the monthly statistical returns

It seems reasonable to accept that large-scale psychiatric statistics contain information potentially of great interest for epidemiology and for purposes of

planning and administration, but their acceptance must be qualified by inquiry into the extent and sources of the inevitable variations, and by constructive criticism to improve their quality Uncritical acceptance of their diagnostic value, as in the studies by Burch (1964*a* and *b*), must mean that the authors' conclusions will be at least as suspect as the quality of the diagnoses themselves.

There are signs, however, that more attention is being paid to the quality of psychiatric statistics, as a result of the increasing emphasis in the last two decades upon the rational planning of psychiatric services The complex system of different types of treatment and care that forms the current ideal of a comprehensive psychiatric service needs careful planning to make the most of the usually inadequate resources available It becomes necessary to estimate numbers and types of patients for very practical reasons, and since the deployment of manpower and money will depend directly upon such estimates, they must be based upon reliable information

The need for reliable diagnostic methods and for the consistent use of psychiatric terms, which are central features of the present study, is now only too apparent

THE RELIABILITY OF PSYCHIATRIC DIAGNOSIS

There is a good deal of evidence to show that under ordinary clinical conditions psychiatric diagnosis is usually an unreliable procedure The whole problem of this unreliability will not be reviewed here in full, since several fairly recent commentaries have covered the subject (Kreitman, 1961, Beck, 1962, Zubin, 1967) It is, however, worthwhile selecting a number of studies which illustrate the different types of inter-psychiatrist variation that can influence diagnostic statistics

In one recent experimental study, the sources of variation were conveniently summarized as 'first, variations at the level of observation and perception by the clinician, secondly, variations in the inferences drawn from such observations, and thirdly, variations in the nosological schemata employed by the individual clinicians' (Shepherd, Brooke, Cooper, and Lin, 1968) Good examples of variations between observers at the perceptual level were obtained by Katz, Cole, and Lowery (1969) who showed short cine films of psychiatric interviews to audiences of psychiatrists For two of the patients used, the audiences showed significant variations in their ratings of 'apathy', in spite of the use of a rating method that ensured that all the raters used the same simply-phrased rating-scales (Lorr and Klett, 1967) High ratings on apathy were associated with a diagnosis of schizophrenia, which raises the question of how early in an interview is a diagnosis made by the clinician, and how much does this affect his subsequent judgements and perceptions One of Katz's films was shown in both the United States and in London, and the American psychiatrists had a significantly lower

threshold for the perception of abnormal behaviour and symptoms than the British

This finding was confirmed by another transatlantic comparison (Sandifer, Hordern, Timbury, and Green, 1969) During a series of evaluations of 23 American patients by eight British and 33 American psychiatrists, Sandifer and his colleagues found that the American psychiatrists reported almost twice as many symptoms as the British A complete cross-over rating experiment in which psychiatrists from two or more cultures all rate in turn patients from the same cultures has not yet been reported, but even without this it seems very likely that these two groups of workers have encountered a difference in rating sensitivity between British and American psychiatrists which will prove to be important and widespread

The final stage in the diagnostic process, that of choosing a term from a nomenclature, is particularly vulnerable to individual differences between diagnosticians, as shown by the study of Ward, Beck, Mendelson, Mock, and Erbaugh (1962) This study was directly aimed at identifying and estimating the relative importance of different sources of inter-observer variation while using the *Diagnostic and Statistical Manual Mental Disorders* (DSM I) (American Psychiatric Association, 1952) Each of a series of 153 patients was seen and diagnosed separately by two different psychiatrists who then met to discuss and identify the sources of disagreement in the 40 cases where this had occurred They concluded that the causes of disagreement were

1 Inadequacy of the nosology, responsible for 62 5 per cent of the disagreement (25 cases)
2 Inconsistency on the part of the diagnostician, responsible for 32 5 per cent of the disagreement (13 cases)
3 Inconsistency on the part of the patient, responsible for 5 per cent of the disagreement (2 cases).

By 'inadequacies of the nosology' they meant that they found that descriptions of diagnostic categories were sometimes overlapping and not mutually exclusive, and that there were inadequate instructions and guidance about how to give precedence to one disorder when two or more were judged to be present In other words, even with a conscientious effort to follow the *Manual*, too much was still left to be decided by individual attitudes and preferences.

In most of the reported studies of the unreliability of psychiatric diagnosis, it is not possible to apportion out the demonstrated variations to the different components of the diagnostic process, they show only a total sum of variations from all possible sources Such studies are valuable none the less, since they illustrate how much variation can occur in widely different settings, whether experimental or clinical

Studies done within the setting of one hospital can bring to light a surprising amount of variation between clinicians One such study was reported by Pasamanick, Dinitz, and Lefton (1959), but it does not seem to have aroused

the attention it deserves, perhaps because of the disarming simplicity of its design It consists merely of a description of the routine hospital diagnoses made upon 538 female patients entering the wards of a well-staffed University hospital The patients went into one of three wards with little or no selection between wards The three resulting groups of patients did not differ in a variety of personal characteristics such as age, race, and socio-economic status, so there was no reason to suppose that their diagnostic breakdowns would be very different In fact, very marked differences in the distribution of major diagnostic categories were found and could be traced to the individual ideas of the three psychiatrists in charge of the sections For instance, one psychiatrist made a diagnosis of schizophrenia in 66 per cent of his patients, the others in 22 per cent and 29 per cent For 'character disorders' one psychiatrist put 57 per cent of his patients into this group, a second 47 per cent, and the third only 15 per cent This study is important because the authors were able to show that these variations in diagnostic habits were not merely of academic interest but were associated with differences in treatment Nor do they mince their words in summarizing their conclusions ' despite protestations that their point of reference is always the individual patient, clinicians in fact may be so committed to a particular psychiatric school of thought that the patient's diagnosis and treatment is largely predetermined'

The stability of diagnosis over time has been the subject of some interest as another aspect of reliability, although a change in the clinical state of the patient is always a possible complication A 1-year follow-up study of 100 patients studied in considerable detail (Masserman and Carmichael, 1938) is often quoted as a source of evidence for the unreliability of psychiatric diagnosis, since 41 of the patients required a revision of the diagnosis at the end of the year It is, however, impossible to tell from the method of study whether the changes resulted from change in the patients, or from differing diagnostic habits among the clinicians concerned In addition, examination of the details of the changes shows that some categories were quite stable, for instance 15 out of 18 patients called 'organic psychoses' kept the same diagnosis, as did seven out of eight labelled 'manic-depressive psychosis' The categories of 'adult maladjustment', and the neuroses, were responsible for most of the instability, for only 18 patients out of 45 in these two groups kept their original diagnosis Schizophrenia was intermediate in its stability Such a wide range of variation makes generalizations unjustified, but rather points to the need for further studies on the stability of the individual categories

There is evidence from a British study that changes in doctor produce most changes in diagnosis over time The mental hospital admission statistics collected by the Ministry of Health since 1948 have made possible a large-scale longitudinal study (Brooke, 1963), and from the 44,047 patients in this group who had their first admission in 1954 a group of 200 were selected who had four admissions within 2 years of their first (Cooper, 1967) It might be expected that in the majority of these patients the admissions would all be

for the same condition, but according to the 3-digit categories of the *International Classification of Diseases* (Seventh Edition) into which the diagnoses were grouped, only 37 per cent retained the same diagnosis throughout By regrouping the diagnoses into eight broad groupings (Schizophrenia, Affective Disorders, Neuroses, Personality Disorders, etc) and by subjecting the case-notes of each admission to a uniform assessment procedure by one psychiatrist, the proportion of patients keeping the same diagnosis rose to 81 per cent Strong correlations were also evident between the number of changes in diagnosis and the number of changes of doctor This study was comparatively crude in that it relied upon the assessment of case-notes, but the sheer size of the reduction in diagnostic changes is impressive.

In view of the length of time that this evidence of the unreliability of psychiatric diagnosis has been common knowledge, it is surprising that the problem was not investigated at an international level prior to the recent work of Sandifer and his colleagues (Sandifer, Hordern, Timbury, and Green, 1968) These authors carried out a series of diagnostic comparisons between 33 psychiatrists in North Carolina, four in London and four in Glasgow. Their study used cine films of 30 brief clinical psychiatric interviews, supported by a written summary of the patient's psychiatric, social, and family history and the results of psychological and laboratory investigations. The differences between the diagnoses of British and American psychiatrists were not very striking and were not in accordance with expectations from Kramer's statistics (Kramer, 1969*a*), for the British and North Carolina groups used the diagnosis of schizophrenia to the same extent (for the North Carolina psychiatrists 18 per cent of all diagnoses, compared with 16 per cent for the British) For manic-depressive psychosis, however, the findings were more in line with expectation, the British psychiatrists using this term twice as often as their American counterparts. In addition, the North Carolina group used the term 'neurotic depressive' far more than the British groups, and the Glasgow group alone made particular use of 'personality disorder' The diagnosticians in North Carolina were apparently typical of their local colleagues, for the authors comment that the 18 per cent of diagnoses of schizophrenia made by the North Carolina psychiatrists is close to the annual percentage of admissions so diagnosed at their parent hospital The existence of regional differences in admission rates or diagnostic criteria is an obvious possibility in a country as large and varied as the United States, and may account for the disparity between the results of this study and Kramer's The issue is considered further in CHAPTER XVIII

In contrast, one of the studies reported by Katz, Cole, and Lowery (1969) revealed a very large Anglo-American diagnostic difference for schizophrenia, at least for the one patient involved. A filmed psychiatric interview was shown to 42 American and to 32 British psychiatrists the patient was an attractive young woman in her middle twenties with a variety of fairly mild symptoms of anxiety and depression, who also complained of difficulty with inter-personal

relationships and the frustration of her ambition to be an actress In spite of one-third of the American psychiatrists making a diagnosis of schizophrenia, none of the British put this forward as the primary diagnosis The predilection of American psychiatrists for symptoms related to or suggestive of schizophrenia was shown by the use of the terms schizophrenia and schizoid personality by nearly half of them, whereas over half of the British psychiatrists used diagnoses with a more affective connotation such as depressive neurosis or emotionally unstable personality Another experiment reported in the same paper showed how wide a variety of diagnoses can be given to a single patient, 44 American psychiatrists between them used 12 different diagnoses after all viewing the same film

This last example from the work of Katz shows how marked the contribution of individual differences between diagnosticians can be towards variation in psychiatric diagnosis, since all the psychiatrists in both countries received exactly the same information and recorded their observations according to the same instructions Because of the lack of objective or quantifiable data, there is no doubt that of all branches of medicine, psychiatry is most prone to this hazard But this necessary examination of glaring examples taken from psychiatry must not obscure the presence of a good deal of inter-observer disagreement in many investigations used in general medicine that are usually accepted without question as objective and reliable Beck, for instance, lists the assessment of emphysema, judgements on the nutritional state of children, estimates of the degree of pathological inflammation of tonsils, and accounts of pulmonary symptoms in medical history taking, as examples of medical procedures that have all been shown to be subject to significant inter-observer error (Beck, 1962) Other examples are the assessment of X-ray films (Garland, 1960), blood pressure readings by sphygmomanometer (Oldham, Pickering, Roberts, and Sowry, 1960, Holland, 1963), and the interpretation of ECGs (Kagan, 1965) In other words, there are good grounds for believing that all varieties of the diagnostic process are at times subject to significant observer variation But because of their specially vulnerable position, psychiatrists must be prepared to devote more attention than other diagnosticians to inter-observer differences and other variations inherent in the diagnostic process

Influenced no doubt by these diagnostic discrepancies, authors such as Masserman and Carmichael (1938), Colley (1960), and Menninger (1963) have gone so far as to suggest that conventional descriptive diagnoses are so unreliable that they are best disregarded. Such pessimism is not justified, because the diagnosticians in the studies described above, and in virtually all the others in the literature, were working under ordinary clinical conditions They had not undergone any special training, and at the most had only a brief acquaintance with an agreed rating-scale or with a glossary of terms At least one study performed even under these minimal conditions of standardization achieved diagnostic consistency, probably because the

diagnosticians involved had all had a similar psychiatric training The study, by Wilson and Meyer (1962), stands in interesting contrast to that of Pasamanick, Dinitz, and Lefton (1959) already noted The distributions of different diagnoses in a psychiatric liaison service in a general hospital for 2 consecutive years were compared and found to be very similar, in spite of a different set of patients and a different set of diagnosticians For instance, the proportion of patients diagnosed as schizophrenic remained at 15 per cent, all types of depression together changed only from 23 per cent to 26 per cent, and personality disorders changed from 38 per cent to 34 per cent (the number of patients dealt with was 128 and 166 in the 2 years)

Much more convincing evidence of the reliability of the diagnostic process, however, is available from specially designed studies, as opposed to those carried out under ordinary clinical conditions One of the early reports on the development and use of the 'Present State Examination' (PSE) which was used in a modified form in the present study, demonstrated that a high degree of reliability can be achieved when diagnoses are made under optimal conditions (Wing, Birley, Cooper, Graham, and Isaacs, 1967) During work primarily directed at establishing the inter-rater reliability of symptom scores, a provisional diagnosis was made by the pairs of interviewers independently at the end of each interview Since the PSE is directed only at the symptoms experienced by the patient over the last month, this provisional diagnosis is not a definitive diagnosis, but it must bear a very close relationship to it in most patients The interviewers had all been trained in the same institute and all used a set of instructions geared to the interview and specially prepared for the purpose of making this categorization. Over a series of 172 patients (about half of whom had schizophrenia, just under half some type of affective illness, and the remainder a variety of other disorders) there was complete agreement on primary diagnosis in 84 per cent, when they were tabulated into 11 categories There was partial disagreement in 7 per cent, and serious disagreement in only 9 per cent, a large proportion of the disagreements involving the notoriously difficult category of 'personality disorder', or distinctions between different subcategories of non-psychotic affective disorders

This high level of agreement showed what could be expected with a special effort to control known sources of inter-observer variation Together with the recent emergence of other reliable clinical methods of interviewing and rating (Spitzer, Fleiss, Burdock, and Hardesty, 1964) and with the development of other types of standardized rating instruments such as Lorr's IMPS (Lorr and Klett, 1967), these results at last raised the possibility of a more systematic examination of the diagnostic process in psychiatry As already noted, it had become increasingly obvious to those concerned with psychiatric epidemiology and statistics in America and in Britain that the differences between some of the mental hospital admission rates were the single outward manifestation of several possible differences between patients, psychiatrists, or hospital and statistical services, all worthy of investigation The use of these standardized procedures as a yardstick to set against the more variable diagnoses used in official hospital admission statistics was an obvious and practical way of starting work in this field.

From: Shepherd, M., Brooke, E. M., Cooper, J. E., & Lin, T., An experimental approach to psychiatric diagnosis. An international study, Acta Psychiat. Scand., Supplement 201, Munksgaard, Copenhagen, 1968

APPENDIX 2

DISAGREEMENTS IN THE MAKING OF A PSYCHIATRIC DIAGNOSIS BACKGROUND AND LITERATURE

If several psychiatrists are presented with the same patient or case-history and asked to make a diagnosis, disagreements are likely to become evident, and these are now well documented *Zubin* (1967) and *Kreitman* (1961) have reviewed and summarised a large number of such studies, in which the emphasis has almost always been upon demonstrating the between-psychiatrist reliability of commonly used diagnostic categories Both these reviewers point out, however, that little is known about the sources of the disagreements

The purpose of this section is to discuss some of the causes of diagnostic disagreements in relation to the components of the diagnostic process, and also to comment briefly upon the relevance of some published studies to possible sources of variation in these component processes There is a pressing need at the moment to improve the reliability and validity of psychiatric diagnosis, and some knowledge of the various components that together constitute a diagnostic interview is a pre-requisite for the identification and study of the major sources of variation and error It is not intended to give an exhaustive and critical survey of the literature, but rather to comment upon a number of selected points, illustrating these whenever possible by investigations directly concerned with psychiatric interviewing. Unfortunately, such studies are surprisingly few The very large number of psychological studies dealing with interviewing techniques, perception, and decision theory, being of only indirect interest, can be given only a passing mention in a preliminary discussion of this type, except where they can be used as examples of a particular principle

It might be as well to begin by suggesting a definition of "making a diagnosis", formulated at a sufficiently general level to cover most eventualities For the purposes of this discussion, the following is suggested

> "The psychiatrist interviews the patient, and chooses from a system of psychiatric terms a few words or phrases which he uses as a label for the patient, so as to convey to himself and others as much as possible about the aetiology, the immediate manifestations, and the prognosis of the patient's condition"

These same elements are also found in diagnostic procedures in general medicine, but there the interview and clinical examination of the patient are usually followed by pathological or biochemical investigations which are expected to give specific and independent clues as to the diagnosis. With the increasing variety and specificity of tests of physiological abnormality, and the simplification or even automation of laboratory procedures, little attention is now given in general medical training programmes to the study of interviewing procedures and the sources of differences between observers, but there are some indications that many medical procedures do not rest upon as firm a foundation as is often assumed Disagreement between experienced clinicians, and also between different opinions of the same clinician presented with the same decision on different occasions, has been demonstrated to a surprising degree for such diverse procedures as physical examination of the chest, the assessment of the need for tonsillectomy, and the interpretation of X-rays and electro-cardiograms (*Garland* (1960), *Fletcher* (1952)) But whatever the situation in general medicine, psychiatric procedures must be much more vulnerable to observer differences, for in dealing with most psychiatric patients there is a lack of external criteria by which the psychiatrist's observations and decisions can be validated, it is usually necessary to rely upon what can be learnt about the patient by means of interviews only Until something is known about how to make the gathering and recording of interview information accurate and reliable, psychiatric diagnosis cannot rest upon a firm foundation.

If the definition given above is accepted for the sake of argument, it is possible to separate out four major stages or components of an interview conducted for the purposes of making a diagnosis It will be assumed that the psychiatrist takes an active role for at least a part of the interview, and also that he does not follow a pre-ordained set of questions

These four elements are, first, the interviewing technique of the psychiatrist, second, the perception of the patient's speech and behaviour, third, the inferences and decisions made by the psychiatrist on the basis of what he has perceived, and fourth, the attachment of a particular diagnostic label to the patient The second and third components together constitute a complex middle stage during which the psychiatrist perceives, classifies, summarizes, and to some degree interprets statements made by the patient. These processes go on in the psychiatrist's mind as the interview is proceeding, and are his guides in the choosing of further questions or lines of enquiry, which in turn result in more information Separation in the manner suggested here is somewhat artificial, but is necessary for the purposes of discussion

An important point concerning all four components should be mentioned at this stage Even from the quite general statements made up to now, it is evident that the system of information extraction and processing formed

by the first three components is essentially sequential in nature The topic at any one moment is to some extent determined by what immediately preceded it, and so divergence between alternative lines of enquiry may be cumulative as the interview proceeds The divergence itself would not be important if it were always possible to ask all questions of all patients, for gaps could be filled in eventually, but this of course does not happen owing to the severe practical limitations on the duration of interviews

These four components will now be examined for possible sources of variation which could give rise to disagreements when psychiatrists are faced with the same patient

1 *Interviewing the patient*

Many psychiatrists consciously adopt a particular style or technique of interviewing on the assumption that the pattern they try to impose upon themselves will have an important effect upon the patient's responses

Style of speech is usually assumed to be highly individual, and the first step in studying the importance of interviewing techniques is to know to what extent constant individual elements can be identified The work of Goldman-Eisler showed that this is in some degree possible, although surprisingly difficult (*Goldman-Eisler* (1951, 1958)) She concluded that "certain relations of time sequences of action and silence in conversation tend to be constant within limits, and characteristic of individuals independent of changing partners and topics" The pattern of silences turned out to be the most easily identifiable interview characteristic in her studies, but fortunately other more positive features could also be identified

What exactly determines an interviewer's verbal behaviour is open to question, and the effect of training in interviewing techniques seems to be still uncertain During a psychiatric or psychological training programme a balance has to be struck between the individual verbal characteristics of the pupil and those taught by the system being learned It is often assumed that voluntary adoption by the interviewer of a particular interviewing technique will play the major part in determining the style of the interview, but there is at least one well-known study which suggests that other and more individual influences may be predominant

Fiedler (1950) used tape-recordings of interviews to obtain judgements about type of therapeutic relationships, the judgements being expressed in terms of comments upon the verbal behaviour and styles of the therapists The therapists studied were chosen from three contrasting schools (psychoanalytic, non-directive or Rogerian, and Adlerian) and it was shown that experts of any one school resembled experts of other schools more closely than they resembled non-experts of their own school The influences of

experience (and presumably age) seemed to be more formative of interview behaviour than were training and affiliation

Whatever their sources, these differences between interviewers presumably must have effects upon the patient Even if the content of a group of questions is the same, variations in the style of questioning and subsequent responses of the interviewer will result in the patient receiving impressions of different emphasis, which may well result in different replies The relative efficiency in terms of the speed and completeness with which information is extracted by different styles of interviewing also remains unstudied

2 *Perception*

The psychiatrist can use only information that he perceives, and some points about perception in general will apply to the interview situation Perception is necessarily a selective process, and involves omitting some of the information available, since all possible stimuli cannot be registered simultaneously Only by attending to and concentrating upon selected stimuli is it possible to make sense out of the environment, and what is omitted from perception is in large part determined by the expectations or motivations operating at the time A simple experiment carried out as long ago as 1904 by *Kulpe* (1904) showed, for instance, that subjects told to report as accurately as possible the number of objects presented, were often unable to answer questions about the colours of the objects when questioned immediately afterwards What is perceived is also often supplemented in order to make the perceived material fit better into the observer's frame of reference Bartlett's studies of remembering of stories and pictures contain many examples of both omission and supplementation (*Bartlett* (1932)) *Newcomb* (1950) refers to this whole process, by which incoming stimuli are sorted out and given meaning, as "structuring a process of omitting some features, supplying others, highlighting one or a few and subordinating the rest in the interests of making sense out of the environment"

These are general principles, and clearly similar to the principles of Gestalt psychology and Lewin's field theory Although usually considered as applicable to the perception of stimuli of a comparatively simple nature, as in experiments on the perception of colours, shapes, numbers and simple movements, there is no reason why the same principles should not hold when the observer is asked to deal with more complex issues such as the judgement of emotions and personality traits. The ideas and concepts upon which the perception of people is based have to go through the same processes of being sorted out and fitted into a frame of reference in order to be used by the observer The application of cognitive theories to the perception of social behaviour and personality forms a significant part of social psychology (e g *Lindzey* (1954)) and here again there are many studies not directly con-

cerned with the psychiatric interview but which illustrate points of common importance

An experiment by *Kelley* (1950) is perhaps a suitable illustration and is of particular interest because it suggested that the expectations about a person to be judged can alter behaviour towards that person, as well as opinions about his personality Kelley informed half of a class of students that a new demonstrator they were about to meet was a "warm" person and separately told the other half of the class that he was a 'cold" person When asked to rate the demonstrator on a number of personality characteristics at the end of the class session, their ratings were biased according to what they had been told beforehand It was also found, although to a lesser degree, that those students who rated the demonstrator favourably in the warm" categories took more part spontaneously in the class activities than the others

3 *Inferences and Decisions*

The setting in which an observation is made, and the presence of definite diagnostic expectations in the mind of the observer, may have an important influence upon the inferences which may be drawn from what is observed Psychiatric diagnoses are particularly vulnerable in this regard, partly because many symptoms are common to different disorders (*Freudenberg* & *Robertson* (1956)), and partly because of the close similarity between some supposedly diagnostic symptoms For instance if it is suspected from an early stage in the interview that schizophrenia will be the final diagnosis, the observation that the patient is rather quiet and slow, with little emotional response, might be regarded as flattening of affect, so confirming the initial diagnosis But to another observer of the same interview who had earlier reached, for other reasons, a diagnosis of a depressive illness, the same behaviour might be interpreted as depressive apathy Thus, depending upon the pre-existing diagnostic set of the observers, the same behaviour is taken to confirm different diagnoses

Inferences about personality traits, the definitions of which are even less well defined and agreed upon than those of psychiatric symptoms, may be made even in the complete absence of perceived evidence in their favour, the "halo effect" and related influences are the chief culprits in this area *Thorndike* (1920) coined the term "halo effect" to refer to a constant error found in ratings of personality traits due to "suffusing ratings of special features with a halo belonging to the individual as a whole" For instance, if a judge is making a number of trait ratings of another person he tends to make first of all an overall decision that the person is generally good or generally bad and then to make the more specific trait ratings so as to fit in with the overall goodness or badness to a much greater extent than

is justified by the evidence presented about the specific traits A very closely related phenomenon which has the same effect of increasing inter-correlation between judgements, is the "logical error" described by *Newcomb* (1931) and *Guilford* (1954) This is due to judges giving a person similar ratings on characteristics which seem to be logically related in the opinion of the judges, i e if a man is judged to be industrious, he may also be called dependable, intelligent and likely to be a good leader because in the opinion of the judge these traits usually go together Other ways in which raters impress their own concepts on the ratings they make of other people were studied by *Levy* & *Dugan* (1960), who emphasised the large extent to which ratings of personality types reflect the perceptual processes of the judges as much as they do the nature of the objects of their judgements The study of *Raines* & *Rohrer* (1954) demonstrated this effect clearly and on a large scale, at least as far as psychodynamic interpretation of psychiatrically healthy young men is concerned A group of experienced psychiatrists interviewed, for twenty minutes each, 886 selected officer candidates, and among other tasks made judgements on each man as to his dominant personality type and his typical defence mechanisms The psychiatrists were found to show statistically significant differences in the frequency with which they observed different personality types in random samples drawn from the 886 men and, in particular, it could be shown from the ratings of 124 men who were interviewed twice that different psychiatrists saw different personality traits in the same man

Sources of spurious correlation between ratings are of very obvious relevance to psychiatric diagnosis, both in the assessment of personality and in the study of mental illness The clustering of symptoms into syndromes is the first stage in the identification of new disorders, whether they are viewed as patterns of reaction or as independent disease processes If the halo and other effects are ignored, there is a considerable risk that artificial symptoms patterns and clusters will emerge The perennial controversy about the nature of depressive illnesses is a good example, and this argument will not be settled until the rating methods upon which the studies are based are properly assessed for the presence of these sources of spurious inter-correlations (*Kendell* (1968))

Many instances of the processes of perceiving and organising that have been discussed up to now are not necessarily regarded as being concerned with conscious selection or decision, the interviewer does not have the subjective experience of making decisions after a period of deliberation. At all stages of the interview, however, inferences and decisions that are experienced as deliberate have to be made upon the basis of the information that has emerged from the initial processes outlined above Decision-making can be regarded as a central feature of most learning situations, and this

has proved a fertile viewpoint is judged by the literature now available on decision theories, problem-solving and the theory of games. Unfortunately the freedom and variability inherent in clinical psychiatric interviewing invalidate any direct comparisons with such investigations, which are usually designed as simple games or single decisions with standard rules and a carefully restricted number of choices (*Messick & Brayfield* (1964), *Shelly & Bryan* (1964)).

Even so, a brief mention of the type of concept used in recent years in this field is worth while, and will serve to illustrate the problems that arise even when the simplest possible type of decisions are being made. Siegel's development of the application of Bayes's Theorem, for instance, emphasised the importance of viewing decisions as a result of the subject maximising both 'subjective probability" and "utility" (*Siegel et al* (1957)). Subjective probability is an estimate by a particular individual of the likelihood that an event will occur. Utility in this context refers to the subjective value or satisfaction a person derives from choosing correctly. The utility of choice variability is less obvious, and refers to the satisfaction a person derives from varying his behaviour. In repetitive-choice experiments, subjects can sometimes be seen to vary their choice among alternatives rather than make the same correct choice repeatedly, even though this means fewer correct decisions. If a purely subjective reason for a choice can be significantly powerful under the conditions of an experiment which militate strongly against its very existence, then similar forces must be expected to be of great importance under the conditions which usually obtain in a clinical interview. Here, every psychiatrist has his own individual concepts and expectations by which probabilities about replies and behaviour will be judged, and there is no immediate way of calling his choices correct or incorrect.

With the growing use of electronic computers in the recording and processing of all types of information, it seems inevitable that the somewhat haphazard decision processes that are responsible for diagnostic decisions will be increasingly compared with and challenged by the more standardised and repeatable decisions that can be carried out by computer programmes (*Baron & Fraser* (1965)). Bayes's Theorem, which is concerned with the probability of a hypothesis before and after the presentation of new data, has been described as "an optimal model for the revision of opinion in the light of new information" (*Edwards et al* (1965)), and as such will presumably feature more and more in the future. It already seems clear that under the conditions of simple experiments involving the estimation of probabilities of sequences, men are inefficient and very conservative compared to a Bayesian analysis (*Edwards et al* (1965), *Edwards et al* (1963)).

Unfortunately as pointed out by several of the contributors to the valuable Michigan Symposium on Diagnosis (*Jacquez* (1964)) it is necessary to know

something about the prevalence of symptoms in the population under study before Bayes's Theorem can be used as a statistical model This is a grave drawback in the case of psychiatric disorders A more promising approach is to use a logical decision—tree model, which is similar to the way in which many clinicians make decisions about diagnoses A series of questions is asked, each of which is either true or false, and the result determines the next question to be asked, while at the same time ruling out one or more diagnoses or groups of diagnoses A computer programme incorporating this method can be used to give diagnoses in which the decisions are necessarily free from inter-oberver differences The final quality and meaning of the diagnoses, however, will always depend upon the quality of the input data Spitzer and his colleagues (*Spitzer & Endicott* (1967)) have made a promising start in the psychiatric field with their "Diagno" programme

There is one recent study by *Gauron & Dickinson* (1966) which is of considerable interest and probably unique, in that it is directly concerned with elucidating to what extent psychiatrists base their decisions upon different items of information Unfortunately, as the authors acknowledge, their presentation of information to the psychiatrists was far removed from the usual clinical setting The information contained in three case histories was broken up into 36 separate and independent categories, each psychiatrist being studied (12 in all) was given a list of the headings of the 36 categories (such as age, sex, family history of mental illness, results of projective testing) and then requested one category at a time in any order they chose After each item of information, they were asked to make a decision about probable diagnosis or diagnoses, and to give a rough estimate of the probability that each one of the diagnoses mentioned would be their final diagnosis They were instructed to continue asking for items until they were satisfied with one final diagnosis When this point was reached, the psychiatrists were also asked to say in their own words why they had reached that diagnosis, and to list the five most important information categories used The decisions recorded after the request of each item were used to rank the categories of information in "actual" importance, and the opinions recorded after each final diagnosis had been made were used to determine "perceived" importance Although "reasons for referral" headed both these rank order lists, it was clear that the psychiatrists were not aware which of the other categories were influencing them the most. For instance, the patient's age was third in perceived importance, but seventeenth in actual importance, projective test results were seventh in perceived importance but second in actual importance, and childhood history was sixteenth in perceived importance but sixth in actual importance

The position of "projective test results" in the rank orders compounded in this study is worth a further mention They were asked for on the great

majority of occasions and as already noted were second in "actual importance" with regard to the final diagnosis In spite of this, they showed a marked tendency to be one of the last items to be requested, being twenty-first in the ranking list for mean order of request. This combination of positions is open to two very different interpretations, both of which give considerable food for thought It could mean that they were merely asked for at the end to clarify and confirm provisional conclusions that had already been arrived at on the basis of the foregoing main body of information, in which case they could perhaps be dispensed with altogether If, however, the psychiatrists really were relying upon these test results to the extent that seems possible, then a good many of the conventional items of information obtained beforehand were in fact redundant and played no real part in influencing the diagnostic choice

These authors also state that inexperienced psychiatrists asked for less information, made more hasty decisions and relied more upon the immediate mental state than their more experienced colleagues, but no details or figures are given to show the basis of their conclusions

They also gave their impressions of how diagnostic styles might be classified, suggesting that the psychiatrists' behaviour could be best interpreted along two dimensions first suggested by *Mendel* (1964)—one of structured versus unstructured method, and one of inductive-logical versus intuitive-alogical thinking.

Gauron & Dickinson do not state whether their diagnosticians were given any instructions as to what sort of a diagnosis they were required to make. If no indication of this was given, then it is likely that some of the differences in style and approach could be due to different psychiatrists having different goals, some might have been attempting to make only an aetiological diagnosis, others to make a descriptive one dealing largely with the present state, and others to arrive at a more comprehensive label as defined at the start of this discussion. In spite of this and other possible criticisms, this study is of great interest since it at last begins to focus attention directly upon the diagnostician in action.

The decision-making processes of 17 probation officers were studied in a very similar way by *Wilkins* (1964), but only one case history and one decision (as to whether probation should be recommended) was involved, ratings of confidence and ease of decision were also obtained. He found no relation between type of decision and the experience of the probation officers, and concluded that the ways in which they sought and used information were characteristic of the individual officers rather than the nature of the information Even reversal of a provisional decision was not related to any particular item of information

The overall effect of these differences in interviewing techniques, per-

ception, decisions and inferences will be to produce differences in the diagnosis made independently on the same patient by two psychiatrists From those studies which report the findings in detail, it becomes clear that if the judgements are made on smaller and smaller components of behaviour this does not necessarily increase the inter-observer reliability, but often appears to lead to more disagreement For instance, *Kreitman et al*, (1961) found on the whole more agreement between the diagnoses reached by two psychiatrists seeing the same patients a few days apart than between ratings of individual symptoms - Although their pairs of psychiatrists reached quite high levels of agreement for organic disorders and functional psychoses, the agreement on the presence of the symptoms upon which these diagnoses were based, such as the presence of intellectual deterioration, delusions and hallucinations, was much lower. Similarly, in the study reported by *Wing et al* (1967), in which a detailed and itemised standard interview was used on a series of patients seen independently by two psychiatrists one week apart, overall diagnostic agreement was very high indeed section scores for depression, worry, tension, etc had a less high inter-psychiatrist reliability, and the individual items of which the sections were composed often showed quite surprisingly low reliability These findings emphasize the necessity to distinguish between the properties of the many detailed and specific judgements that are made during an interview, and those of the final diagnostic terms chosen at the end as the best available summary There may be the most curious and startling differences in the observation of two different psychiatrists, but they may yet agree upon a diagnosis This is because a diagnosis is essentially a summary of many different observations and as such is likely to be more reliable than any of its components or sub-groups

The fact that a number of currently used overall diagnostic terms can be shown to be highly reliable when used under ideal conditions should not, of course, be taken to mean that there is no need to study the unreliable nature of the individual items of information upon which they are based, for in spite of the possibilities of high agreement in the use of major diagnostic categories, no one would suggest that the disorders coming under such terms are homogeneous groups in any but the most superficial manifestations Before further progress can be made in sub-dividing and validating clinical groupings for purposes such as assessment of therapy and prognosis, it is essential to begin to sort out those items which lead to reliable judgements from those which do not, and to examine the reasons behind such differences

The Choice of a Diagnostic Label

Even if the many sources of variation already considered are avoided,

two psychiatrists will find communication about a patient difficult or impossible unless they use similar diagnostic terms to summarise their opinions. If they use terms from systems of sufficiently different nature and emphasis, however, they cannot necessarily be said to be disagreeing, for they may merely be making complementary statements, each dealing with quite different aspects of the patient or the illness. This lack of comparability of diagnostic schemata is at the present time a major international problem, and *Stengel*'s study undertaken on behalf of the WHO shows the nature and magnitude of the difficulties involved (*Stengel* (1959)). *Stengel* illustrated his comments by collecting together 17 classifications in current use in various countries, and it is worth while reiterating the principal points that emerged from Stengel's study as responsible for the incomparability of different systems, for they also figure in practice as possible reasons for disagreement when using any one system. A common and glaring deficiency is that no glossaries or definitions are usually provided which would give an indication of the intended meanings and implications of the terms listed. In addition, it is usually unclear whether an attempt is being made to classify people or illnesses, and some systems appear to try to do both. Finally, almost all the systems in current use are mixtures of aetiological, descriptive and often interpretative terms, only one of which can be given precedence as the main diagnosis.

When faced with difficulties as fundamental as these, it becomes clear that principles of a simple and theoretically ideal system of classification, as put forward by *Hempel* (1959) are quite unattainable for psychiatry at present. He points out that, ideally, a classification divides a set of objects into subclasses which should be mutually exclusive and jointly exhaustive. Different principles have to be used in specifying psychiatric conditions, even if only descriptive "operational definitions" of illnesses are considered, since the signs and symptoms upon which present-day descriptions are based show a very large overlap, even between conditions such as "psychosis" and "neurosis", which many psychiatrists consider to be fundamentally different (*Freudenberg* & *Robertson* (1956)). It is worth while considering a commonly used classification at this point, such as the I.C.D. (1965), so as to illustrate the several and complex types of decisions involved in its use. The I.C.D. contains groups of illnesses under the headings of "psychoses", "neuroses", and "personality disorders". It seems likely, for instance, that for many psychiatrists, the mere presence of hallucinations or delusions is sufficient to result automatically in a diagnosis of "psychosis" of some sort; in most cases this is a comparatively easy judgement about their presence or absence. To distinguish between the various types of neuroses, however, it is often necessary to rely upon the relative severity of a particular symptom amongst the many which are commonly present in the majority of such

patients, here the deciding factor is relative severity, rather than presence or absence A third type of judgement, that of time duration, may be crucial in making a distinction between some types of neurosis and a personality disorder The judgements of severity and duration instanced above are of a particularly difficult nature, since they require the observer to impose cutting-off points and discontinuities upon what may appear to be continuous data, rather than to merely identify discontinuities which are already there This doubtless makes an important contribution to the unreliability of these diagnostic categories (*Kreitman* (1961), *Zubin* (1967)) Even these simple examples highlight the very complex nature of most current classifications, and the need to separate out different types of judgements This is very closely connected with the second and third points noted by Stengel above, concerned with the distinction between the classification of people or of illnesses, and the mixing up of aetiological, descriptive and interpretive terms in most systems The present patchy and uneven state of psychiatric knowledge is presumably responsible for these problems being so prominent There seem to be no reasons, other than secondary ones such as medical tradition or statistical simplicity why it should be thought desirable to make only one type of statement about a patient In many cases, in order to do justice to the findings, it is necessary to make separate statements about, for instance symptoms, personality and intelligence

This particular point emerged with some emphasis from the study of *Ward et al* (1962) Their study is worth noting in some detail, since it is one of the very few in the literature which attempts to examine reasons for diagnostic disagreements with reference to the system of classification being used A panel of four psychiatrists prepared themselves by several preliminary discussions, during which they reached a consensus regarding the specific criteria for each of the nosological entities outlined in the "Diagnostic and Statistical Manual Mental Disorders" of the American Psychiatric Association The psychiatrists were then randomly paired so that each of a series of 153 patients was seen and diagnosed separately by two different diagnosticians, and subsequently the psychiatrists met to discuss and identify the major cause for their disagreement in the 40 cases where this had occurred The causes for disagreement were divided into three main categories (i) inconstancy on the part of the patient, which was judged responsible for 5 % (2 cases) of the disagreement found, (ii) inconstancy on the part of the diagnostician, responsible for 32.5 % (13 cases) of the disagreement; and (iii) inadequacy of the nosology, responsible for 62.5 % (25 cases) of the disagreement Since these judgements were made by those concerned with making the actual diagnosis, it is possible that an independent judge may have allotted the blame for disagreements quite differently, but it is reassuring to some extent that at least the patients were held responsible

for very little disagreement The major cause of disagreement for which the diagnosticians held themselves responsible was "weighting symptoms differently" (7 cases), other causes in this category were 'different material elicited from the patient" (2 cases) and "different interpretations of the same pathology" (2 cases) The chief inadequacy of the A P A system (responsible for 25 disagreements out of the 40) was said to be the necessity of making a forced choice of a predominant major category, "and the common difficulty here was having to determine the relative predominance of psychoneurotic disorder and personality disorder with both entities present" Other nosological sources of disagreement were unclear criteria (10 cases) and the requirement of unpractically fine distinctions, (3 cases)

Too great an emphasis upon fine distinctions which are difficult to identify and agree upon can be a serious drawback in a system of classification, even if the major divisions are unambiguously defined The presence of too many sub-groups is particularly liable to give the spurious impression of many disagreements at a point in time, or many changes of diagnosis with the passage of time The latter effect was particularly noticeable in a study done on a special group of patients taken from a British cohort study (*Brooke* (1960), *Cooper* (1967)) In a group of patients with a total of four admissions in two years, only 20 % retained the same diagnosis throughout the four admissions, when the I C D 4 digit categories were recorded This rose to 37 % if only the 3 digit categories were used, and when the 3 digit categories were arranged in eight larger groups (Schizophrenia, Affective Disorders, Neuroses, Personality Disorders, Addictive Disorders, Non-senile Organic, Senile Organic, and Miscellaneous), the proportion keeping the same diagnosis throughout rose to 54 %

Simple observations such as these, which re-inforce similar observations by *Ward et al* (1962) and *Kreitman et al* (1961), raise the question of what criteria can and should be used for diagnostic groupings The purpose of most currently used classifications is not stated, but many appear to be attempts to provide an all-purpose classification, based upon a number of different principles For psychiatric purposes, as in general medicine, the most valuable single classification would be based on aetiology, but a lack of knowledge of the aetiology of most psychiatric conditions prevents this Other methods of classifying illnesses and patients can clearly be of value, but a classification can only be used properly for the purpose for which it was designed.

Finally, a glossary or set of definitions and instructions is clearly of over-riding importance, for unless such an aid is provided and used in an agreed manner, different diagnosticians may be unknowingly using the same words to describe different phenomena, or vice versa. The provision of what appears to be an adequate glossary of terms, however, is not in itself

a complete solution The non-technical words and phrase that are necessarily used to define the technical psychiatric terms are always open to different shades of emphasis and interpretation which may turn out to be important diagnostically In addition, different parts of a definition may be given different emphasis by different diagnosticians even though they agree exactly upon the meaning of the terms, and agree upon the presence of the defined symptom Agreement must be reached upon the relative importance and implications of symptoms, as well as upon their presence or absence