

VOLUME 59
NUMBER 4

WHOLE No. 274
1945

Psychological Monographs

JOHN F. DASHIELL, *Editor*

An Investigation of a General Normality or Control Factor In Personality Testing

By

PAUL E. MEEHL
University of Minnesota

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Publications Office

1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

ACKNOWLEDGMENTS

THE WRITER gratefully acknowledges his indebtedness to his adviser, Dr. Starke R. Hathaway, and to Dr. Howard F. Hunt, both of whom contributed many interesting and helpful ideas in conversations regarding the problem investigated.

Special gratitude is due to Alyce I. Meehl for constant encouragement and for invaluable assistance in the collection and analysis of the data.

CONTENTS

	Page
Acknowledgments	iii
I. Introduction and Problem	1
II. Theoretical Considerations Regarding MMPI	7
III. Derivation of the N-Scale	15
Plan of Investigation	15
The "Criterion Normals"	15
The "Criterion Abnormals"	18
The Item Analysis	20
Content of Items on N Scale	22
IV. Reliability and Differentiating Power of the N-Scale	25
Reliability	25
Differentiation of the Criterion Groups	25
Distribution of N for Unselected Normals and T-Scores Based Upon This	27
Deviation of the "College Level" Group on the N Scale	29
Application of N-Scale to a "Test" Group of Normals with Elevated Neurotic Triad Scores	32
Application to Cases with Elevations Other Than on the Neurotic Triad	37
V. Miscellaneous Properties of N and Its Relation to Certain Other Variables	41
N Considered as an Inverted "Lie" Scale in Terms of the Scores of a Group of Clinical Abnormals Showing Normal MMPI Profiles	41
Results of Application of N to Unselected Abnormals	42
The Relationship of N to Age and to the Other Variables of the MMPI	44
Item Overlap Between N and Other MMPI Scales	46
An Examination of the Ten Most Potent Items on N	47

Ten Most Powerfully Discriminating Items of N, with the Response Scored for the More "Normal" Indicated	47
Item Overlap of N with the "Normal" Component of the Humm-Wads- worth Temperament Scale	47
A Minor Experiment in "Blind" Diagnosis Utilizing N	48
VI. Psychological Interpretations of N	51
VII. Summary and Conclusions	59
Bibliography	61

AN INVESTIGATION OF A GENERAL NORMALITY OR CONTROL FACTOR IN PERSONALITY TESTING¹

CHAPTER I

INTRODUCTION AND PROBLEM

THE PRESENT investigation had its origin in several converging lines of thought, beginning with a consideration of Rosanoff's theory of temperament as described briefly in his well-known manual (30), and related specifically to the problem presented by persons obtaining markedly abnormal scores on the Minnesota Multiphasic Personality Inventory² but managing to stay out of the hands of the psychiatrist. Although it is expected on purely statistical grounds that about 1 in 40 people in the "general population" will receive scores more than two standard deviations above the mean on any given scale, still it is of interest to consider such deviates with regard to their psychiatric condition. This is particularly true because some of these scores are so markedly deviant, *i. e.*, four or five standard deviations above the "normal" mean, that they cannot reasonably be looked upon as merely the tail cases of a normal distribution from the same supply as the normals whose statistics define the probabilities in question. Furthermore, it is not possible to dismiss all of these cases as simply "test misses" (not in itself a very descriptive term) when closer acquaintance typically shows them to be characterized by rather abnormal amounts of the traits in question

on clinical grounds. For example, we find a "normal" subject who shows a depression score 3.5 standard deviations above the mean of unselected "normals," who has no language handicap or intelligence deficit which would invalidate his test results, and who complains of feeling blue, as if nothing was worth while, and of being filled with a continual ruminative anxiety. Nonetheless, we find him attending his classes, achieving scholastic success, making adequate social adjustments, and in general behaving in such a way that no psychiatrist would consider him worthy of therapeutic attention unless he had no really ill patients as competitors. On the other hand, one can find numerous patients in the psychopathic unit of the University Hospitals who are diagnosed as psychotic or reactive depressions and whose depression has left them effectively incapacitated for ordinary social and economic interaction to the extent that they are in an institution under psychiatric care, yet with depression scores a standard deviation or more below that of this first case considered. The immediate problem is "How is such a thing possible?"

Rosanoff's theory of temperament, which was psychometrically embodied in the Humm-Wadsworth Temperament Scale, may be briefly summarized at this point. He considers three major components tending toward abnormality as manifested in full-blown psychiatric conditions to be present also in lesser degrees among "normal" individuals. These components he calls the *antisocial* compo-

¹This paper is a revision of the thesis submitted to the Department of Psychology of the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy, March, 1945.

²Henceforth the Minnesota Multiphasic Personality Inventory will be referred to for brevity as MMPI.

ment, characterized by "imposed" morality and a tendency to hysteroid and psychopathic manifestations; the *cyclothymic* component, characterized by mood swings, emotional lability, and a tendency to manic-depressive reactions; and the *chaotic sexuality* component, identifiable roughly with the tendency to schizophrenia. Relatively independent of these three abnormal components he posits a fourth factor which is the "normal," inhibiting, or controlling factor of temperament. The theory is that if one possesses a sufficiently great amount of this general normality or control factor in his make-up, he will be able to stay out of psychiatric difficulties and will successfully restrain any deviation that may exist in one of the others. He admits that the evidence for this theory is at present sketchy and inadequate, but still suggestive enough to warrant its serious consideration and investigation. The lines of evidence which he adduces in its behalf are ontogenetic, pharmacologic, pathological, and data from the phenomena of senile involution.

The ontogenetic finding which suggests the presence of a general inhibiting and controlling factor of personality is the appearance of antisocial, cyclothymic, and chaotic-sexual behavior in children which in most cases is "outgrown," and which when it appears in children does not afford ground for such pessimism as to ultimate outcome as the same behavior does when appearing in the "matured" adult. Rosanoff suggests that this is due to a differential rate of maturation of the (genetic) factors for the normal component in different persons, and that the appearance of such "abnormal" trends in children who turn out to be normal in adult life is due to the greater slowness of maturation of this inhibiting,

controlling factor of temperament. Those in whom the genetic basis for such a factor is basically weaker never grow out of the abnormal manifestations and as adults have a poor prognosis.

In the pharmacologic realm, Rosanoff stresses as especially significant the wide individual differences in reactivity to alcohol. In persons who are normally controlled and inhibited, the effect of paralyzing the higher nervous centers through alcohol is to release very different kinds of behavior. In some persons, chaotic sexuality makes its appearance, *e. g.*, the isolated appearance of homosexual manifestations in superficially "normal" males under the influence of alcohol. Others show antisocial, still others cyclothymic behavior. The hypothesis is that in the normal condition these components are already present in all persons (although in varying degrees), and that the direction of the uninhibited behavior is a function of the relative strength of the ordinarily controlled and inhibited tendencies in either of these three directions.

The argument in the case of data from organic pathology and senile involution is practically identical. In the early stages of diffuse cerebral lesions, before the general mental deterioration of advanced cerebral pathology has supervened, phenomena of a schizophrenic, cyclothymic, and antisocial nature occur. It is common knowledge in psychiatry that although certain kinds of response may be expected more often than others, depending upon the nature of the pathology, nevertheless there is considerable variation in the behavior syndrome which occurs when the central nervous system is ravaged by organic disease. For example, two persons have been adequately controlled and socially adapted

prior to becoming ill. Both individuals become parietic in the late stages of syphilis, and one shows euphoria, expansiveness, and many of the symptoms of an atypical affective psychosis. The other may instead become irascible, paranoid, and hallucinated. "Here again, these observations suggest that there exists in the cerebrum an inhibiting and controlling apparatus; that the functions of this apparatus are among the most vulnerable of all the cerebral functions and are among the earliest to be lost in cases of diffuse cerebral lesions; and that the particular manifestations that appear in a given case, upon the loss of the inhibiting and controlling functions, depend on the particular temperamental components that have already existed in the individual, though in a state of complete or partial latency." (30, pp. 667-668.)

Although this conception of temperament is not widely accepted as such, anyone who works with psychiatrists hears them frequently classifying patients into categories that are at least closely similar in their connotation to Rosanoff's control factor. Thus two patients with a similar picture of monosymptomatic conversion hysteria may be sharply separated on the grounds that one has more integration, is better controlled, "keeps her neurosis in hand" to a greater extent, and similar characterizations.

The problem as originally posed could be stated something like this: Is there a fairly stable trait (variable, characteristic) of persons, with regard to which there exist wide individual differences, that is relatively uncorrelated with abnormal components of personality but which, if present, lessens the likelihood that a large amount of some *abnormal* component will result in psychiatric breakdown?

If there is, how can it be measured? If it can be measured, what correlates does it have, what is its behavioral dynamic nature, and what are the people like who have a lot or a little of it?

The theoretical importance of such a factor is too obvious to require discussion. On the other hand, even if its psychological nature were not elucidated satisfactorily, the practical importance of having a means of measuring it would be considerable. It would constitute another technique for reducing the errors of prediction and classification using a psychometric instrument such as MMPI. It is also conceivable that it would have prognostic and therapeutic significance, for example, in estimating the probabilities of staying well after discharge following psychotherapy. In the industrial and other situations in which inspection of a test profile is expected to yield a prediction of successful adjustment in the prospective situation, any score which would differentiate those persons who have deviant tendencies but can be reasonably expected to remain "adjusted" by some minimal standard from those who with scores no more deviant can be expected to prove inadequate, would be of tremendous advantage. For these several reasons both theoretical and pragmatic the development of such a scale (either of a very general nature or more specifically in respect to some particular test) would be of considerable value.

A survey of the literature fails to reveal any empirical investigations of real relevance and merit, with almost no exceptions. A careful study of articles in the *Psychological Abstracts* under the headings abnormality, adjustment, character, constitution, control, delinquency, frustration, inhibition, neuroticism, normality, personality, persistence, psychoneu-

rosis, psychopathy, rigidity, stability, stress, temperament, volition, and will yields practically no material devoted to this aspect of personality testing, although there appear a number of discussions regarding such a hypothetical factor of personality, under various guises and names. Of greatest (and in fact the only obvious) relevance is the work of Humm and Wadsworth, constructing a temperament scale with Rosanoff's theory in mind (17, 18, 19, 20, 21, 22). Of slight relevance one may mention studies of "persistence" (15), the theory of "frustration tolerance," (31, 32, 33), and Freeman's work on a "psychological plimsoll mark" (8). Freeman's is the only one of the last three which uses a question-answer test to measure the hypothetical factor involved, the test in this case being the "normal" component of the Humm-Wadsworth scale.

This last scale consists of 38 items selected empirically by the authors on the basis of overall differentiation between a group of so-called "normal" persons who were defined so by being company employees adjusting satisfactorily to their work and showing no evidence of abnormal trends in their case studies, and a heterogeneous group of patients committed to state hospitals as the "abnormals."

In the application of the Humm-Wadsworth scale to clinical diagnosis, the assumption was essentially based on the theory of Rosanoff, in that a subject with a deviant score on a given scale, *e. g.*, schizoid-autistic, would not be considered as probably maladjusted to a serious degree if his "normal" component were also high, since the latter would act as an inhibitor or "brake" on the appearance of his abnormal tendency. Whether one considers this to have worked or not

depends upon the evaluation he places on the Humm-Wadsworth, concerning which the evidence is somewhat conflicting (2, 5, 6, 18, 19, 22, 23). At any rate the problem of whether the scale called "normal" measures the hypothetical normality factor of Rosanoff's theory is not solved this easily, since such a set of items might serve the purely psychometric function required by the theory but on an entirely different basis.

The items on the Humm-Wadsworth scale are not psychologically homogeneous in any obvious way. However, for the vast majority of these items two things at least can be said. Firstly, they are scored for "normality" when answered in the direction that a naive person would consider more normal and adjusted by inspecting the item content, though this fact means less than many persons might think when one considers what is found in purely empirical item analyses such as done by Humm and Wadsworth or by the authors of the Minnesota Multiphasic Personality Inventory. Humm and Wadsworth do not give the statistically preferred response for the general population, so it is impossible to say whether the response called "normal" is that of the majority of normal persons. Nevertheless the method used by these authors in item selection was in this case the differentiation of their heterogeneous group of "abnormals" (defined by being institution inmates) from a group of persons rated as normal and making satisfactory vocational adjustment (17, p. 167). So that whether the item response scored "normal" on the Humm-Wadsworth is the response given by the majority of normal persons or not, at least it is known to be the response given significantly more often by normal persons than by abnormal per-

sons considered as a class. Many of the items on the Humm-Wadsworth normal scale are found, in slightly altered form, in the MMPI. Inspection of the scoring of these items on MMPI shows that 88 per cent of the Humm-Wadsworth "normal" items which appear in almost identical form in MMPI are responded to by the majority of the MMPI general population sample in the same direction that is scored "normal" on the Humm-Wadsworth. This fact, combined with the item content, suggests rather strongly that to achieve a high "normal" score on the Humm-Wadsworth it is necessary to avoid saying psychiatrically "bad" things about yourself of a sort which are so general among all varieties of abnormals that they show a differentiation when this latter group is considered without further breakdown. Examples of the H-W "normal" items are such responses as the following:

Have your hardest battles been with yourself? (Yes)

Has more than one person called you hotheaded? (No)

Have you lost out in several undertakings by not making up your mind quickly enough? (No)

Have you several times worked under people who seemed to have things fixed so that they would get the credit for good work, and those under them would be blamed for mistakes? (No)

Have you at times felt obliged to ask your friends to help you personally even though you could not return the favor? (No)

Do you think most people inwardly dislike putting themselves out to help other people? (No)

Is it always best to keep your mouth shut when you are in trouble? (No)

Have you ever taken up a line of work which called for close attention to fine detail, to the exclusion of other activities? (No)

Have you ever stayed away from another person because you feared doing or saying something you might afterwards regret? (No)

Have you ever felt that difficulties were piling up to such an extent that you could not overcome them? (No)

Have you at times been certain that people were talking about you although you couldn't prove it? (No)

Although there are, as always in a test of this sort, some responses of a puzzling and seemingly irrational character scored in a certain way, this list of items (every third item in the H-W normal scale) indicates, I think, the general tendency for the "normal" scored response to consist of a denial of a trait or symptom which if present would indicate maladjustment. This observation may seem utterly trivial but the importance of it will become clear when the interpretation of the results of the present study is discussed.

A further relevant finding is that the so-called "normal" component shows substantial negative correlations with several of the other components as found by three investigators (2, 6, 23), although for some unaccountable reason the coefficients between the normal component and other components are reported as positive by Humm and Wadsworth. The highest negative correlations (ranging from $-.53$ to $-.78$) reported by these other three investigators are with the depression and paranoid scales. The correlation of the "no-count," as would be expected from the list of items in the sample above, is markedly positive with "normal."

It is not here denied that the H-W "normal" component is measuring the hypothetical inhibiting, controlling "brake" factor of Rosanoff's theory. But it is obvious that the differentiation of heterogeneous abnormals from normal, adjusted persons by a given item does not establish prima facie that it is measuring such a component. The scale, the

internal consistency of which is indicated by Humm and Wadsworth's split-half reliability coefficient of .82 (17), might well be measuring some other component of personality than the one hypothesized by Rosanoff and still function statistically in the way suggested by the authors. To mention only two possibilities, it might be measuring the overall freedom from reality-distortion that appears in verbal behavior, so that a general collection of abnormals would score low. Or again, it might measure (as will be suggested on more evidence below) some "test-taking" trait of such a nature that while not itself intimately related to a specific psychiatric syndrome, it would be related to the tendency of persons, whether normal or abnormal, to answer items in a certain way. In which case a high "normal" component occurring in a person who was also high on some pathological component might operate to suggest improved chances of his being clinically normal in the same way the "no-count" does the converse. I do not wish to defend any particular hypothesis regarding the nature of the underlying personality trait sampled by the H-W "normal" items, but it must be understood for what follows that the differentiation of normals and abnormals by a given item is, as it stands, a rather ambiguous fact. So far as I know, the only empirical finding which tends independently to support the Rosanoff theory as an interpretation of the psychological meaning of the H-W "normal" component is the study of Freeman (8)

mentioned above.

In the present investigation the initial attack on the problem was made in terms of one particular personality test, the Minnesota Multiphasic Personality Inventory, which has been described in detail elsewhere (9, 10, 11, 12, 13, 14, 24). In brief summary, this is an empirically constructed set of scales furnishing scores on nine personality variables named as follows: Hs (hypochondriasis), D (depression), Hy (hysteria), Pd (psychopathic deviate), Mf (masculinity-femininity), Pa (paranoia), Pt (psychasthenia), Sc (schizophrenia), and Ma (hypomania), together with three scales used to determine the probable trustworthiness of the other nine. These last three are: ? (number of items answered "cannot say" and thus effectively removed from all scales on which they appear), L (for "lie," measuring the deliberate or unconscious tendency of the sorter to put himself in a favorable light) and F (tendency to give statistically infrequent replies, usually suggesting lack of understanding or haphazard sorting). Before beginning the empirical investigation it was felt desirable to formulate as well as could be done a priori the logically exhaustive alternative ways in which a person could come out with a deviant score on the Multiphasic on some scale and yet avoid being in a psychiatric hospital nonetheless. This requires a preliminary consideration of the theory of the trait employed in work with this instrument, and will be dealt with at some length in the next chapter.

CHAPTER II

THEORETICAL CONSIDERATIONS REGARDING MMPI

IN TERMS of Allport's distinction between a "biosocial" and a "biophysical" concept of a trait, it can be said without hesitation that the attitude prevalent among those who work clinically with the MMPI is clearly biophysical. Although the diagnosis of a psychosis must necessarily be made at the time of decision in terms of the "social stimulus value" of the patient to the psychiatric staff, nevertheless this social stimulus value is not the intended *content* of the diagnosis. The best testimony to this fact is the very frequent occurrence at a case conference of statements such as "He looks to me like a schizophrenic but as you say, he *may not be*." It is regrettable true in the field of psychiatry that precisely what it means to *be* a schizophrenic is by no means clear. For those who are compulsively operational this is a frustrating state of affairs, but I do not see any advantage in the approach which says "We do not yet know the internal state which is a necessary and sufficient condition for belonging to the class of schizophrenics, which we would better define if we could; therefore schizophrenics are all those labeled as such in a psychiatric diagnosis." In some respects the situation is very similar to that which existed in the case of a disease-entity such as "general paralysis of the insane" before Noguchi and Moore found the spirochete in the brains of paretics. If we could now go back and get Wassermann, gold colloid, etc., tests on all individuals diagnosed G.P.I. from 1850 to 1900, we should find a number who have nothing in common with the majority so classified except certain similarities of

symptomatology. We know that if the pathological process is to be the real criterion which specifies an entity, these persons were misclassified. If the intent of the diagnosers of 1850 was to put those persons together who belonged together because they had the same "disease," on these non-luetic persons they are now known to have been wrong. It is true that as long as their entity was only defined by the symptom syndrome, they cannot be said to have been "wrong" in a semantic sense; but inasmuch as the cases were etiologically and pathologically heterogeneous, these former investigators would now be willing to concede that all of the cases should not have been classified together, although at the time they had no way of knowing which ones.

Another analogy can be found in the case of units of measurement in electricity. Some writers may take the coulomb as the basic unit, defined in terms of the deposition of silver at the cathode of a silver salt solution. In terms of this the current can be defined as time rate of passage of amount, the resistance in terms of current for a given EMF, and so on. The manifestations of the passage of an electric current may be considered from the standpoint of such diverse operations as the heat generated in a resistor, the amount of deflection of a compass needle, the rate of deposition of silver, the intensity of illumination of a bulb, and the like. It is fortunate for the physicist that the dependence of these different phenomena on the underlying "real" process is sufficiently close and straight-forward so that he does not have

a serious problem of "choosing" one of them as his *definition* of the amount of current, in order to be able to give an immediate answer to the question "How much current" in those cases in which the results of these diverse operations and observations failed to agree. Suppose now that a Maxwell's demon exists in the silver solution and insists on confounding us by arbitrarily stopping and starting ions in their passage to the cathode. Another demon exerts variable amounts of force on the needle of a galvanometer. Still another interferes in some way with the measurement of illumination from our bulb, and so on through all the operations defining "amount of current." The result would be a failure of these diverse procedures to yield values of the "current" that would be isomorphic with the underlying reality (flow of electrons), and hence also they would fail to be in good correspondence with one another.

Now the point I wish to make is that it is still not necessarily the best scientific procedure for the confused physicist to say "Well, there doesn't seem to be any decent covariation in these phenomena, so I guess there isn't any underlying common class of events involved. Operationally speaking, I have to define what I mean by flow of current unambiguously, so I'll pick galvanometer deflection. In the future, whatever else is observed, whenever the galvanometer is deflected so much, I shall define that as the flow of so much current. Any discrepancies with other facts are just unfortunate, that's all." This approach is positivistically elegant but in my opinion not calculated to bring about the most fruitful scientific thinking, particularly in a field which is in an embryonic state and in which the best criteria for

inclusion of a definition or concept are not even yet known except dogmatically. Such covariation as *does* remain after the demons have done their work might be fruitfully utilized in an effort to get tentative interpretations including interpretations of the failures of correlation, and to improve the apparatus so as to keep the demons out. If the latter can be managed, the "true" underlying flow of electrons will finally be arrived at.

To take a typical example in the clinical use of the Multiphasic, one might consider the case of hysteria. It has long been known by psychiatrists that aside from the presentation of conversion symptoms which define the clinical entity named hysteria, there is a more or less well defined personality structure characteristic of the persons so diagnosed. At the more superficial levels one speaks of the tendency of the hysteric to shut out unpleasant facts from his conscious thinking; the "emotional immaturity" showing up by parasitism, sexual frigidity, overdependence upon parents, selfishness in love relationships; a certain beautiful indifference to the symptoms; Rosenzweig's "impunitiveness" as a reaction to frustration; and so on. A psychoanalyst goes on to characterize even deeper layers of the personality—phallic genital level, etc. "Hysteria" diagnostically refers to the production of certain disease-mimicking symptoms, yet a "hysteroid" temperament is discernible and is in fact utilized in diagnosis, *e.g.*, as between a diagnosis of hysteria and hypochondriasis.

This clinical finding is supported by the empirical findings of test item analysis arrived at in studying the responses to the MMPI of persons so diagnosed. There is in the set of items of the hysteria scale a group referring to somatic com-

plaints, the core trait of the clinical entity as the medical man perceives it. But there is also a set of items reflecting a tendency to deny psychiatric and personality problems, typified by responses such as saying "false" to such items as "I frequently have to fight against showing that I am bashful," "I get mad easily and then get over it soon," "Often I cannot understand why I have been so cross and grouchy," and so on. The fact that this set while self-correlating is slightly negative in correlation with the somatic set does not prevent the sum of the two sets from differentiating hysterical persons better than either one could do it alone. Although persons in general tend to answer these items in a negatively correlated way, yet a person with a clinical hysteria has a tendency to reverse this negative correlation and be a deviate in the same direction with regard to both. It would seem tentatively plausible that the "hysterical" person is characterized by his deviation on several unrelated (or in this case actually negatively related) components of personality. How those components had best be defined can only be found by an empirical investigation of their clinical correlates, the way they hang together, and how to give them the most psychological meaning.

For the present purpose it will be convenient to make a clear distinction, therefore, between the operationally measured score achieved on a *scale* with a certain name on the one hand, and the hypothetical amount of the *trait* possessed on the other. In what follows if the biophysically intended trait is discussed it will be referred to by its name, whereas the score value will be indicated by the initial letters of the scale as is common in present discussion of the MMPI

at this institution. Thus, in the case of hysteria we have first, "hysteria," the clinical entity showing conversion symptoms and the like; secondly, we have hypothesized the "hysteroid component" (or components) which predispose to hysteria but are not identical with the appearance of symptoms; and thirdly, we have "Hy" which is the obtained Multiphasic score. Allport's consideration of the "individual" trait as meaning that every person's hysteroid component is unique and cannot truly be arranged on any such continuum will be theoretically conceded but can obviously not be treated of in measurement procedures of the present nomothetic sort.

The construction of the MMPI has been detailed elsewhere and will only be summarily treated at this time. The procedure has been with trivial exceptions a thoroughly empirical one which implicitly denies the usual assumption that one can deduce the verbal behavior of a certain sort of person in a test-taking situation from a priori considerations alone. Allied with this denial is the attitude that the verbal type of personality inventory is *not* most fruitfully seen as a "self-rating" or self-description whose value requires the assumption of accuracy on the part of the testee in his observations of self (26). Rather is the response to a test item taken as an intrinsically interesting segment of verbal behavior knowledge regarding which may be of more value than any knowledge of the "factual" material about which the item superficially purports to inquire. Thus if a hypochondriac says that he has "many headaches" the fact of interest is that he *says* this, and that fact would not be controverted by any investigations establishing that he has interpreted the word "many" in a statis-

tically rare fashion and actually does not suffer headaches any more frequently than the ordinary person. The now prevalent pessimism with traditional question-answer personality tests because they all require the subject to be both honest and objective in his self-descriptions simply does not apply because it is what he *says* of himself that is diagnostic rather than what is in fact the case.

With this orientation the procedure of construction of the MMPI follows in a straightforward fashion. Consider any homogeneous clinical group for the detection of which a scale is to be constructed. Then the items on such a scale are selected by the empirical determination of response-frequencies for this group as compared with those for persons in general. It may be the case that other groups than the one intended also deviate in their responses to certain items. This is again determined empirically and such items are either eliminated or a new comparison of the group desired with the erroneously discriminated group is made to create a "correction scale" (10, 11). By this technique there have up to the present writing been constructed the nine scales mentioned.

Before embarking on the experimental material in the present investigation it is of interest to formulate the possibilities of a "test miss" in so far as they can be seen a priori. Granted the two simple and well established facts that (1) the scales tend to identify the people for whom they were named, (2) they do not by any means do so infallibly, the question arises "Why does it miss when it does?" I have tried to set down what seems to me an exhaustive list of the ways in which this could theoretically occur, keeping the above distinction be-

tween a scale value and the biophysical trait value in mind. Some of these, especially I-B, are not always sources of error, as should be clear from the preceding discussions, but lead to errors of certain kinds in certain cases.

- I. "Errors of measurement" as regards the single trait score itself. This includes all those cases in which the personality variable underlying the scale responses does not actually exist in the amount indicated by the score.
 - A. The answers given do not correspond to the facts as the patient himself sees them.
 1. Inadequate understanding of the questions due to feeble-mindedness, organic brain disease, reading difficulties, bilingualism, etc.
 2. Failure to cooperate in the task required.
 - a. Deliberate misrepresentation—"lying."
 - b. Responding more or less at random due to boredom, hurry, fear of results.
 3. Unusual interpretation of the meaning of the questions of a sort *other than those interpretations related to the personality trait it is desired to detect.*
 - B. The patient does not see the facts as they are.
- II. The testee actually possesses the underlying trait to the extent indicated by the score, but he fails to manifest it symptomatically and thus appears as a "test miss."
 - A. An external situation or a non-personality trait exists in his life which is especially favorable for adjustment and thus he remains free of symptoms. Family protection, sheltering, financial inheritance, Murray's "gratuities" such as beauty, high intelligence, etc.
 - B. Other compensating personality traits exist.
 1. Other traits which can act to inhibit the one in question. For instance, a potential psychopath is kept out of trouble by a concomit-

- ant psychasthenia that makes him be careful.
2. The hypothetical "general control" factor posited by Rosanoff.

It is obvious that some of the above sources of error are minimized by techniques included in the MMPI itself. For example, the case of deliberate misrepresentation is to some extent at least overcome by the inclusion of the Lie scale and its more recent experimental forms. A marked lack of understanding such as that which might occur in an almost illiterate or foreign-speaking subject will almost inevitably raise the F-score and thus make the interpreter suspicious. It should be clearly understood that the use of categories such as "unusual interpretation" and "not seeing the facts as they are" does not contradict the empiristic verbal-sample attitude previously stated.

Some varieties of unusual interpretation, as in the case of frequency of headaches, are correlates of the trait in which we happen to be interested, and the unusual interpretation is actually a basis for our discrimination. On the other hand, there may be interpretive differences among persons which are correlated with entirely different behavior variables including those which have no appreciable personality relevance. For instance, if one asks the question "How many are *several books*" he gets a considerable variation in the connotation people put upon this quite neutral, non-personological quantifier. Aside from the psychiatric factors which determine such differences in interpretation in MMPI items, this other kind of difference will also operate and obscure the results. It is a superficial answer to say that how a person interprets the word "several" is a reflection of his personality, since it

may be a reflection of a personality trait which is minimally or even negatively correlated with the one for whose sampling the particular item was statistically selected.

It might be advanced against the foregoing considerations that the nature of the statistical methods used to select items is such as to systematically exclude such factors. Naturally the relative frequency of occurrence of this kind of non-psychiatric determiner of unusual interpretations has been made low by the setting of high critical scores in the neighborhood of two standard deviations in a "normal" population which by definition must include those individuals who are psychiatrically normal but make such "misinterpretations" when confronted with the items. Thus it would be contended that all of those factors which are psychiatrically irrelevant but weight the probabilities of a given item response have been taken care of in the setting up of norms. The statistical cogency of this argument may be fully conceded without detracting in any way from the value of investigating those cases who belong in the "missed" group no matter where the critical line is drawn. The point is that we would prefer to reduce still further the probabilities of a miss in either direction, and these combined probabilities are a function of the total overlap of groups as well as of the arbitrary critical score set. Thus, if you want to "catch" more of the actual cases of schizophrenia, you have to avoid missing some of those formerly missed. This means a lowering of the critical score, and immediately a rise in the frequency of normals erroneously called schizophrenic. You have to pay for an improvement in one direction by a loss in another. But if there

were procedures for isolating *among* your "normals" some of the sub-groups who achieve relatively deviant scores without being trait-deviant, these sub-groups could then be systematically identified and the net result would be a reduction of errors for any specified critical level chosen in practice.

The entire problem can be stated succinctly in terms of the probabilities involved in sampling from a collective of a defined character. The relative frequency of a "plus" response for any specified item is determined for the large and heterogeneous collective called "normal" defined by their not being in a hospital or under the care of a physician. If a person belongs to this collective, the probability of his saying "Yes" to an item indicating abnormality is empirically known. Combine several such items scored in a given way, and the probability of a member of the given collective saying "Yes" to all of them is small. Combine a sufficiently large number of such and the probability of his saying "Yes" to all of them may become vanishingly small. For some other group, *e.g.*, the set of all hypochondriacs so diagnosed, this probability is empirically known to be much larger. Now *among* the whole set of "normals" there are various sub-sets whose presence requires us to raise our critical score in order to avoid miscalling them hypochondriacs. Some of them are potential and semi-hypochondriacs, but not all of them. It is in these sub-sets of non-hypochondriacal high-scoring normals that we are interested. Their high scores are a function of some variables we do not at the moment measure, and if we could isolate them among all the normals we would find ourselves with a little sub-collective for which the relative frequency of the

significant responses would be different from normals generally. That they are not a very sizeable minority is necessarily implied by the very definition of a "deviant" score, of course; but they are the important minority who produce the test misses.

To reiterate at this point that "The statistical frequencies take care of all that" is to ignore the fact that sub-sets with diverse relative frequencies may be combined to produce a total collective in which the relative frequency is greater than some of the sub-sets and smaller than others. Any procedure for the identification of the members of such sub-sets will constitute an improvement in the measurement. It is necessary to realize that there is no such thing as "the probability" in the empirical sense of frequencies. If probability is a relative frequency—and in this type of analysis it cannot reasonably be interpreted in any other way—then "the probability" is undefined until a class for which the frequencies are to be determined is specified. With every additional fact one specifies a narrower sub-collective for which the relative frequencies continually oscillate as the membership is reduced. The "true" probability is not to be thought of as unfortunately undetermined for some experimental or technical reason but rather as being simply undefined until the limits of the class are specified. Furthermore, the set of such probabilities all stand on the same footing and cannot be discriminated among as to which is "best." If an insurance actuary wishes to define the probability that I will die in the next year, he may begin with the relative frequency of deaths per year for white married males at my age. If we add the fact that I am of Norwegian ancestry

the probability will change, say for the better. If we add that I have a mitral regurgitation it changes again, for the worse. If we add that I have no cardiac enlargement, it changes again, for the better. If we add that I make a practice of running upstairs, it changes again for the worse, and so it goes. It is a mistake to assume that the successive values reached by such a procedure *correct* the probabilities previously arrived at, since the latter are based upon frequencies for larger classes and hence do not refer to the same quantity. We have a set of probabilities each of which is correct for the class to which it refers, and the isolation of a sub-set for the class for which the relative frequency is different cannot be said to refute the former statement in any meaningful sense.

It may be asked "What probability should be chosen in practice?" The answer to this question is given by Reichenbach (29)—one should choose the most restricted class for which the number of cases is still large enough to yield stable relative frequencies. To quote his example for the simplest case,

"Imagine a class A within which an event of the type B is to be expected with the probability $1/2$; if we wager, then, always on B, we get 50 per cent successes. Now imagine the class A split into two classes, A_1 and A_2 ; in A_1 , B has a probability of $1/4$, in A_2 , B has a probability of $3/4$. We shall now lay different wagers according as the event of the type B belongs to A_1 or to A_2 ; in the first case, we wager always on non-B, in the second, on B. We shall then have 75 per cent successes."

Although this example is taken from the probability of attributes, the principles apply in all essentials to the case of continuous variables.

With this orientation one can see the so-called "validity" scales of the MMPI

in a new light, namely, as procedures for the detection of persons who, although they belong to the "normal" collective, also belong to sub-collectives for which the relative frequencies of certain item responses (and hence of raw scores) are not the same as they are for the normal collective as a whole. If only 4 per cent of unselected "normals" achieve a score above two standard deviations on a given scale, we shall erroneously segregate that 4 per cent if we were to use the test by itself for diagnosis. If among that 4 per cent are some who yielded deviant scores because they did not understand the English language, these persons may be detected by another measurement—the scale F. If they constitute any appreciable proportion of this 4 per cent, the use of the F-scale becomes profitable. Such a use of F for the isolation of a sub-collective with atypical frequencies among the normal in no ways contradicts the overall figure of 4 per cent which remains perfectly correct as a frequency obtained in the appropriate (larger) class.

It is not unthinkable that the further development of clinically useful personality inventories of the question-answer type may hinge in considerable part on the practical application of this line of thought. If, as now appears probable from our evidence, part of the discriminating power of such test items arises from what is essentially a "projective" element in the test-making, we shall do well to retain "ambiguities" with deliberate intent. But there are probably few verbal ambiguities which cannot be settled projectively by the testee as a function of several different factors. Ordinarily only one of these factors is of interest to us. The purpose of "validity scales" and "correction scales" is to

partial out, so to speak, the influence of the others. Horst has discussed this same problem with mathematical rigor under the name "suppression variables" and "clearing variates," as follows (*italics mine*):

"From the point of view of the variables involved, the foregoing treatment suggests a way of looking at the problem of statistical prediction which has not been previously emphasized. The conventional way has been to regard the prediction problem as consisting of two sets of variables, viz., the variables to be predicted, that is the dependent variables; and the variables from which predictions are made or the independent variables.

"The foregoing analysis suggests the utility of regarding the independent variables as consisting of two sets. The one set should have appreciable correlations with the dependent variable. The other set should have negligible correlations with the criterion or dependent variable but appreciable correlations with the other independent variables. For convenience, we may designate the first set of independent variables as the prediction variables and the second set the suppression

variables. These latter have been called "clearing variates" by Mendershausen.

"From a common sense point of view, it is not difficult to explain what is happening in this type of prediction system. In most practical prediction problems, *it is impossible to find useful prediction variables which do not have appreciable components independent of the criterion variable*, so that we are usually predicting part of the criterion and also other components independent of the criterion. What we need in this case is another set of variables independent of the criterion but correlated with those components of the prediction variables which are independent of the criterion. By means of the second set we should be able to suppress the irrelevant components of the prediction variables." (16, p. 434).

The present investigation will be seen from the evidence presented to consist essentially in the empirical isolation of a set of items on the MMPI which act in the role of suppression variables for at least some of the abnormal components, although the behaviorial nature of the components thus suppressed remains largely mysterious.

CHAPTER III

DERIVATION OF THE N-SCALE

PLAN OF INVESTIGATION

IN ACCORDANCE with the usual empirical procedure of scale derivation which has been found fruitful in the development of the MMPI, the following plan of study was employed in the present investigation:

1. A group of cases was selected by the fact that the profiles showed deviations that would ordinarily be labeled "abnormal," in spite of the apparent freedom of the subjects from incapacitating psychiatric involvement. These cases will be referred to hereinafter by the phrase "criterion normals."

2. A second group of cases was selected so that their profiles would match as closely as was practicable the profiles (case by case) of the first group, but this second group was drawn from the hospital population and thus was composed entirely of individuals whose psychiatric involvement was such as to result in their being in a psychopathic unit. This group is referred to hereinafter as the "criterion abnormals."

3. An item analysis was carried out on all 495 items¹ of the MMPI and those items which showed a statistically significant differentiation between the two groups were brought together on one key to form a scale.

4. The central tendency and variability of the resulting scale was obtained for a random group of "normal" persons, and standard scores calculated for further easy reference.

5. The differentiating properties of this scale were studied by systematic ap-

plication to various special groups, such as the original criterion groups, "test" normals and abnormals from diverse sources, college students, WPA workers, persons of different chronological ages, and so on.

6. The internal consistency and test-retest stability of the scale were determined by the usual procedures.

7. The intercorrelation between the scale and the other scales of the MMPI were calculated for both normal and abnormal populations.

8. A minor experiment in the selection of normal from abnormal profiles with and without the scale in question was performed.

9. From an analysis of the preceding findings plus an inspection of the items on the scale, tentative interpretive possibilities are suggested and possible experimentation for this purpose proposed.

THE "CRITERION NORMALS"

The first step was the selection of a group of cases showing abnormal profiles by usual standards but not under psychiatric care. As is the case with all work on the MMPI it was necessary to accept a very crude and superficial definition of "normality." Ideally, one would prefer to be able to have intimate psychiatric knowledge concerning the behavior and history of the persons so designated, but this is obviously impossible for practical reasons if it is desired to have a large and representative enough group to carry out an item analysis. Furthermore, it is necessary to avoid equating "normal" to "well-adjusted" in some optimal meaning of the

¹ Only 495 items were involved because of the original 504 items available on old forms, 9 have been eliminated in the present box.

latter, since what is of clinical importance is in a fundamental sense this very crude brute fact of having managed to stay out of the hands of a psychiatrist. Thirdly, it is desirable in view of the previous era of overoptimism regarding the power of personality tests to define the situation so as to make it as "hard" for the test as possible, an end which is certainly facilitated by the blanket assumption that *all* persons who are not under a doctor's care or in a psychopathic ward are by that token more "normal" than *all* of those who are. In the following discussion "normal" means that the profile is that of a person who at the time of examination was not under a doctor's care.

The entire collection of "normal" cases ($N = 691$) in the records was canvassed in order to locate all profiles showing any of the eight components (Mf has been ignored throughout the present study) with a T-score equal to or exceeding 70.² These 691 normal cases make up the standardization group upon which the other scales have been developed, and their composition has been previously described (9, 10, 11, 12, 13).

Since it was a priori doubtful whether the suppression variables for all eight of the abnormal components would turn out to be the same, the original item analysis was confined to those normals from this group who showed elevations ($T \geq 70$) on any of the three scales of the so-called "neurotic triad" (Hs, D, or Hy). In the event that the same suppression variables might operate in the case of other scales, this fact would be subsequently elicited; whereas if such

were not the case the confounding of factors due to pooling several varieties of deviant-scoring normals might result in a failure to locate discriminating items. The first group studied, therefore, consisted of persons scoring $T \geq 70$ on at least one (and possibly more) of these three scales, but no systematic effort was made to avoid cases in which some other scale was also elevated. The exclusion of all but "pure" neurotic triad profiles would have reduced the number available for study beyond a workable limit.

From this preliminary pool of deviant scoring normals a smaller group was selected by the application of certain important restrictions regarding the probable validity of the profiles. Because it was clear already that the number of cases for analysis would be quite small, the restrictions regarding probable validity were made fairly rigorous. In addition it was considered that to loosen these restrictions might result in a mere duplication of the F-scale, a suppression variable already extant. Accordingly the following two validity restrictions were imposed in the selection of a smaller set from the group consisting of all normals showing any $T \geq 70$:

1. In order to make practically certain that no case was included in which the elevation of profile was due to carelessness or lack of understanding, it was required that the F-score of any case accepted should not exceed $T = 60$. It was recognized that this would exclude an unknown proportion of individuals whose F was more elevated not because of non-cooperation or misunderstanding but because of validly "unusual" responses reflecting the same personality trends as the elevated profile suggested. But for the above reasons it was felt preferable to lose this group than to in-

² All MMPI scales are expressed as T-scores,

$$\text{where } T = 50 + 10 \frac{X - \bar{X}}{SD}$$

clude cases where cooperation or understanding was doubtful or in which the subject had become careless toward the end of the testing period.

2. In order to avoid an excessive frequency of "Cannot say" responses in doing the item analysis on so few cases, as well as to exclude cases of doubtful validity indicated by such an excess, no case accepted showed a "?" score of $T > 64$.

No restrictions were placed upon L (lie) because the error introduced by including persons who consciously or unconsciously put themselves in a favorable light would be opposite to the error one wished most to avoid—that of including cases in which the deviant profile was spuriously *high* due to some factor or other.

The imposition of these restrictions reduced the total number of cases for study to only 42, of which 17 were males and 25 were females. No effort was made to make the sample "representative" with regard to such variables as age, sex, or socio-economic status, since this would have further reduced its size. The ages ranged from 20 to 45 with a mean age of 32.9 years (32.2 for the males and 33.3 for the females). Except for such minimal comments as were written on the record sheet attached to each test, no further material was available regarding these persons. With the possible exception of a self rating by the subject as to whether he considered himself "nervous" or not, this additional information was of no value. Since this latter self rating is essentially a condensed variant of a great mass of such self judgments contained in the test itself it also is of little value, although it is worth noting that 18 of the 42 cases stated that they considered themselves to be "nervous."

TABLE I
Mean T-Scores on All Scales for the 42
Criterion Normals

	Both Sexes * (N=42)	Males (N=17)	Females (N=25)
?	52.0	52.3	51.7
L	53.0	51.5	54.0
F	53.0	52.9	53.0
Hs	66.1	66.9	65.5
D	66.8	66.1	67.3
Hy	64.1	63.3	64.6
Pd	56.5	55.2	57.4
Pa	55.4	55.6	55.3
Pt	60.6	60.6	60.6
Sc	56.4	55.4	57.1
Ma	48.5	47.5	49.2

The mean T-scores for these 42 cases as well as for the two sexes separately are presented in Table 1.

It will be noted that although every one of these cases was selected for showing either Hs, D, or Hy with $T \geq 70$, the mean profile is within the limit usually called "normal," since individuals do not tend to exceed 70 on all three in most cases. Twelve cases show scores ≥ 80 , the other 30 scoring between 70 and 79. Although the profiles were selected on the basis of the neurotic triad and apparent validity only, one notices that with the exception of Ma for males, all scales show a mean above the population average 50, and the general profile pattern is quite similar for the two sexes. The three validity indicators are very comfortably low.

We have here, then, a group of 42 persons who were cooperative and understood their task at the time they took the inventory;³ who score two or more standard deviations above the mean of the general population on one or more of the eight abnormal components; and who were not under the care of a physi-

³From now on the term *valid* as applied to profiles will be used in this restricted sense, and has no connotation of "agreement with other criteria."

cian at the time. While it is true that these are merely the deviates of the "normal" population on which the T-scores were derived, the important thing to keep in mind is that persons whose profiles are no more deviant than these can be found in the hospital with an incapacitating psychiatric involvement. It is in comparison with these latter persons that the present group of "normal deviates" are of interest. The dichotomy "normal" and "abnormal" may be artificially made sharp by the definition in terms of "in or out of a hospital." The dividing line in the case of the measured (scale) component is even more arbitrary, and although the problem has been stated in terms of the critical score 70, it is clear that if there were no "abnormals" with scores as low as this the occurrence of a small group of scores among normals which exceed this critical line would merely exemplify a statistical truism and the group involved would have little practical interest for us. As long as the analysis was ultimately to be carried through on matched groups in any event, one could probably have chosen some other arbitrary value as critical for selection of profiles, either higher or lower than 70.

THE "CRITERION ABNORMALS"

The 42 profiles obtained from the criterion normals were then considered one at a time with reference to the profiles of the whole available population of hospital cases. Each "normal" profile was matched as closely as possible for all 8 components, as to sex, and as to age (always within 10 years). In spite of the laborious canvassing of over 400 abnormals which was repeated with each of the 42 criterion normals, it was of course impossible to achieve an ex-

tremely close match to each criterion normal when eight variables in addition to age and sex were involved. Most attention was paid to the three scales of the neurotic triad, but all eight scales were taken into account to some extent. Organic cases were excluded.

As regards the validity indicators, it was necessary in the case of the criterion abnormals to eliminate "lying" or at least that aspect or type of lying which is detected already by the L scale. For this reason no case was included in the criterion abnormal group if its L-score exceeded 60, with one exception which was unavoidable if any reasonable matching was to be made on the other variables. The "?" was not allowed to rise above 56, nor the F above 62. (These limits were not chosen arbitrarily to begin with, but are the upper limits finally obtained when matching was completed with the importance of these three variables being kept in mind during the matching process.)

As to clinical diagnosis, all 17 of the male matched abnormals were diagnosed psychoneurosis, as were all but three of the 25 female matched abnormals. These three cases were diagnosed manic-depressive manic (questionable diagnosis, atypical case); psychopathic personality with hysteria and anxiety; and paranoid condition with recurrent depression.

The ages of these criterion abnormals range from 19 to 47 with a mean age of 32.3 years (31.9 for men and 32.5 for women).

The mean T-scores for the 42 criterion abnormals as well as by sexes separately are presented in Table 2, together with the corresponding scores for the 42 criterion normals with which they were matched.

These results are graphically repre-

TABLE 2
Mean T-Scores for the 42 Criterion Abnormals, Sexes Separately and Pooled, Compared with Those of the 42 Criterion Normals with Whom They Were Matched

	42 Criterion Normals	42 Criterion Abnormals	17 Male Abnormals	25 Female Abnormals
?	52.0	50.7	50.2	51.1
L	53.0	52.9	52.4	53.3
F	53.0	51.8	52.0	51.6
Hs	66.1	62.9	61.4	63.9
D	66.8	68.3	69.9	67.2
Hy	64.1	65.9	63.4	67.5
Pd	56.5	53.7	52.5	54.5
Pa	55.4	55.7	56.1	55.4
Pt	60.6	55.5	56.1	55.0
Sc	56.4	52.9	53.5	52.5
Ma	48.5	51.3	52.4	50.6

sented in the accompanying profiles. (Fig. 1) It will be noted that the most marked difference in both sexes occurs in the case of Pt (psychasthenia), in which the normals are about one-half a sigma above the abnormals. This difference may be partly the reason for the correlation later found between the derived N scale and Pt. On the other

hand, it is difficult to see how such a systematic trend arose as a result of the matching process unless there were some factor in or associated with the Pt scale which contributes to "normality" in the presence of elevated scores on the neurotic triad.

Of the eight differences between means of the criterion groups, three are statistically significant at the 5 per cent level. Pt is significantly higher for the normals ($P < .001$), Pd is significantly higher for the normals ($P < .01$), and Sc is significantly higher for the normals ($P < .01$). It is worth noting although inexplicable that in all three comparisons the criterion normals have the more "abnormal" (elevated) score.

The differences show that the matching procedure, in spite of the laborious profile-by-profile technique employed, was not completely successful in eliminating systematic differences between the two groups. The matching was as good as could be reasonably expected under

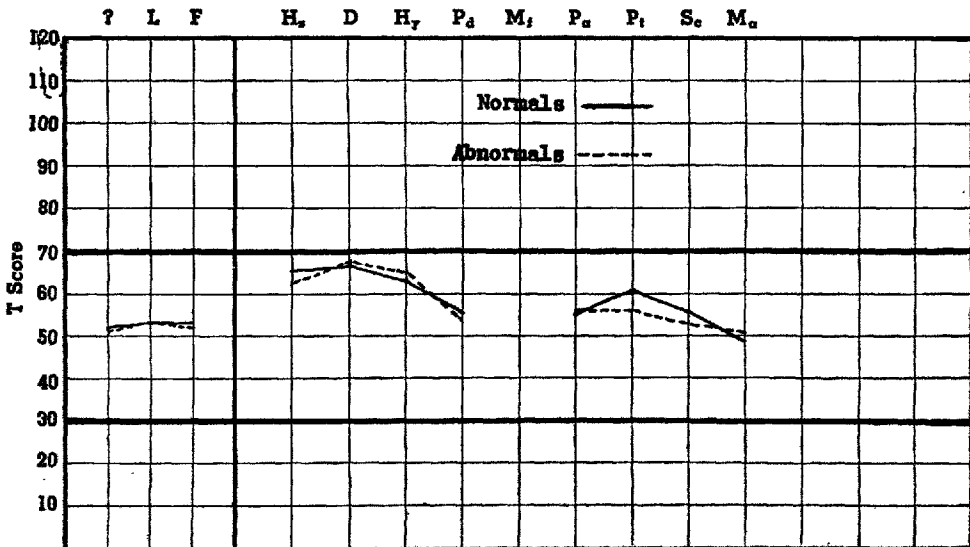


FIG. 1. Mean profiles of the 42 criterion normals and the 42 criterion abnormals with whom they were matched.

these circumstances, and the differences, while statistically significant in three instances at least, are not in the direction which would be serious for the present investigation. Correlations subsequently found with other scales or random samples of normals and abnormals are made ambiguous to interpret to some extent because of these differences in the original groups used in item selection. Nevertheless it must be clear that the correlations cannot be summarily dismissed as being due to this matching factor, since it is necessary to explain the systematic matching difference which can hardly have been due to any bias on the part of the matcher. A "correlation" of variables in the supply might result in such systematic differences in matching equally easily as the converse.

THE ITEM ANALYSIS

We now have two groups of valid individual profiles, one being from a set of persons who exhibit "deviant" MMPI profiles in terms of the arbitrary critical score 70 but are not in a psychiatric ward or under a physician's care ("criterion normals"); the other from a set of persons with valid and no more deviant profiles than the first group but in the psychopathic unit because they have not been able to make an adjustment outside ("criterion abnormals"). The next step is to determine whether there are any items which show a significant differentiation between these groups. If any such normality or suppression variable as has been provisionally hypothesized should actually exist, it seems a priori likely that at least some among the 495 items of verbal behavior might be at least in part a function of this variable. It is not to be expected that a scale composed of items so selected will

be either behaviorally or mathematically homogeneous. At the present stage of analysis it would be foolish to insist that it should be. More refined statistical procedures can be applied subsequently if the end of factorial purification is deemed worthy, although it is not a self-evident proposition that such is to be desired in any case. Any "homogeneity" of the first scale derived by this method consists precisely in the method of its derivation, namely, that it tends to distinguish persons who get abnormal scores and are abnormal from those who get equally abnormal scores but are normal.

In scoring the MMPI a "significant" response to any given item is defined as that response (whether "Yes" or "No") which is made by a minority of the standardization group. These responses are indicated with a red *plus* (+) on the answer sheet and are also commonly referred to as "plus" responses. In what follows the terms *significant* and *plus* will be used hereafter in this special sense, whereas the term "abnormal" will be used to designate the response characteristic of the criterion abnormals in the present study. As will be seen later, the two indications are not in perfect agreement by any means, rather the reverse being true.

After tallying and summing the "significant" responses for each of the 495 items for both criterion groups, these sums were then converted to proportions, giving the per cent significant responses to each item by the criterion normals and by the criterion abnormals. All items for which the difference in proportion of significant responses between normals and abnormals reached or exceeded .20, which is approximately at the 5 per cent level of significance in

samples of 42 cases, were examined further by more exact tests. In a few cases items in which the proportions were very close to .00 or 1.00 were further investigated even though the differences were less than .20 because of the greater statistical significance of differences among proportions approximating these limits.

The preliminary survey yielded 103 items which showed differences in the neighborhood of .20 between proportion of plus responses for the two criterion groups. The significance of the differences was then determined more exactly by the t-test for significance of difference of two proportions, using the formula:

$$t = \frac{p_1 - p_2}{\sqrt{\frac{N_1 p_1 + N_2 p_2}{N_1 + N_2} \left(1 - \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2} \right) \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

All items for which the value of *t* from this formula equaled or exceeded 1.96 were considered significantly differentiating.

It is true that this formula, while an improvement over the older formula for the standard error of a difference of two proportions, still involves approximations when it is applied to proportions in the neighborhood of .00 and 1.00, for which the sampling distributions are so markedly skewed except for very large values of *N*. However, here also the error introduced is in a harmless direction, inasmuch as for such very deviant values of *p* the above formula for testing significance tends to underestimate the real significance of a difference, so that one is much more likely in such cases to miss a significant item than erroneously to segregate an actually non-differentiating one. In the case at hand, we are not especially desirous of locating all the

items which show "statistical" significance when they are answered in the same direction by practically every member of both populations, for in such instances the "significance" arises merely from the deviant values of the proportions and the effective discriminating power of such an item when included in a scale is hardly worth the effort of scoring it.

An examination of the differences between proportions for those items showing *p* near .00 or 1.00 using the normalized C.R. of Zubin's nomographs (34) shows that not more than six items would have been included had this more liberal test been employed.

On applying the criterion of signifi-

cance to the original set of 103 items, a smaller group of 78 items was finally selected which met this criterion. Of these 78 items, 9 show differences 3 times or more their standard error; 62 show differences between 2 and 3 times their standard errors; and the remaining 7 items show differences between 1.95 and 2.00 times their standard errors (just reaching the 5 per cent level). Of the 7 items with a C.R. between 1.95 and 2.00, 4 are significant for one sex but not for the other. But in these cases the differences are in the same direction for both sexes and not less than .12 in any case. This set of 78 items, scored as indicated in the next section, constitutes the scale hereinafter referred to as *N*.

The initial "N" was originally chosen as a reference to "normal," having in mind the theory of Rosanoff previously mentioned and the analogous scale called *N* by Humm and Wadsworth. It seems

rather clear from what follows later that this N is "normality" only in the sense of a suppression variable or correction scale, and not in the original sense of a positive personality variable of inhibitory or "controlling" nature.

The 78 items on the N scale are as follows, the letter T or F indicating the direction in which the criterion *normals* tend to respond as contrasted with the criterion *abnormals*. In other words, those persons who show deviant profiles on the other MMPI scales but have remained in the community out of psychiatric hands, have a greater tendency to give the indicated response than those of similar profile but in the hospital with a psychiatric disturbance.

CONTENT OF ITEMS ON N SCALE

(These items are listed according to their numbers as given in the back of the revised manual of directions, 1943. Indicated answer is more characteristic of the criterion normals.)

- A-15 I am troubled by attacks of nausea and vomiting. (F)
- A-20 I have had attacks in which I could not control my movements or speech but knew what was going on around me. (F)
- A-30 Peculiar odors come to me at times. (T)
- A-32 I can read a long while without tiring my eyes. (F)
- A-39 I have no trouble swallowing. (F)
- B-9 I am almost never bothered by pains over the heart or in my chest. (F)
- B-51 I have been quite independent and free from family rule. (F)
- C-7 My relatives are nearly all in sympathy with me. (F)
- C-21 I hate to have to rush when working. (T)
- C-25 I have often lost out on things because I could not make up my mind soon enough. (T)
- C-33 It takes a lot of argument to convince most people of the truth. (T)
- C-34 When I am feeling very happy and active, someone who is blue or low will spoil it all. (T)
- C-35 I liked school. (F)
- C-46 The only interesting part of the newspapers is the funnies. (T)
- D-9 Religion gives me no worry. (F)
- D-13 I feel sure there is only one true religion. (T)
- D-15 I believe there is a God. (T)
- D-17 I believe in a life hereafter. (T)
- D-18 I believe in the second coming of Christ. (T)
- D-20 The only miracles I know of are simply tricks that people play on one another. (F)
- D-52 I think most people would lie to get ahead. (T)
- E-11 At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it. (F)
- E-13 Policemen are usually honest. (F)
- E-15 It wouldn't make me nervous if any members of my family got into trouble with the law. (F)
- E-24 I prefer to pass by school friends, or people I know but have not seen for a long time, unless they speak to me first. (T)
- E-27 At parties I am more likely to sit by myself or with just one person than to join in with the crowd. (T)
- E-28 I love to go to dances. (F)
- E-49 It does not bother me that I am not better looking. (F)
- E-51 I do not like to see women smoke. (T)
- E-52 People often disappoint me. (T)
- F-4 I am easily embarrassed. (T)
- F-6 I am not unusually self-conscious. (F)
- F-7 What others think of me does not bother me. (F)
- F-8 It makes me uncomfortable to put on a stunt at a party even when others are doing the same sort of things. (T)
- F-13 I have sometimes stayed away from another person because I feared doing or saying something that I might regret afterwards. (T)
- F-17 I feel unable to tell anyone all about myself. (T)
- F-28 People have often misunderstood my intention when I was trying to put

- them right and be helpful. (T)
- F-31 I easily become impatient with people. (T)
- F-34 Criticism or scolding hurts me terribly. (T)
- F-47 Often, even though everything is going fine for me, I feel that I don't care about anything. (T)
- F-48 I very seldom have spells of the blues. (F)
- F-53 I have often felt badly over being misunderstood when trying to keep someone from making a mistake. (T)
- G-1 I feel anxiety about something or someone almost all the time. (T)
- G-3 I believe my sins are unpardonable. (F)
- G-14 I tend to be interested in several different hobbies rather than stick to one of them for a long time. (T)
- G-17 I like to keep people guessing what I'm going to do next. (T)
- G-27 I am often said to be hot headed. (T)
- G-28 I am not easily angered. (F)
- G-29 I get mad easily and then get over it soon. (T)
- G-31 At times I feel like smashing things. (T)
- G-32 At times I feel like picking a fist fight with someone. (T)
- G-33 I have the wanderlust and I am never happy unless I am roaming or traveling about. (T)
- G-38 When I leave home I do not worry about whether the door is locked and the windows closed. (F)
- G-40 In walking I am very careful to step over sidewalk cracks. (T)
- G-45 Often I cross the street in order not to meet someone I see. (T)
- G-49 When someone says silly or ignorant things about something I know about, I try to set him right. (F)
- H-2 I have often thought that strangers were looking at me critically. (T)
- H-18 If given a chance I could do some things that would be of great benefit to the world. (T)
- H-28 I get anxious and upset when I have to make a short trip away from home. (T)
- H-34 Lightning is one of my fears. (T)
- H-35 A windstorm terrifies me. (T)
- H-36 I am not afraid of fire. (F)
- H-41 I have no fear of water. (F)
- H-43 I do not worry about catching diseases. (F)
- H-46 I am afraid to be alone in the dark. (T)
- H-47 I am afraid when I look down from a high place. (T)
- H-52 I have no fear of going into a room by myself where other people have gathered and are talking. (F)
- I-5 Horses that don't pull should be beaten or kicked. (T)
- I-10 The future is too uncertain for a person to make serious plans. (T)
- I-11 It is great to be living in these times when so much is going on. (F)
- I-14 I frequently ask people for advice. (T)
- I-15 My plans have frequently seemed so full of difficulties that I have had to give them up. (T)
- I-17 I wish I could get over worrying about things I have said that may have injured other people's feelings. (T)
- I-18 I often must sleep over a matter before I decide what to do. (T)
- I-30 My way of doing things is apt to be misunderstood by others. (T)
- I-38 I often think, "I wish I were a child again." (T)
- I-40 I am entirely self-confident. (F)
- I-41 Once in a while I think of things too bad to talk about. (T)

Of the whole group of 78 items, it should be noticed that by far the majority of them are answered in the "significant" or "plus" direction by the criterion normals more than by the criterion abnormals. There are only 14 exceptions to this trend, namely, items A-15, A-20, C-21, D-15, D-17, D-18, E-11, E-15, E-51, G-3, G-14, I-5, I-14, J-41. Furthermore, it is evident that *in the great majority of the items the response most characteristic of the criterion normals is not only the statistically unusual one but is the one which on face inspection of the item would be considered the "bad" or "un-*

adjustive" response. Although this kind of observation is of less value than is sometimes believed in tests of this sort, nevertheless the very marked tendency for the criterion normals to say "bad" things about themselves more than the matched abnormals is of considerable interest.

A scoring key was next constructed in which the direction indicated above as characterizing the criterion normals more than the criterion abnormals was the scored response. Thus, if a subject sorts card A-15 into the "False" category he receives one raw score point for N ("nor-

mality"); while if he sorts card A-30 into the "True" category he receives one raw score point. High raw scores on N, therefore, indicate "normality" whereas low raw scores indicate "abnormality," at least as far as persons having deviant profiles are concerned. In what follows, the raw N-score of a person, therefore, means the number of responses he made to the N items which agree with the response indicated by the letters T and F in the foregoing list—the number of times he responded in the direction preferred by the criterion normals.

CHAPTER IV

RELIABILITY AND DIFFERENTIATING POWER OF THE N-SCALE

RELIABILITY

FOR 50 normal males whose mean and SD were practically identical with that of the sample of "men in general" on which T scores were subsequently based, the odd-even reliability of N was .81, which indicates a fair degree of internal consistency of the items. For 50 normal females of high school age retested at intervals varying from 4 to 14 months, the test-retest reliability was .74, which is somewhat lower than the typical test-retest reliabilities reported thus far for MMPI scales, although these previous reliabilities were based upon a different sample.

In the case of personality measurement the desirability of high test-retest reliabilities is dependent in part upon whether "function variability" itself is great or small. If the hypothetical "control" factor of Rosanoff were involved, we should require any alleged measure of it to show a fairly good retest stability. Since we do not know the psychological nature of what is measured by N, it is not possible to say to what extent the coefficient of .74 is lowered by test "unreliability" and to what extent it would have to be lowered to some degree below unity by a function fluctuation which if not detected by the scale would be indicative of the latter's invalidity.

In the case of split-half reliability the same problem presents itself as was previously referred to in connection with the dynamic inhomogeneity of the items on the Hy scale. Further "purification" of N will have to be undertaken for other reasons, but not simply to increase the internal consistency which may or may not

be valued, depending upon the practical situation.

DIFFERENTIATION OF THE CRITERION GROUPS

When the scoring key for N had been constructed the first step was to score the criterion cases on the whole key to determine whether or not the scale was able to differentiate them sufficiently well to justify applying it to other test groups. That the differentiation of the criterion groups should be good is only a necessary and by no means a sufficient condition for accepting such a scale, since the most important test lies in its application to "test" cases, other than those cases upon which the item analysis was made.

The distribution of scores for the criterion normals and criterion abnormals is discussed in the next section. It will be noted that the separation, while by no means perfect or as clean as could be desired, is certainly significant and the overlap not excessively great. It is necessary to avoid undue leniency toward the instrument being studied by attempting to "explain" all exceptions to the rule desired, so that the overlap that exists must be taken at its face value in evaluating the separation. Nevertheless, it should at least be noted that it would be of considerable interest to have information regarding the adjustment of those persons who constitute the low tail end of the distribution of criterion normals. For this reason it is perhaps not out of place to refer to the meager data available on the information sheets attached to their records, from which we learn that the three male criterion normals

TABLE 3
Comparisons of Means and Variances by Sex and Psychiatric Status
A. Tests for homogeneity of variance by sex and psychiatric status

	N	d.f.	V	F	P
<i>42 criterion normals</i>					
Sex group:					
Male normals	17	16	120.6518		
Female normals	25	24	100.0833	1.206	$P > .05$
<i>42 criterion abnormal</i>					
Sex group:					
Male abnormal	17	16	68.7330		
Female abnormal	25	24	98.8433	1.437	$P > .05$
Criterion group:					
42 normals	42	41	108.5271		
42 abnormal	42	41	89.9891	1.206	$P > .05$
B. Significance of difference of means of males and females					
	Mean	SD	Diff	t_{diff}	P
Normals:					
17 male normals	42.1765	10.656			
25 female normals	45.4000	9.802	3.2235	.985	$.30 < P < .40$
Abnormals:					
17 male abnormal	22.8824	8.043			
25 female abnormal	27.5200	9.741	4.6376	1.627	$.10 < P < .20$

having the lowest raw scores on this distribution (20, 21, and 27) were all unemployed at time of testing. Of the three female criterion normals showing raw scores of 26, one is a college student (the significance of which will be elaborated subsequently), another is unmarried at age 30 and has a deviant church affiliation. All three of these female normals consider themselves "nervous" according to the record sheet.

More precise statistical tests of the sig-

nificance of difference between the groups show that the variances of the males and females do not differ significantly, nor do their means, within either the normal or abnormal group. Nor does the variance of the normals differ significantly from that of the abnormal. (Summary of data in Table 3.)

Consequently the sexes were pooled and the t-test applied to test the significance of difference between means of the normals and abnormal. The mean

TABLE 4
Comparison of Means on N Scale for the 42 Criterion Normals
versus the 42 Criterion Abnormals

Group	N	Mean	SD	Mean _{diff}	C.R.	P
Criterion normals	42	44.0952	10.293			
Criterion abnormal	42	25.6429	9.373	18.4523	8.487	$P < .001$

score of the 42 normals was 44.10, with a SD of 10.293; the mean score of the 42 abnormals was 25.64, with a SD of 9.373. The test of significance between these two means shows $t = 8.48$, which, with 80 degrees of freedom, is clearly significant ($P < .001$). Data for these various comparisons are shown in Table 4.

Inasmuch as the difference is taken between matched groups, a more appropriate test of significance is the t based upon differences between matched members of each pair. When this analysis is made, we find that the best estimate of the SE_{diff} is 1.5167 which yields t_d of 12.42. Of the 42 pairs involved, in only one case did a given normal have a lower N-score than the criterion abnormal with whom he was matched. This was a female college student whose case was mentioned above in regard to overlap.

The median of all 42 normals is 46, which is exceeded by one case from the abnormals. Considering the sexes separately, no male abnormal reaches or exceeds the median of the male normals; and one female abnormal reaches or exceeds the median of the female normals. The inference from these findings was that the N-scale had sufficient differentiating power on the criterion groups to warrant further study on other groups.

DISTRIBUTION OF N FOR UNSELECTED NORMALS AND T-SCORES BASED UPON THIS

It was necessary to make a preliminary standardization in terms of T-scores even though it was realized that the scale would not be used in its first form. This latter was required for convenient interpretation when it was found that although the sex differences are not significant in the criterion groups (although, it will be noted, in the same direction

among both normals and abnormals) there were significant differences between the means of men and women sampled at random from the general population of "normals."

Inasmuch as the T-scores to be derived were merely for convenience in the present study and would not be used in practical work, it was sufficient to base the standardization on a smaller number of cases than usually is employed—in this case upon 100 males and 100 females. The standard error of the mean given by 100 cases is only about 1 point if we assume the SD obtained to be reasonably accurate at around 10 raw score points. Accordingly, 100 cases were selected at random from the "normal" files, taking 30 cases from each age group, 16-25, 26-35, 36-45, and the remaining 10 cases from the age group 46-55. Cases with $L \geq 70$, " $?$ " ≥ 60 , and $F \geq 60$ were excluded as of doubtful validity; otherwise the first consecutive 30 cases in each age group were taken from files.

For the 100 normal males, the mean N-score was 29.14 with a SD of 9.55; for the 100 normal females the mean was 35.59 and the SD 9.80. This difference is in the same direction as that found in the criterion cases, although in this case the difference is based upon sufficient cases to reach statistical significance. The variances do not differ significantly as seen in Table 5, but the means are significantly different.

It should be pointed out that the foregoing difference is not only statistically stable but is of such magnitude as to be of practical importance and warranting separate norms; since a difference of six raw score points is 6/10 of a standard deviation for both groups. Since high T-scores on the MMPI indicate abnormality (in the sense that higher scores increase

TABLE 5
Test of Significance of Difference of Means and Variances for
100 Normal Males and 100 Normal Females

Group	N	Mean	V	F	Mean _{diff}	C.R.	P
Males	100	29.14	9.55	1.026*	6.45	4.712	P < .001
Females	100	35.59	9.80				

* The F of 1.026 is not significant at the 5 per cent level, so the variances are considered not to differ.

the odds that the person is or will be psychiatrically ill) this custom has been preserved in the present scale, although there are arguments on both sides as to its advisability. The relation of *raw scores* to *T-scores* is, therefore, an *inverse* one, contrary to that which obtains in all the other MMPI scales. If a person showing elevated scores on the personality components has a *low* raw score, he responds to the N items in the direction more characteristic of the criterion abnormal group; and this indicates a greater probability of his being clinically abnormal than if he had responded more frequently in the scored direction on N. Hence such a low raw score is given a high T score in the table. If a profile shows an elevation on any of the personality components, the higher the T-score is, the greater is our suspicion of an actual, behavioral deviation in respect to the syndrome indicated by the particular scale on which the elevation occurs. Whether high T-scores on N (*i.e.*, low raw scores) indicate a likelihood of abnormality in the absence of any significant elevation on the personality components proper remains to be seen; that is, an elevated N-component by itself is not *prima facie* suggestive of abnormality, as is obvious from the method of scale derivation.

This procedure has what may appear to be an unfortunate consequence in that with a mean of around 30 to 35 and a SD

of 10, the upper limit of possible T scores is fixed by the fact that no raw score is negative, so that no one can get a T score above 80. This fact is comparable to the *lower* limit of possible T scores that exists on the MMPI scales, and is related to the nature of the items which yields a marked positive skewness on these other scales. It is possible that with further refinement it will be found advisable to construct a set of T-scores based upon the distribution of normal persons with abnormal profiles, rather than upon the general unselected normal population. This is especially likely in view of the fact that the scale seems to have little significance by itself as regards the differentiation of normals from abnormal, in which case the only rational application of it would be to the group (both normals and abnormal) who score high on the personality components. If this were done one would not bother scoring N on any profiles except those showing a suspicious elevation on one of the eight personality components, and the T score would specify the position of such a person on the distribution of all "normals" with such deviant profiles.

In the case of both males and females, the unselected normals achieve a mean score about half way between the means of the normal and abnormal criterion group means for their respective sexes. The relationships are shown in Table 6.

In the case of the males, the criterion

TABLE 6
Mean T-Scores of Unselected Normals, Criterion Normals and Abnormals, Both Sexes

Group	Mean T-Score on N Scale
Male unselected normals	50
Female unselected normals	50
Male criterion normals	37
Male criterion abnormals	56
Female criterion normals	40
Female criterion abnormals	58

abnormals average about 1.9 sigma "above" their matched normals, in terms of the variability of the unselected male population; whereas the female criterion abnormals average about 1.8 sigma "above" their matched normals, in terms of the variability of the unselected female population.

Combining the sexes after converting to T-scores, the following table of cumulative frequency for the entire group of criterion normals shows their overlap with the unselected normal population (Table 7).

TABLE 7
Distribution of T-Scores for 42 Criterion Normals

T-Score	Frequency	Per Cent	Cumulative Frequency	Cumulative Per Cent
10 < T ≤ 20	1	2	1	2
20 < T ≤ 30	8	19	9	21
30 < T ≤ 40	20	48	29	69
40 < T ≤ 50	6	14	35	83
50 < T ≤ 60	7	17	42	100

No case among the criterion normals obtained a T of over 60 on the N-Scale. The median T score for the entire group is 33.5, sexes pooled, and 83 per cent of the cases do not exceed the mean of the unselected normals.

As regards the criterion abnormals, Table 8 gives the comparable percentages for them. No case among the criterion abnormals obtained a T as small as 30

on the N-scale, and only one obtained a T as small as 40, whereas 69 per cent of the normals fell at $T \leq 40$. The median T of the entire 42 criterion abnormals was 59.

TABLE 8
Distribution of T-Scores for 42 Criterion Abnormals

T-Score	Frequency	Per Cent	Cumulative Frequency	Cumulative Per Cent
T ≤ 20	0	0	42	100
20 < T ≤ 30	0	0	42	100
30 < T ≤ 40	1	2	42	100
40 < T ≤ 50	9	21	41	98
50 < T ≤ 60	14	33	32	76
60 < T ≤ 70	16	38	18	43
70 < T ≤ 80	2	5	2	5

DEVIATION OF THE "COLLEGE LEVEL" GROUP ON THE N SCALE

In all of the work with MMPI it has been found necessary to take special precautions regarding the college population and college-educated persons. The extent to which these persons deviate in their responses to many items from the general population is sufficiently great to necessitate elimination of items which on other grounds appear to be good differentiators for the variable being studied, but also unfortunately "differentiate" college people from people in general. In spite of the care exercised in this respect in construction of the personality scales of the MMPI, it is still observed that college means differ somewhat from those of persons in general. The typical college student, for example, shows a deviation below the mean on Hs and D, whereas he tends to show an elevation on Hy. It was suspected that the college (and apparently "college-educated") group might show considerable differences from people in general on the N scale, particularly when the application of N to some of those cases which

originally aroused interest in the possibility of such a scale failed to reveal as high scores on N as would be suggested by their being clinically normal in spite of having deviant profiles. Accordingly, males and females of the college population were sampled in an effort to determine the magnitude of this difference.

For a sampling of 50 college males the mean raw score on N was 21.40 which corresponds to a T score of 59, not quite a full standard deviation above the mean for the general male population. The SD of these 50 male college students was 8.938, as compared to 9.55 for the general unselected male normals. The difference between these means is significant ($t = 4.746$ $P < .01$) but the difference between the standard deviations is not.

For a similar sample of 100 college females (from which, however, profiles showing any $T \geq 70$ except Mf has been excluded for another study) the mean raw score on N was 25.93, which corresponds to a T score of about 60, again about one SD above the mean of the general female normal population. If the cases with "abnormal" profiles are included, the mean rises slightly to 27.16, corresponding to a T score of 59. The SD of the entire group of 122 female college students is 8.079 which in this case does differ significantly ($P < .05$) from that of the general female normal population.

The explanation of this rather pronounced deviation from the mean of the general population is not clear. The fact of this deviation, however, is of great importance for two reasons. First, it has a bearing upon the interpretation of the N scale, of such a nature as to dovetail with other evidence in casting serious doubt upon any interpretation

of the scale in terms of Rosanoff's "normal" component. Secondly, it indicates either that the scale cannot be applied to college students or else that if so applied it must be interpreted using a special set of norms.

The frequency of college graduates among those test and criterion cases which seemed to be test "misses" (i.e., were out of the hospital in spite of elevated MMPI profiles but also showed an elevated T on the N scale) suggested that *having been* a college student might alter the likelihood of a high score even though the person was not still in college at time of testing. This suspicion was strengthened by the fact that a group of college graduates studied with the MMPI in an industrial concern showed the usual college pattern even though the age range was considerable.

To test this hypothesis there were selected at random from a sample of profiles obtained in this industrial concern the profiles of 50 college graduates (engineering) of age 25 or over, disregarding the profile itself. The mean age of this group was 31.64 years, as compared with 30.95 for the 100 unselected male normals from the general population with whom they are to be compared. The age range was from 25 to 46, with a SD of 5.102 years. We, therefore, have a group of males who are all college graduates and have been out of college an average of 8 years, and whose mean age is very close to that of the general male normal sample used in getting the tables for T.

For this sample of 50 male college graduates the mean N score (raw) was 20.9, which corresponds to a T score of 59, which is exactly that of the sample of college males still in school at time of testing. The SD of raw scores was 6.688, which differs significantly ($P < .01$) from

the variability of the 100 unselected male normals from the general population, on which the T scores were based. The mean raw score of 20.9 differs significantly from the mean 29.14 of the unselected general population males, showing a t of 3.10 ($P < .01$). Unfortunately, a corresponding study for females out of college for some years at the time of testing could not be made due to lack of convenient case material. Considering the similar elevation of female college students, it is provisionally taken as probable that some degree of elevation at least would be maintained in later life, although this will be investigated when a suitable group of profiles is readily available for study.

It appears from this evidence that the elevated T scores of college persons are not a function wholly of their age, nor of being "in college" at the time of testing. The question then arises, is it legitimate and desirable to employ separate norms in the application of N to the college group (henceforth the "college group" will be used to designate college students and college graduates alike) or should the scale not be applied to them at all?

If we had some way of knowing *why* the college group deviated from the general population, we would know how to answer this question. But since the reason for the deviation is not known, it is difficult to decide. The proposal has been made by certain college counselors that a T score of 60 on one of the personality components "means as much for a college sophomore as a T of 70 would mean for a person from the non-college general population." The rationale of this attitude is questioned by those who are most intimately connected with the application of the MMPI in clinical diag-

nosis. That a T of 60 may mean as much in terms of its statistical rarity among college sophomores as one of 70 in the general population is admitted and is quite trivial. The important thing to determine is whether in the long run one can expect a college student scoring 60 on, say, Hs to have as many bodily complaints and be as refractory to modification as a non-college person with a score of 70. It must not be forgotten that scales such as those of the MMPI acquire whatever non-statistical meaning they possess from the clinical description of those extreme deviates who make up the diagnostic categories for which the various scales are named. "Hypochondriasis" refers initially to a certain behavior syndrome found in a small number of neurotic persons; the burden of proof is upon anyone who assigns any other than a somewhat "watered" meaning of this term in describing milder deviates scoring in the "normal" although elevated range. If it is actually the case that college students and persons who have graduated from college exhibit significantly less body-consciousness and somatic complaining than the general population, there would seem to be no good reason for applying a special set of norms to college people based upon their own group.

In the case of the present scale this reasoning is not so straight-forward. The function of the N scale is not to measure a "personality" component which is of intrinsic interest to us (although it may ultimately turn out to do so), but rather to distinguish persons showing deviant profiles but psychiatrically adjusted from persons showing equally deviant profiles who are psychiatrically maladjusted. Since we do not know how the scale works (i.e., the psychodynamics of re-

sponding to items which bring about such discrimination as empirically occurs) it is not clear whether separate norms would facilitate this in the case of college students or not. Actually it would be necessary to make a separate "test" study of college level abnormals and normals to see whether college normals with deviant profiles score relatively higher raw scores on N than abnormals, in which case the use of separate college norms would be profitable in practice. This has not been done, mainly because of the great difficulty of selecting cases from the abnormal files on the basis of level of educational attainment.

Nevertheless, it is at least suggestive to consider the performance of the college "normals" who deviate on the other personality components of the MMPI in terms of their N scores as related to college group means and variabilities. For 22 college girls showing any T score of 70 or greater (excluding *Mf*) but "normal" so far as known, the median raw score on N is 32, which is at about the 86th percentile for a sample of 100 female college students, and the mean raw score is 32.73 which is approximately one SD above the female college mean. Such a score, on the other hand, would be about one-third of a SD below the mean raw score of women in general ($T = 53$).

Among a group of over 100 employed male college graduates there were found 9 persons who showed a T score of 70 or greater on any of the eight MMPI personality scales excluding *Mf*. These persons, while probably not so well adjusted as the non-deviant cases, were at least "normal" enough to make an adequate occupational adjustment at a fairly high level and stay out of the hands of a psychiatrist. Their mean raw score on

N was 28.67 which corresponds to a T of 50 for the general population. However, in terms of the distribution of N scores for the entire group of employed college graduate men in which they occurred, this mean is over 1.1 SD above the latter's mean, and would be equivalent to a T score of about 39 if separate norms were used. This latter T score is not far from the T of 37 found in the original criterion group of male normals showing abnormal profiles.

Further evidence regarding the desirability of separate norms for college persons will have to be collected. Until this evidence is forthcoming, however, the same norms will be used in order not to favor the scale unduly; however, it will at least be pointed out in what follows that in several cases "misses" are not clearly "misses" because the cases in question would not be misses had the T scores on N been used based upon special college norms.

APPLICATION OF N-SCALE TO A "TEST" GROUP OF NORMALS WITH ELEVATED NEUROTIC TRIAD SCORES

As a further study of the differentiating power of N it was applied to a new "test" group of deviant-profiled normals from two sources—a file from a group of WPA workers and a set of scores contributed by a medical detachment at Fort Snelling Hospital. While the WPA group might be thought atypical to such an extent as to be unusable for such purposes, the findings on the other MMPI scales and the item frequencies for this group do not indicate that they differ in any appreciable degree from the general population in those aspects of personality sampled by the inventory.

Since the derivation of N was based upon criterion cases showing elevations

on the neurotic triad Hs, D, and Hy, the ideal test cases should be those showing these elevations. Only a small number of such cases was found in the WPA and army groups combined, namely, a sample of 22 male individuals who were making an adequate adjustment psychiatrically as far as was known, but who deviated to the extent of $T \geq 70$ on Hs, D, or Hy. Six cases were included which probably should have been left out on the basis of information subsequently collected; three of these were found to have been "under a doctor's care" (com-

As may be seen from an inspection of the distribution, there is a gap from 43 to 50 above which lie seven cases, of which five belong to the six cases mentioned above as possibly deserving of exclusion. The one case with $T = 64$ is a college graduate and was under a doctor's care when tested. The case at $T = 56$ and that at $T = 50$ are both college graduates. The cases at $T = 52$ and $T = 55$ were both "under a doctor's care" at time of testing. Two of these three college graduates are indicated on the attached record as "nervous," and

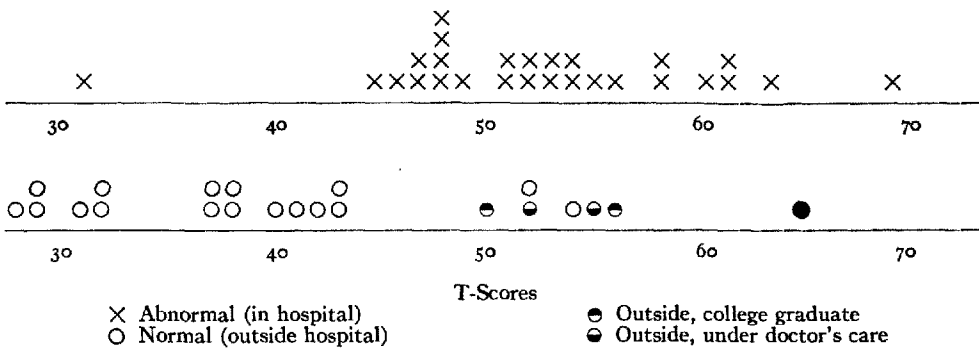


FIG. 2. Distribution of T-scores for 22 male "test" normals with elevated neurotic triad but out of the hospital; compared with scores for 27 test abnormals in the hospital with diagnosed psychoneurosis and elevated triad.

plaint unknown) at the time of testing (a fact which would exclude them from the "normal" category according to the standards on which the MMPI was originally constructed) and the other three cases were college graduates.

The distribution of these cases is shown in Fig. 2. The distribution placed above it will be discussed later. The median T score of these 22 "test normals" with elevated neurotic triad but out of the hospital was 40.5, and the mean T score was 42, with a SD of 9.585. So this test group falls at .8 to 1 SD below the mean of the general population, as compared with the criterion normal males at about 1.3 SD below.

the one with $T = 56$ was unmarried at age 45. With these considerations in mind it may be said that the "test" group results compare rather favorably with the results of the original criterion cases.

Had the six doubtful cases been excluded to be more lenient in judging the N scale instead of rigorous, the median T score would have been 37, at the mean of the original male criterion normals. The mean T score in that case would have been 37.8, again very close to that of the criterion normals.

Another way of looking at these same data would be to convert the scores of the three college cases into appropriate T scores assuming that separate college

norms should be used. In this case the college cases with T's of 64, 56, and 50 would instead score T's of 56, 48, and 41, respectively. Using these transformed scores instead of the original ones, the median T score of the entire 22 test normals would be at 40. Only 5 cases would now have a T above the mean T of 50, and no case would have a T score above 55. (See Table 9.)

TABLE 9

Frequency Distribution for 22 Male "Test" Normals from WPA and Army Medical Detachment Group, Showing Elevated Neurotic Triads. The College Cases Have Been Expressed in Terms of College Norms

T Score	Frequency	Cumulative Frequency	Per Cent	Cumulative Per Cent
20	3	3	14	14
30	7	10	32	46
40	7	17	32	77
50	5	22	23	100
60	0	22	0	100

It seems justifiable from the preceding analysis to state that at least so far as the neurotic triad is concerned, a "normal" person who shows an elevated profile but manages to stay out of psychiatric hands will average a T score in the neighborhood of 40 or about one SD below (one SD above in raw score terms) the mean of the normal population. As the T score of a deviant-scoring profile rises above 50 the chances of its owner being free of psychiatric involvement to the extent of hospitalization become progressively smaller; whereas a T score as high as 60 will make these chances very slim. Putting it in another way, in terms of whatever component the raw scores on N measure, a person on the average needs an amount of this component about 1 sigma above the mean of the general population if he is to remain

"normal" in spite of elevated scores ($T \geq 70$) on the neurotic triad. As one reaches and then begins to fall below the general population raw score mean on this factor, the odds increase that one is not a member of the "normal" group. If one scores above 70 on Hs, D, or Hy and falls as much as one SD below the mean of the general population, the chances that he is psychiatrically involved are high.

It should be pointed out again at this time that the great crudity of division into "normal" and "abnormal" on the basis of the only available objective criterion "intramural" or "extramural" with reference to a psychopathic ward inevitably worsens the apparent differentiating power of a scale, particularly one such as the present. Although practically all persons who are "in" a mental hospital with psychosis or psychoneurosis deserve to be there, the converse is not true for those who are "outside." Detailed study of the individuals used in testing such a scale would almost invariably work in its favor by finding independent evidence of psychiatric difficulties in persons otherwise counted as "exceptions" to a rule. As an example of this I might mention one of the few cases in which a person who would be counted as such a test "miss" was accessible to me. I was shown a profile of a college student with an elevation in the region of $T = 90$ on the Hy scale, with Hs, D, and Mf also up but not above 70. I scored his test on the N scale and found to my discomfiture (this being before the extent of college atypicality on N was realized) that his N score was about 65, contrary to a provisional hypothesis that 60 was a rough limit of T scores on N to stay out of the hospital if one had a deviant profile on one of the other scales. I

mentally recorded this case, which I had come across more or less accidentally in a social way, as a "miss" for the N-scale. Shortly thereafter I had occasion to meet the individual in question personally, without his knowing that I had seen his MMPI profile. In the course of conversation I inquired as to his draft status. He informed me that he had been in the army but had been discharged for "functional asthma." I asked whether he had any other ailments, and he went on to describe the terrible headaches he had so frequently for which the doctors could find no explanation, the backaches which were equally troublesome, and the fact that he had had insomnia and was quite a "nervous person." Had this opportunity for direct clinical contact not presented itself, this case would have appeared merely as a misplaced X on a frequency distribution and called a "miss" for the scale.

In the evaluation of the results applying N to the "test" group of normal deviates on the neurotic triad, one needs to consider the fact that in the original criterion group of *abnormals*, the profiles were in a sense "artificially" held down by the nature of the matching technique involved. That is, due to the rarity of extremely elevated T scores among normals, the matched profiles of the criterion *abnormals* were simply not allowed in the matching process to attain the great elevations ($T = 90$ and up) which do occur in the hospital neurotic population. In a certain sense this is of no consequence, inasmuch as if it is established that T scores above this upper limit almost literally *never* occur in a valid normal profile the problem of distinguishing "normals" with such elevations does not exist. That is, that a scale such as N

should "work" only within the semi-pathological range, say, of $T = 70$ to $T = 90$ (which includes two full standard deviations) would be no cause for complaint if everyone scoring above 90 were practically certain to be pathological. It might be objected that the normal curve integral does not decrease very rapidly between a score at two sigma and a score at four sigma. This statement has a superficial plausibility, but it is based upon the erroneous assumption that the "probabilities" in question are those of a normal curve of distribution for the "normal" population. The point must be clear that the "probability of a man's being correctly called abnormal from his profile alone" is not a function of the normal distribution alone but depends to an equal degree upon the distribution of scores for abnormal persons and depends most of all upon the actual relative frequency of normals and *abnormals* as defined by the non-test criterion. The solution of this problem, which is really one of inverse probability, requires the application of Bayes' Rule and as usual the relevant knowledge for this application is inaccessible to us in practice. In point of fact, if one calls a case "abnormal" (by some arbitrary non-test dividing criterion) whenever one finds a T above the 5 per cent level, he will not be right 95 per cent of the time by any means. The relative frequency of clinically normal and abnormal persons is such that the best guess is probably "normal" with any T score between 70 and 80. There may only be 5 persons in 100 who score above 70 on Sc among the normal population; but there are much fewer than 5 schizophrenics in 100 among the general population.

To illustrate this by a specific example, consider a population in which the rela-

tive frequency of actual schizophrenics (defined, say, by the presentation upon careful psychiatric examination of symptoms so marked that the examiner would be willing to attach this as a diagnosis) is as high as 5 per cent, which is, of course, much higher than really exists. Assume that it is as rare for an actual schizophrenic to hold his T score on Sc down to 70 as it is for a "normal" person to achieve a score as high as 70 or above, say, again 5 per cent. Then if one were to draw a sample of persons at random from such a population and label every case a schizophrenic who showed a score ≥ 70 , he would not be right by any means in 95 cases out of 100 but in about 67 out of 100. The "probability that a man scoring as high as 70 is a schizophrenic" is only .67, and 1 in 3 times such a diagnosis will be erroneously applied to a man who is in fact "normal."

For this reason it is quite reasonable to concern oneself with the group of normals scoring between 70 and, say, 90, as of sufficient frequency to justify study, even though the mathematical properties of the normal curve (which do not apply to MMPI scales for the most part anyway) are such as to indicate to a superficial examiner the practical uselessness of studying the group in that range. Out of 1000 "normal" cases in the files one may expect to find, say, 50 profiles showing scores above 70 on any given scale, and yet not find a single case (except invalid profiles with high F) scoring above 90. As was pointed out earlier in the introductory remarks, it is this group of 50 persons at whom the present study is aimed.

Nevertheless, the mean profile of the criterion abnormal is almost completely within the range of two standard devia-

tions usually called "normal," and, therefore, it is worth while to study the N scores of a group of "test" abnormal of appropriate neurotic triad diagnosis and test scores whose profiles are allowed to vary freely into the higher ranges. The distribution of these abnormal must be compared with that for the 22 test normals we have just finished considering in terms of their relation to the general population norms alone.

From the abnormal file were selected at random 27 male abnormal with an elevated neurotic triad (Hs, D, or Hy ≥ 70) and a clinical diagnosis in harmony with such a configuration in that the diagnosis was required to be psychoneurosis, either hypochondriasis, hysteria, reactive depression, or mixed type. The group thus selected turned out to consist of 11 hypochondriacs, 8 hysterics, 6 psychoneurosis mixed, and 2 reactive depressions. Cases with ? or F scores over 70 were not allowed. This group of "test" normals is also stringent in its requirements for the present scale, because whereas deviations in the upward direction were unrestricted in this group, all cases used as test abnormal showed elevations on the neurotic triad, whereas a number of hospitalized abnormal with such a diagnosis fail to do so.

The mean raw score on N for this group of abnormal was 26.44, which corresponds to a T score of 53. The SD of the raw scores was 7.057. The mean of 53 is about $1/3$ sigma below the mean T score of 56 found in the male criterion abnormal, whose profiles were "held down" artificially by the matching technique; but the difference is not significant ($t = 1.507$, $p > .10$).

The distribution for these 27 random and unrestrained profiles was presented in Fig. 2 for comparison with the distri-

bution of N scores for the 22 male "test" normals. It can be seen that the overlap is considerable, 6 of the test normals reaching or exceeding the median of the abnormals. The separation of this test series should be compared with the separation shown by the distributions for the criterion normals. It should be pointed out again that the highest scoring seven test normals include three college graduates one of whom was under a doctor's care and two other persons under a doctor's care for unknown ailments. With this in mind the overlap does not appear quite so great.

In spite of these handicaps, the difference between the means of these two distributions is 10.37 raw score points, which is statistically significant ($t = 4.268$, $P < .01$). It can be said with confidence that even if the matching process did hold down the T scores of the criterion abnormals to a slight degree still it did not hold them down to such an extent that the differentiation is removed when the profiles are not so restrained.

A comparable test study with females could not be carried out because of the very small number of female normals with elevated neurotic triads available other than those already used in the criterion study. There were found only seven female cases in the WPA sample with elevated triads, of whom one was a Negro college graduate (social worker) and another was under a doctor's care, considered "nervous," and separated from her husband at the time of testing. This left only five cases for whom both clinical normality and applicability of norms could be reasonably assumed. The mean T score of all seven cases is at 50—the mean of the general population. If the college case is changed to college group T-score, the mean would be 48.

Whether female test cases would show a differentiation better or worse than the males cannot be predicted upon present evidence.

APPLICATION TO CASES WITH ELEVATIONS OTHER THAN ON THE NEUROTIC TRIAD

Due to the small number of deviant profiles available for study in the normal file, it was not possible to apply the N-scale to several new "test" groups of typical normals with deviant profiles on the neurotic triad, as would have been the best procedure. However, it will be remembered that in the original selection of deviant scoring normals, study was confined to those showing deviations on the neurotic triad, Hs, D, and Hy. The remaining set of deviant scoring normals constitute a possible "test" group, although they have the disadvantage of testing two hypotheses at once. For if the N-scale fails to discriminate them in the desired direction, it would be unclear whether this was because too many of the original items were merely selected due to sampling errors, or whether the same "suppression variables" are simply not applicable to the other scales as to the neurotic triad.

In order to get a sufficient number of cases it was necessary to relax somewhat the previously stringent restrictions as to validity. In this group, F scores were allowed to rise above 60 before discarding a profile as invalid, although no case retained had an F over 66. There was also an examination of the profile as to its "form," which is, for persons acquainted with the clinical application of the MMPI, a good indicator of whether the F is "validly" high or an indicator of confusion or non-cooperation. (Of course, these procedures were ap-

plied before any determination of the N score was made on the cases.) The result of this screening was the selection of 21 "normal" males showing at least one T score equal to or greater than 70, but not on any of the three neurotic triad scales, and 21 "normal" females with similarly deviant profiles. (The total number of "test" cases is 42 by pure coincidence, and it must be clearly understood that none of these 42 test cases is among the 42 of the criterion group used

per cent, which compares very favorably with a corresponding 83 per cent for the criterion group. Here also, no case was found with a T greater than 60 on the N-scale. The summary of these cases is presented in Table 10.

In spite of the fact that the central tendencies of the T scores on N rise when the scale is applied to the new group, the relative rarity of scores above the general population mean of 50 is maintained. This would suggest a fair

TABLE 10
Distribution of T-Scores on N-Scale for 21 Male and 21 Female "Test"
Normals Showing Elevated Profiles Other Than on Hs, D, or Hy

N	Frequency			Cumulative Per Cent		
	Males	Females	Both	Males	Females	Both
$T \leq 20$	0	0	0	0	0	0
$20 < T \leq 30$	3	0	3	14	0	7
$30 < T \leq 40$	6	8	14	43	38	40
$40 < T \leq 50$	11	9	20	95	81	88
$50 < T \leq 60$	1	4	5	100	100	100
$60 < T \leq 70$	0	0	0	100	100	100

in item selection.) Due to the relaxing of the F restriction, three cases with elevated neurotic triad scores were included in this group who had previously been excluded from the criterion normal sample. The other 39 test cases deviate on non-triad components only.

The mean raw score on N of the 21 male test cases is 38.09, SD 7.993; for the 21 females the mean raw score is 41.10 and the SD is 6.89. These means correspond to T scores of 41 and 44, respectively. So we see that in both sexes, the mean T score of the test cases rises about 4 points from that of the criterion normals on whom the item selection was based. The median T score is 41 for males, and 44 for females; 42.5 for the sexes pooled. The shift seems to be from the region below $T = 40$ into the region between 40 and 50. The proportion of cases which do not exceed $T = 50$ is 88

applicability of the scale to cases other than the triad deviates, were it not for the fact that the raw score on the N scale correlates to a rather marked degree with Pt and Sc and to a lesser but still positive extent with some other scales in the normal as well as abnormal population. For this reason the mere presence of deviation on these other components tends to be associated with a deviation on N (appearing as a low T score) regardless of whether the person measured is in the hospital or out. Therefore, the tendency of the present group to score low on N does not establish its utility as applied to cases other than those elevated on the neurotic triad, until we investigate the magnitude of a similar lowering effect on the cases who are in the hospital. When this is done results are not nearly so encouraging, and in fact indicate that while stable statistical differences exist,

the overlap is too great to make the scale of any appreciable practical value in such cases.

A group of profiles for 25 males and 25 females was extracted from the file of abnormals, the requirement being that the profile must deviate on some other variable besides the neurotic triad, that the peak scores must be on some other variable, and that the clinical diagnosis must not be hysteria, hypochondriasis, or depression. Profiles with ? or F scores above 70 were also excluded as probably invalid. These cases were of various diagnostic groups, including psychoneuroses other than those mentioned above, manic-depression, schizophrenia, psychopathic personality, paranoid condition, behavior problem, and simple adult maladjustment. The sex difference which appears in the goodness of separation is such as to warrant separate treatment.

For the 25 male abnormals the mean raw score was 29.92, corresponding to a T score of 46, and a median raw score of 31 (T = 48). For the 25 female abnormals the mean raw score was 29.92 corresponding to a T score of 56, and a median raw score of 31 (T = 55). These comparisons are presented in Table 11.

In spite of the significance of these differences from a statistical point of view, the overlap is excessively great and the

differentiation, while indicating something of theoretical interest regarding the generality of at least part of the N scale as a suppressor, is not sufficient to be of practical utility. Inspection of the plotted distributions (not presented here) indicates that the difference of the sexes lies in the much greater tendency of the male abnormals to achieve low T (high raw) scores than is the case with the female abnormals. As regards the normals, in both sexes the "test" normals tend to remain below the population average; and this tendency appears also when they are compared with the central tendency of the 50 abnormals. In the latter comparison, we find only 14 per cent of the female normals reach or exceed the median T score of the abnormals; and 19 per cent of the male normals reach or exceed the median T score of the abnormals. But at the other end of the distribution whereas only 8 per cent of the female abnormals fall as low as the median of the female normals, we find that fully 1/3 of the male abnormals fall as low as the median of the male normals. In other words, if one has a T score above 50 on the N scale with any nontriad component deviating to a T of 70 or more, he is likely to be abnormal. If one has a T above 60 on N in such a case, he is practically certain to be abnormal. But

TABLE 11
Comparisons of the 42 "Test" Normals with Non-Triad Elevations and 50 Random Abnormals with Non-Triad Elevations

Group	Median Raw Score	Mean Raw Score	Mean T Score	Mean Raw Score Diff	t _{diff}	P
			<i>Males</i>			
21 "test" normals	38	38.10	41	5.50	2.135	.05 > P > .02
25 abnormals	31	32.60	46			
			<i>Females</i>			
21 "test" normals	41	41.10	44	11.18	5.110	P < .01
25 abnormals	31	29.92	56			

a low T score, at least for males, by no means guarantees that one will not be abnormal, whereas it does seem to afford some such assurance in the case of females.

The net result of the investigation of this test group can be summed up as follows: The mean T-scores on N of "normals" showing elevations on other scales than the neurotic triad is somewhat higher than that of the criterion group of normals. The appearance of T scores above 50 is equally rare in the test group, as is also the case for scores above 60. The incidence of scores in the 40 to 50 range increases over that of the criterion normals. The differentiation of normals from abnormals is not nearly so good as this finding might suggest, because abnormals with deviations and diagnoses other than on the neurotic triad obtain considerably lower T scores on N than the criterion abnormals and are in the case of the male abnormals actually below the mean T of the general population. There is a statistically significant difference between the mean T scores of normals and abnormals with non-triad elevations, but the overlap is great, particularly in the case of the males. It does not appear that the amount of discrimination is sufficient to have any hope that the N scale will be of practical utility for these groups in its present form. The statistical stability of the differentiation, however, suggests that further refinement, especially in the removal of those items which contribute most to the high correlation with some of the other scales, might improve this state of affairs. It is, of course, very difficult for

those items which must be operative in producing such discrimination as does exist to show up as clearly powerful differentiators when their influence is systematically masked by items showing a high correlation with the scales for abnormal components which are being considered. That even with the existence of such high correlations with Pt and Sc as will later be indicated, the scale is able to produce a statistically although not practically significant differentiation suggests that a further search for a highly generalized suppression variable would be rewarded.

It might be thought that the exclusion of cases showing T above 70 on ? or F from the "test" normals is undesirable since such profiles do occur in practice. The present position is that for this purpose (evaluating differentiation) every effort should be made to include only valid profiles, and that the clinical significance of apparently invalid ? and F scores may be subsequently determined once the power of the scale is clearly seen. If the normal records called invalid because of their unusual form combined with suspiciously high ? or F scales were included, the differentiation would not become worse but on the contrary would have been greatly improved. A group of 26 cases originally excluded from the "test" normals because of invalid looking profiles and high F or ? scores earns a mean T score of 37.31 (38.64 for males, 35.75 for females) which is considerably lower than the value for these means actually obtained by the more rigorous selection. Among these 26 cases no profile shows a T above 50 on the N scale.

CHAPTER V
MISCELLANEOUS PROPERTIES OF N AND ITS RELATION
TO CERTAIN OTHER VARIABLES

N CONSIDERED AS AN INVERTED "LIE"
SCALE IN TERMS OF THE SCORES OF
A GROUP OF CLINICAL ABNORMALS
SHOWING NORMAL MMPI
PROFILES

IN VIEW of the slight though insignificant difference between the N scores of abnormal with freely varying profiles and those with profiles "held down" artificially by matching, together with the negative correlations later to be reported between the two chief "lie" scales and raw score on N, it seemed possible that giving few responses scored on N might be allied to the same tendency which gives rise to elevated lie scores. Furthermore, there are actually 7 of the 22 items on L_6 (an unpublished lie scale in process of development) which appear on the N scale also, scored in the opposite direction in every case. One partial test of such a hypothesis is to determine whether the T scores on N of clinical abnormal showing "normal" profiles tend to be elevated.

The file of profiles for hospitalized female abnormal was entered at random and the first 35 consecutive cases failing to show any MMPI personality component over 65 were extracted. (Actually, the first 19 cases so selected showed no T score over 60 on any MMPI scale; but it was necessary to raise the boundary to 65 in order to get more cases.) All cases with T over 58 were eliminated from consideration, since elevated T scores frequently mean spuriously low profiles achieved by simply failing to answer large numbers of significant items. The clinical diagnoses of these patients were variable, in-

cluding psychoneuroses (all types), depression, psychopathy, schizophrenia, mania. The mean raw score on N for these 35 patients with "normal" appearing profiles was 26.63, and the SD 9.565. The T score corresponding to this mean would be about 59. Another group of 16 female abnormal (this time showing no scores above 60) taken from files for recent patients shows a mean raw score of 25.5, corresponding to a T of 60. For the entire group of 51 female abnormal with normal profiles, the mean raw score is 26.27, corresponding to a T score of 60, and the SD is 9.594. This mean raw score differs significantly from that of the 100 unselected normal females ($t = 5.515$, $P < .01$).

A similarly chosen group of 26 male abnormal showing no score over 60 showed results in the same direction but of lesser magnitude. The mean raw score for male abnormal with normal profiles was 24.92, corresponding to a T for males of only 54, and a SD of 7.4519. This raw score mean is of borderline significance statistically, the t in comparing it with the mean of 29.14 for unselected normal males being 2.079 with 124 d.f., so that $.02 < P < .05$. Why there should be a smaller difference in the case of males is not clear, nor is it clear why the combining of over 300 male abnormal profiles and 500 female should have resulted in a rather pronounced difference in the pattern of frequencies for clinical diagnoses among those managing to hold their profiles down to "normal." Over half of the females in this study were diagnosed psychoneurosis, 3 were schizo-

phrenias, 3 psychopaths, 2 each of simple adult maladjustment, involuntional depression, manic depressive depression, and 1 each of paranoid state, addiction, behavior problem, and manic. In the case of a sample of males taken in the same way (consecutive cases in the file showing "normal" profiles) there were only 5 psychoneuroses, 5 schizophrenias, 7 psychopaths, 2 manic depressed, 2 paranoid state, 2 homosexual, 2 alcoholic, and 1 behavior problem. Quite possibly the difference in the relative frequency of cases which would be *expected* to show elevations on the neurotic triad is the major cause of the difference found between the sexes in the functioning of N as a sort of "lie" scale.

The median T score of the 51 female abnormal with normal profiles was 60, and only 14 per cent showed T scores below 50. In the case of the males the median was 53.5 and 27 per cent showed a T score below 50.

In summary, it may be concluded that whatever the mechanism may be, the N scale functions to a significant extent inversely as a variety of "lie" scale not in the sense of indicating deliberate and conscious deception but some tendency, however produced, to avoid putting oneself in an unfavorable or abnormal light when answering the items on MMPI. Female abnormal who fail to show any of their personality components on the eight scales deviating by more than 1.5 sigma ($T = 65$) tend on the average to reach a T of 60 on the N scale. Male abnormal with similarly "normal" profiles otherwise show the same tendency to elevated T scores on N but to a considerably weaker degree, averaging a score only about $3/10$ of a sigma above the mean of men in general ($T = 54$). Only about 1 female abnormal in 6 or 7 can

hold her profile down to 65 on all scales and keep her N score below $T = 50$. About 1 male abnormal in 4 is able to achieve this, however.

The occurrence of T scores in the region 55 to 65 is, of course, so frequent among the general normal population that the present scale could not have any practical utility in the detection of lying when inspecting profiles for abnormality. It is only after being possessed of the information that the patient is abnormal in spite of his apparently normal profile that it occurs to us to interpret an N of, say, 55 in such a way as to "explain" the low profile. The chief importance of the present finding regarding the function of N in a rough way comparable to L and L_6 is with reference to its ultimate theoretic interpretation.

RESULTS OF APPLICATION OF N TO UNSELECTED ABNORMALS

Upon the Rosanoff hypothesis which originally suggested the present investigation, one might plausibly argue that if the N scale were measuring even in a crude way this so-called "normal" controlling component of personality, the mean score of a group of persons defined simply by their psychiatric abnormality should be elevated (i.e., raw scores on N should be low). The elevation of T scores on N might not be expected to be as great for unselected abnormal persons as for those in the criterion abnormal group even on Rosanoff's hypothesis. For since, as was pointed out previously, the criterion abnormal had their MMPI profiles "artificially" held down by the matching process, we might in effect have been selecting these abnormal who were clinically abnormal in spite of fairly small amounts of deviation in the specific abnormal components

which N is supposed to control. Such a group would be expected on Rosanoff's hypothesis to be characterized by an extremely small amount of the "normal" component, since such relatively slight deviations on the MMPI scales (assuming these scores for the moment to be valid) have resulted in their being psychiatrically incapacitated.

male abnormal is actually higher than the corresponding raw score mean for the male normals, but this difference is not significant statistically. The variances do not differ significantly by the F-test.

Because of the sex difference these results are difficult to interpret in any consistent way. It would seem that the fail-

TABLE 12
Comparisons of Means and Variances of 50 Random Abnormals of Both Sexes with the General Population Sample

Group	N	Mean Raw	T	SD	Diff. of Means	C.R.	P
Male normals	100	29.14	50	9.550			
Male abnormal	50	30.14	49	9.457	1.00	.608	P = .23
Female normals	100	35.59	50	9.804			
Female abnormal	50	30.62	55	10.284	4.97	2.834	P < .01

Following this line of reasoning we selected 50 abnormal from both sexes from the file of hospitalized patient's profiles. Cases with organic involvement or with diagnosis unrecorded or doubtful were excluded. No restrictions were placed on P, L, and F in this sample, because of another use to which it was intended to put the data. The diagnoses were, of course, very heterogeneous, as was intended, the largest single category being psychoneurosis.

The means and the variabilities of these abnormal are given in Table 12, together with a test of the significance of their deviations from the unselected normal sample upon whom the T scores were based.

We find that the mean raw score of the female abnormal is about 1/2 an SD below that of the female normals on which the T scores are based, and that this difference is statistically significant at the 1 per cent level. The mean of the

ure to obtain any difference in the case of the males constitutes stronger evidence against any interpretation in terms of Rosanoff's theory of control than the slight although stable difference in the case of the females constitutes evidence in favor of such an interpretation. In view of the other findings on the average amount of deviation found in abnormal in MMPI research, a deviation of .5 sigma does not lend much support to the suggestion that these female abnormal are clinically abnormal because they do not have enough of the hypothetical "control" or normality variable to stabilize themselves. It must be remembered that this T score of 55 is still almost 1/2 SD nearer to the mean of unselected normals than the mean of female college students. The smallness of the deviation and the lack of any difference in the case of the male group lead us to doubt seriously (especially in conjunction with the evidence from intercorrelations and item

content to be discussed later) whether the results can be construed as supporting the original hypothesis. That there is any deviation at all in the case of the one sex can be "explained" on simpler grounds than the Rosanoff hypothesis, as will be seen shortly.

THE RELATIONSHIP OF N TO AGE AND TO
THE OTHER VARIABLES OF
THE MMPI

Further evidence against any interpretation of N as a positive "normality" component of personality comes from a study of its relation to the other MMPI scales and to chronological age. For the 100 random male normals on whom the

group (16-25) were medical students and such a college group, as pointed out previously, tend to show low raw scores on N.

The intercorrelation of the N scale with the other scales of MMPI and with a few unpublished or abandoned scales was also investigated. In Table 13 below are presented the correlation coefficients for raw scores on N with certain other measures on MMPI, for the sexes separately and for normals and abnormals separately. The variable C_h is the no longer employed "correction scale" for the old hypochondriasis key (10), and L_6 is a more subtle "lie" scale which is still in the process of validation.

TABLE 13
Correlation of Raw Score on N with Other Variables

	C_h	L_6	L	F	H_s	D	Hy	Pd	Pa	Pt	Sc	Ma
100 male normals	.79	-.76	-.17	.43	.58	.02	-.25	-.33	.27	.84	.71	.37
100 female normals	.78	-.65	-.11	.28	.54	.46	.11	-.33	.21	.82	.72	.31
50 male abnormals	.82	-.82	-.28	.60	.55	.47	.21	.46	.58	.80	.74	.36
50 female abnormals	.78	-.84	-.21	.42	.37	.56	.17	.37	.53	.80	.69	.31

T scores were standardized, there is evidence of a progressive rise in mean N score with chronological age in the general population. For the age range 16-25 the mean N raw score was 23.73, as compared with a mean of 34.30 for the age group 36-45 in the same normal sample. For all 100 male normals there is a correlation of .38 between chronological age and raw score, which is significantly different from zero ($P < .01$). For the 100 female normals this correlation is only .16, which is not statistically significant in its deviation from zero ($P > .10$). The means for females also do not show as much change, being 33.67 for the youngest age group versus 38.80 for the oldest.

The difference in the case of the males is in part due to the fact that about half of the normal males in the youngest age

For the 100 normals, all correlations above .20 are significantly greater than zero ($P < .05$); for the 50 abnormals, those above .27 are so significant (7). The correlations with F and L probably underestimate the intensity of association since the distribution of these two variables is so markedly skew that the conditions for the Pearson r being a maximally descriptive statistic of association are assuredly not fulfilled. The importance of an exact measure of covariation was not such as to warrant the application of more appropriate descriptive statistics, however.

Of particular interest in the table above are the findings which bear upon the psychological interpretation of the N scale. Although detailed discussion of this matter will be given in the following

chapter, some emphasis may be placed on certain correlations here. Of particular interest are the following relationships:

1. The high positive correlation of N with the old correction scale C_n .
2. The marked to high negative correlations of N with L_n , and the fact that these correlations are higher in the case of the abnormals. Also allied to this the negative correlations, although low, with L.
3. The high positive correlation of N with Pt and Sc.
4. The relatively low correlation (significantly negative for normal males) with Hy, despite the positive association with Hs.
5. The positive correlation with F.
6. The general tendency to significant and in some cases marked positive association with the various abnormal components, there being only one negative correlation in the table except for the lie scales.

The fact that N correlates low and in one group significantly negative with Hy and yet correlates fairly markedly with Hs suggested another correlational study. The Hy and Hs scales have many "somatic" complaint items in common. The Hy scale differs not only in that it does not share quite *all* of the Hs somatic items with the latter (which in itself would not account for this correlational difference), but more importantly the Hy scale contains a large number of non-somatic items, including those "denial of the psychiatric" responses that were mentioned in Chapter II. This group of items, as was pointed out there, have the common property of denying psychiatric involvement and assert oneself to be extraordinarily well adjusted socially. These items are responded to by the hysteric in the statistically "normal" direction, and hence are recorded "O" in scoring the MMPI. Examples of these

items and the responses scored for Hy are:

- C-25 I have often lost out on things because I couldn't make up my mind soon enough. (F)
- D-52 I think most people would lie to get ahead. (F)
- E-43 When in a group of people I have trouble thinking of the right things to talk about. (F)
- E-54 I can be friendly with people who do things which I consider wrong. (T)
- F-25 I resent having anyone take me in so cleverly that I have to admit that it was one on me. (F)
- G-40 In walking I am very careful to step over sidewalk cracks. (F)
- H-10 I commonly wonder what hidden reason another person may have for doing something nice for me. (F)

Because of the relations in the correlation table, plus the growing hypothesis (after abandoning Rosanoff's) that the N scale was at least in part detecting a tendency to say psychiatrically undesirable things about oneself even if unwarranted, it was felt that there might be a correlation between N and these "I-am-psychiatrically-fine" responses found in hysteria, and that this correlation should be negative. Accordingly, the correlation was computed for a sample of 60 normal females and 60 normal males, disregarding their profiles. However, there are 5 items out of the 18 which are common to Hy and N (all five incidentally scored oppositely on the two scales). To eliminate the effect of this overlap of items another correlation coefficient was determined, this time based on N versus the 13 remaining items of Hy after the five overlapping were removed. The results of these calculations are in Table 14.

These four correlations all differ very significantly from zero ($P < .01$) although they do not differ significantly from one another in passing from the

TABLE 14
Correlation of Raw Score on N with the "O"
Items of the Hy Scale, for 60 Normals of
Each Sex, Both with and without
the 5 Items Common to Both

Group	N Versus the Entire 18 "O" Items	N Versus the 13 Non-Over- lapping Items
60 normal males	-.72	-.68
60 normal females	-.69	-.59

first to the second columns ($P > .30$). We observe that there is a marked inverse association between raw scores on N and raw scores on the non-somatic psychidenial items of the Hy scale, and that this relation is only to a small or insignificant extent a function of the actual item overlap between the two scales being correlated. This fact may be said to substantiate the prediction which prompted the investigation of the relationship, and to some degree to support the hypothesis regarding the nature of N.

THE ITEM OVERLAP BETWEEN N AND OTHER MMPI SCALES

As is usual in MMPI scales, there is considerable overlap of items derived on N and on other scales. Of the 78 items scored on N, 42 also appear on at least one other scale of the eight personality components, whereas the remaining 36 do not. Some of these remaining 36, however, appear on the lie scales or on the old C_h scale. In the case of overlapping items the mere number of common items is slightly misleading inasmuch as often the items are not scored in the same direction on both scales. Table 15 shows the number of items common to N and each of the other scales, and indicates the frequency of items in which the direction scored on N is the same or reversed.

For most of the personality scales

proper the overlap is quite small. The largest amount of item community for any of the personality components proper is only 7 per cent. In many cases it will be observed that the distribution of responses scored same and opposite is nearly equal. The most marked deviations from this trend are on Pt, C_h , L_e , G, and "+." In these five cases, every item on N is scored either consistently the same or consistently the opposite as it is on the other scale compared with N. In the case of the first three scales in question, the direction of scoring is in line

TABLE 15
Item Overlap between N Scale and
Other MMPI Scales
(Number of N Items = 78)

Scale	Num- ber of Items	Com- mon Items	Num- ber Scored the Same	Num- ber Scored Oppo- site
L	15	1	0	1
F	64	6	2	4
Hs	33	3	2	1
D	60	9	4	5
Hy	60	9	3	6
Pd	50	6	5	1
Pa	40	2	1	1
Pt	48	7	7	0
Sc	78	4	2	2
Ma	46	4	0	4
Mf	60	6	3	3
C_h	48	13	13	0
C_a	11	2	1	1
L_e	22	7	0	7
G	62	16	16	0
+	56	11	11	0

with the correlations found for the scales as a whole. Pt and C_h are positively associated with raw scores on N, whereas L_e is negatively associated. The significance of the item overlap with the unpublished scales G and "+" is not clear because the psychological (and even clinical) nature of these scales is unknown, they having been derived by purely statistical methods analogous to factor analysis without regard for any criterion

or clinical significance. So far as any evidence exists regarding G and "+" they represent some sort of a general factor of abnormal answering (whether due to test-attitudes or personality factors proper is not known) which permeates the MMPI items rather generally.

AN EXAMINATION OF THE TEN MOST POTENT ITEMS ON N

There is not any very obvious psychological interpretation of the whole group of N items which can be easily made from an inspection of their actual verbal content. As has been pointed out previously, they do have the common property (aside from the differentiating one which was used to identify them) of admitting to personal traits, attitudes, and experiences which are not admitted by the majority of the general population and have the character in most cases of being "undesirable" or "maladjusted" responses. It is of interest to examine the ten most powerful differentiators as judged by their proportional differences in the criterion groups. These ten most discriminating items are as follows:

TEN MOST POWERFULLY DISCRIMINATING ITEMS OF N, WITH THE RESPONSE SCORED FOR THE MORE "NORMAL" INDICATED

- G-35 I liked school. (F)
- E-49 It does not bother me that I am not better looking. (F)
- E-52 People often disappoint me. (T)
- F-7 What others think of me does not bother me. (F)
- F-47 Often, though everything is going fine for me, I feel as though I don't care about anything. (T)
- G-28 I am not easily angered. (F)
- H-35 A windstorm terrifies me. (T)
- H-36 I am not afraid of fire. (F)
- H-41 I have no fear of water. (F)
- H-52 I have no dread of going into a room

by myself where other people have gathered and are talking. (F)

E-52, F-47, G-28, H-35, H-36, H-41 do not appear on any of the personality scales proper of MMPI. Two of them appear on C_h and one appears on L₆ (scored reversed). All ten of these most potent items are responded to in the statistically "unusual" fashion more often by the criterion normals than by the abnormals; and as is apparent from reading them, they are all responded to by the normals in the direction which would be considered superficially indicative of mental ill health and maladjustment. The responses to H-35, H-36, and H-41 are certainly very mysterious as indicators of "normality," as are the others also to a variable extent. The writer is not prepared to defend any hypothesis as to the common property of these items psychodynamically, nor even to assert that they have a common property. There may be several "suppression" variables operating via these different items such that it would be a mistake to attempt an interpretation to hold for all of them. This question will be discussed further in Chapter VI.

ITEM OVERLAP OF N WITH THE "NORMAL" COMPONENT OF THE HUMM-WADSWORTH TEMPERAMENT SCALE

Since the "normal" component of the Humm-Wadsworth Scale was designed by the authors to measure the hypothetical normality or control factor of Rosanoff which gave rise to the present investigation, it is of interest to inquire what relation, if any, holds between the Humm-Wadsworth normal scale and the present. There are no correlational data available at the present writing, but some idea may perhaps be gained by a study of item overlap between N and the nor-

mal component of the Humm-Wadsworth. Of the 78 items on N and the 38 on the normal component of the Humm-Wadsworth, we find that there are seven items in common with almost identical wording. The items in question are indicated with their item numbers in the two tests:

<i>N-scale from MMPI</i>	<i>Normal scale of Humm-Wadsworth</i>
C-25	73
D-13	109
D-52	94
F-13	282
G-27	28
I-15	97
I-30	279

Although the number of items is too small to draw any definite conclusions, it is worth mentioning that in every one of these seven items, the response scored for "normal" on the present N-scale is opposite to the direction scored for "N" on the Humm-Wadsworth. It seems unlikely from this finding alone that the two are closely related, and it may be that their correlation is negative as would be suggested by this finding.

A MINOR EXPERIMENT IN "BLIND" DIAGNOSIS UTILIZING N

The distributions of criterion and test cases and the tests of significance indicate whatever amount of validity the N scale possesses for the discrimination of deviant-scoring normals from deviant-scoring clinical abnormals. However, a minor experiment was done involving an attempt to discriminate a group of profiles obtained from supposedly "normal" persons from a group obtained from hospitalized abnormals, using the N-scale as a basis. Three separations were of interest:

1. The separation obtained when the

judge merely looked at the profiles and was asked to pick out the actual abnormals, in the absence of any N score on the profile.

2. The separation obtained when the judge was in situation (1) except that the N-score had meanwhile been added.
3. The separation obtained when the profiles were mechanically sorted in terms of the magnitude of the N-score only, with no "judgment" based upon profile form and amount of elevation being permitted.

In order to reduce the difficulty and the fatigue of making the separation, the number of cases employed was not great; and this fact increases the difficulty of attaining statistical significance, by some tests at least. There were chosen at random (every third case, omitting college cases) from the entire group of deviant scoring normals (persons with any $T \geq 70$ but out of the hospital) 22 cases for "test normals." These were all males because of the lack of sufficient test cases among the females, which would have resulted in too large a share of the normals being members of the criterion group, thus unduly biasing results in favor of the scale. The profiles for these 22 male "normals" showing deviant scores on at least one MMPI scale were then drawn by a third party.

This third party also recorded profiles for 22 hospitalized abnormals taken at random from the hospital files, excluding cases with almost certain invalidity judged by elevated ? and F. Since the writer, who was to be one of the judges, knew a priori that all of the "normals" would have at least one score $T \geq 70$ (having been selected for that reason), it was necessary to include among the abnormals only those having abnormal profiles. The result is naturally to make the problem more difficult for the N-

scale, since from the previous findings we know that abnormals with non-deviant profiles tend to score high T's on the N-scale. All of the present 44 profiles, in other words, were "deviant" in that at least one of the eight personality scales showed $T \geq 70$.

These 44 profiles were randomized by the third party to the experiment and presented simply as profiles with T-scores indicated at the bottom, but without any score on N having yet been recorded on them. Two judges were tried in the separation, and the separation of the first judge was, of course, not known to the second. One of the judges was Professor S. R. Hathaway, co-author of the MMPI; the other was the present writer. The information accessible to the judge, besides the profile, was the statement that of the 44 cases, 22 were in fact the records of persons clinically abnormal and hospitalized; whereas the remaining 22 were profiles of persons who, so far as known, were psychiatrically "normal" and were not under a doctor's care at the time of taking the test but were out in the community. The judge was instructed that he must divide the 44 into 22 he guessed "normal" and another 22 he guessed "abnormal." This restriction as to how many could be put in each category was imposed on Professor Hathaway because the writer, of course, knew the group to be divided evenly while he was making his judgments.

There was a difficulty arising from the entire design which showed up in making the statistical analysis, namely, that the judges were sufficiently "good" at discriminating the normals from the abnormals regardless of N that in order for a significant difference in proportion correct to be algebraically possible with 44 cases the discrimination using N would

have to be practically perfect.

After the classification had been made by Professor Hathaway using the profile without N, the writer made his classification without N. Then the N scale was added by the third party, and the writer attempted another separation. The second separation by the writer was done several days after the first, but it is possible that the earlier sorting (although its correctness was not checked) exerted its influence on the latter. Nevertheless, the writer was found in subsequent analysis to have made a total of 18 changes (nine exchanges) in the second sorting.

For purposes of statistical analysis of the results of these three separations (two by the same judge but differing in that the N scale was on the profile in making the second separation) the null hypothesis was assumed to hold for a four-fold table. The goodness of separation was tested by the deviation of values in the four-fold table from this null hypothesis, using chi-square as the test of significance. On the hypothesis that the judges could not actually discriminate, the four entries in the table should be equal within errors of sampling. A fourth separation was carried through in which the entire set of 44 profiles was simply arranged in order of magnitude of the T-score on N, a median drawn between the 22nd and 23rd case so ordered, and the top half of the cases mechanically labeled as "abnormal" with the bottom half labeled "normal." The division occurred at between $T = 45$ and $T = 46$, about one-half sigma below the general population mean. This latter fact probably reflects in part the arbitrary exclusion of abnormals with non-deviant profiles, who occur in some considerable number in practice and would almost certainly have raised the median

score on N had they been included. Suffice it to say that this fourth separation was used to compare the results of a purely statistical and mechanical use of N with the judgment of persons more or

cally significantly, at least with this number of cases involved) by adding N to the profile. The most important finding here is probably that the N scale used in a mechanical fashion does al-

TABLE 16
Results in the Four Separations in the "Blind" Diagnosis Experiment

Separation	Per Cent Correct	Chi-Square	P	Significance of Deviation from Chance
1. Hathaway, without N-score	59	1.455	$P > .16$	Not significant
2. Meehl, without N-score	68	5.818	$P = .025$	Borderline
3. N-scale, working mechanically	73	9.091	$P < .003$	Significant
4. Meehl, judging with N-score available	77	13.091	$P < .001$	Significant

less "expert" with regard to the clinical interpretation of MMPI profiles.

The results of these four separations are given in Table 16 below, in terms of the value of chi-square from the four-fold table as well as simply the per cent of "correct" judgments out of 44.

We see that the first judge does not succeed in discriminating the normal from the abnormal profiles significantly better than chance would allow. The second judge does somewhat better, achieving a deviation from chance that has borderline statistical significance between the 5 per cent and the 1 per cent level. The separation of the second judge is improved somewhat (but not statisti-

most as well as the better judge when he makes use of it, and does better than either judge working in its absence. It should be pointed out that the second judge may have been aided somewhat unconsciously, both in his matching with and without the N-scale, by having seen a few of the profiles several months previously in another context, although it is doubtful whether this could have exerted any marked effect. In any case, the results using N alone are of the greatest significance in the above table. Even including cases with non-triad elevations and artificially holding all abnormal profiles up, it appears that N has sufficient differentiating power to be of practical value.

CHAPTER VI

PSYCHOLOGICAL INTERPRETATIONS OF N

THE PREVIOUS treatment of the N-scale in this paper has been confined almost entirely to a consideration of its statistical properties. Dynamic functions have been loosely attributed to the score in remarks such as "An abnormal female with all personality components under $T = 65$ cannot easily hold her T-score on N under 50" and similar comments; but no serious effort has been made to give a theoretical and psychodynamic exposition of what N is behaviorally and how it produces the discriminations which it does.

The present chapter is only a tentative effort in that direction. It is characteristic of research on the MMPI that very little attention is paid to the actual content of items, nor is it usually asked "How" patients of a given description come to respond in the way they empirically are found to do. In some cases, as in many of the items of the D-scale and all of the items on the Hs scale, the relation of the scored response to the clinical entity discriminated is "obvious," i. e., one can see (or so he thinks) *why* a depressed or hypochondriacal person would come to respond in the way indicated by the scoring key. In many if not most cases, however, this relation is either tenuous or altogether lacking. In the case of N we are in the unfortunate position of having to reason backwards from the properties of the scale and, to a lesser extent, from the nature of the items, to a guess as to its psychological "nature." From the findings in regard to other scales, it must be obvious that interpretations must of necessity be held with only very slight confidence in so far as they are based

upon study of the items. It needs also to be pointed out that all of the 78 items probably do not "function" in the same way to perform the empirical discriminations found, so that further statistical and experimental study of N needs to be done in order to isolate, at least for theoretical reasons, the components of which it is very likely a composite.

What we know of N experimentally is that it seems to enable us to distinguish persons who show deviant MMPI profiles but remain clinically "normal" from persons with equally deviant profiles who do not. This tendency is well established in the case of the scores on the neurotic triad, but it also holds up statistically although not usefully for the other non-triad components. There is only a slight tendency (none for one sex) for it to discriminate between normals generally and abnormals, when the rest of the MMPI profile does not enter into consideration.

It seems appropriate at this time to summarize the reasons which have accumulated throughout the preceding pages for not considering N, at least the major part of it, as measuring the "normality" factor hypothesized by Rosanoff and measured (purportedly) by the N-component of Humm and Wadsworth. I submit that the following findings render an interpretation of N in these terms extremely unpalatable:

1. The responses scored for normality on the present N-scale are in the vast majority of instances the statistically abnormal responses from the standpoint of what most persons say to these items.
2. The responses scored for normality are with practically no exceptions re-

sponses which admit to personal traits and attitudes of the sort that are psychiatrically undesirable and indicative prima facie of maladjustment and unhappiness, assuming them to be an adequate description of the person.

3. The raw score on N, which tends to be *high* among "normals" who have deviant profiles but low among abnormals, shows a marked and statistically significant deviation in the low direction in unselected college students and in adults from the normal population who have previously attended college. To reconcile this finding with the interpretation in question it would be necessary to assert that college students and college-educated persons have for some unknown reason considerably less normal, inhibiting, and controlling components in their personalities than persons in general, and in fact that such college persons have even less of this normalizing factor than an unselected sample of psychotic and neurotic persons.

4. The N-scale differentiates only to a slight extent between female abnormals and females from the general normal population, and it does not differentiate male abnormals at all. Regardless of the magnitude of the other specific components, on the Rosanoff hypothesis it would be expected that at least on the average, persons who have broken down psychiatrically would have less of the "normal" component than an unselected group of normals.

5. The raw scores on N correlate to a significant extent with the abnormal components of MMPI, both among normals and among the abnormals. In every case but one these correlations are positive, and in some cases (e. g., Pt) they are very high. There is no rational way to interpret such a finding in terms of a

positive "normal" component of the kind in question.

6. The raw scores on N correlate high negative with a subtle lie scale, and to a lesser but still negative extent with the old lie scale. While this fact does not contradict the Rosanoff interpretation of N, still it remains quite unexplained on that basis.

7. In the case of the non-triad components of abnormality, it has been pointed out that if a person shows deviations on these components and has a *low* raw score on N, he is almost certain to be psychiatrically incapacitated to the extent of being in a hospital. Whereas if he has *high* raw scores on N, he is by no means guaranteed against such a condition. The hypothesis in question posits a relation working both ways—too little "control" will result in break-down in the presence of slightly elevated abnormal tendencies; but conversely, an excess of control is supposed to prevent the appearance of abnormality even when such tendencies are present.

8. There are seven items on the present scale which occur also, with slight modifications of wording in some cases, on the Humm-Wadsworth "normal" scale. All seven of these items are scored in the opposite direction for "normal" on the present scale from the way they are scored on the Humm-Wadsworth. To the extent that the latter scale discriminates between normals and abnormals generally and functions in the way these authors describe it to, this fact argues against making an interpretation in terms of Rosanoff's theory for the present case.

For these reasons the interpretation of N in terms of a positive, dynamic personality component of "normality" or "control" which inhibits or regulates

those components of temperament which might otherwise result in psychiatric upset must be discarded. This is not to be construed as meaning that Rosanoff's hypothesis is false, nor that the Humm-Wadsworth normal component does not measure Rosanoff's variable, but merely that the present N-scale cannot be interpreted psychologically in that way. It would seem that *if* such a factor as Rosanoff's does exist, somewhere among the 495 items studied one should find items loaded with it. And this latter may actually be the case, since the reasons adduced against such an interpretation of N apply to the scale "as a whole," i. e., as manifested in the correlational and discriminating properties of the entire set of 78 items, which are probably not, as has been pointed out previously, psychologically homogeneous. It is conceivable that the minority of items which are "O" responses on the N scale are more akin to the kind of item that would sample Rosanoff's hypothetical normality factor. On the other hand, it may be that these items would turn out to be of the same nature as some of those on the Humm-Wadsworth—merely denials of symptoms of such severity or diagnostic portent as would hardly occur in normal persons. Thus, the first two items of N are what one would a priori expect to find on a "normal" scale:

"I am troubled by attacks of nausea and vomiting." (F)

"I have had attacks in which I could not control my movements or speech but knew what was going on around me." (F)

If all of the items on N were of this sort, the scale would be trivial and thoroughly uninteresting theoretically, but it would not be difficult of interpretation. If such items preponderated, one would simply have a case of listing those self-reported

symptoms which would practically never occur in normal persons but would frequently occur in psychoneurotics of the sort who score high on the neurotic triad. To be "normal" would raise your score in answering such items in the indicated direction simply because to be "normal" means being relatively free of such symptoms. From the evidence that has been adduced up to now, it is clear that the major part of N cannot be interpreted in that light either.

Another possible interpretation of N is that it is an "insight" scale in some sense. The persons who are in the hospital with Hs scores of 75 differ from those who are out of the hospital with Hs scores of 75 in that the latter have more "insight" into their own behavior and motivation and hence avoid the actual development of symptoms, even though they do in fact possess an equal amount of "hypochondriasis-component." This insight variable also shows up in a willingness to admit to oneself the sort of thing which appears on the N-scale.

This hypothesis is not nearly so easy to refute as the hypothesis which posits Rosanoff's positive control factor. In fact, a number of the findings can be brought to bear in support of it. The negative correlation with L and L₆ and with the 18 "O" items of the Hy scale would fall in line with such an interpretation; for it has always been emphasized by the MMPI authors that the "lie" scale does not really mean *lie* but a tendency to put oneself in a good light, whether conscious or unconscious. The tendency of N to show elevated raw scores in cases with high Pt to a much greater extent than in cases with elevated neurotic triad might also be brought as evidence, since the psychasthenic is notoriously more "insightful" at least

verbally than the hypochondriac or the hysteric.

Opposed to such an hypothesis, however, are numbers of bits of data. First of all, on this assumption there is no reasonable account of the scores obtained from college persons. We must assume that the average college student or college-educated person is about one SD below the mean of the general population in terms of how much "insight" he has into himself, which is conceivable to be sure but it is not supported by any other evidence. In the second place, we must assume that if a patient in a psychopathic unit with a psychosis or psychoneurosis is not deliberately selected for having a "normal" profile (as were the criterion abnormal to some degree), he has as much of this insight as the average non-hospitalized person—fully as much in the case of male abnormal and only slightly less in the case of female abnormal. In other words, the insight hypothesis fails to account for the finding that abnormal show up as having less insight *only* when their other scores are restrained from appearing abnormal, but not when these other scores are allowed to vary freely. What makes the difference here is the scores, not the presence of abnormal symptoms themselves, it would appear. If I am enough of a hypochondriac so that my relatives bring me to an institution for psychiatric treatment, I am likely to have as much "insight" as anyone else in the general population of normal persons, if this interpretation of N is correct. It is only when (in spite of being a hypochondriac) I fail to show very elevated scores on the neurotic triad that my "insight" may be expected to be poorer than average—and according to the findings, not very much poorer at that. In this line of thought, the insight

postulated comes to be primarily in terms of the responses to test items, and not in terms of whatever kind of insight might protect a person with abnormal tendencies from developing the clinical symptoms of a neurosis.

The correlation of N with Sc and, in the abnormal, with Pa does not tend to support such an interpretation either, nor do any of the positive correlations with abnormal components, nor those with G and "+." Furthermore, it should be remembered that some of the scales on MMPI are "symptomatic" themselves and cannot readily be regarded as measuring basic *components* which appear or are inhibited from appearing, as the case may be. Thus, in the neurotic triad scales on which the N-scale was originally derived, to get a high score often simply means to complain of those symptoms which, if actually present, would entitle one to be called whatever diagnosis is written above the score on the profile. To obtain a high score on Hs, for example, it is necessary and sufficient to say that one often feels weak all over, that one's hand shakes when he tries to do something, that one's head seems to hurt all over, that one is bothered by stomach trouble often, that one is not in so good health as his friends, and the like. These statements are the *complaints* which, if he believes them and talks about them and acts as if they were real, define the hypochondriac clinically. How is it possible to say that the "normals" who have sufficient insight as measured by N avoid being hypochondriacs in spite of answering the items in this way, when answering the items in this way is per se hypochondriacal? Similarly, in the case of depression, to say "I cry easily," "I brood a great deal," "I don't seem to care what happens to me,"

"I am not happy most of the time," and so on, would seem to be *prima facie* evidence of actual depression, if the person is using these words in the way in which most people use them to describe their psychological condition.

Lastly, consider the N items themselves. Persons who receive deviant scores on the neurotic triad but say that they are easily angered, that a windstorm terrifies them, that they are afraid of fire, that they have a fear of water (to mention some among the most potent), have a much better chance of being out of the hospital than those who do not assert these unpleasant things about themselves. It seems stretching the point considerably to say that for a person to admit that a windstorm terrifies him shows he has more insight than another person who does not admit this, and that, if they both have an equal amount of hypochondriacal, depressive, or hysteroid components, the former is more likely to remain out of the hands of a psychiatrist. It seems much more parsimonious to the present writer to doubt that all the "normals" who said "A windstorm terrifies me" were merely manifesting more honesty and insight than the abnormals who were equally terrified; and I suspect that neither group was especially terrified but that those who *said* that they were terrified were not actually so at all. This interpretation brings us to the last hypothesis, which will be defended as being somewhat more plausible than the others.

In terms of the list of possible ways in which a "test miss" can occur (Chapter II), it seems to the writer that the most reasonable place to assign the source of N's variance and discriminating power is under Group I, the "Errors of Measurement" category. This includes the

cases in which the personality variable underlying the scale response does not actually exist to the extent indicated by the score. In terms of this category, the "normals" with depression scores of 85 are not out of the hands of the psychiatrist because they have a lot of "control" which holds this depression component in check; nor are they outside because they have failed to show certain very gross signs and symptoms as part of their depression; nor are they outside because they have enough "insight" to keep their depression from incapacitating them. They are most probably outside because they are simply not nearly so *depressed* (in the biophysical sense) as the people inside, and their depression score of 85 is spuriously high because of some other personality or verbal factor. The present hypothesis is that it is this factor of "plus-getting" that is detected by the N-scale.

In terms of the list of possibilities given under "Errors of Measurement," categories A-1 and A-2 have been systematically excluded in the present study by the insistence upon very rigorous standards of validity in terms of profile form, ? and F scales. The only other possibilities are "Unusual interpretation of the meaning of the questions of a sort *other* than those interpretations related to the personality trait in question," and "The patient does not see the facts as they are." It has been pointed out previously that unusual interpretation as such, and not seeing the facts as they are as such, do not necessarily give rise to test misses, because of the deliberately projective element in the whole theory of the MMPI. But there may be some personality traits, as well as some quite superficial verbal patterns, which are almost unrelated to the variable with which a given item is

loaded and for the discrimination of which it was empirically selected. The traits and patterns are the "irrelevant" components of the predictor as described by Horst (16), and the hypothesis I wish to consider seriously is that the N scale acts as a suppression variable for some of these irrelevant components.

The detailed character of these tendencies I am, of course, unprepared to elucidate. The crudest description would be to say merely that they involve the tendency to get high scores on personality tests of the MMPI variety, which is saying little more than the empirical findings *except* that it points the way to an interpretation in terms of test-taking as such instead of one in terms of psychiatric personality components and dynamisms as did the other hypotheses. There have been observed by Professor Hathaway and the writer a number of persons who were excessively preoccupied with their own psychiatric status, who were ruminative and anxious about their own state of mind and their own feelings, who displayed the "psychasthenoid temperament" in a modified and watered form without being really psychasthenic in the sense of interfering compulsions and obsessions. Many of these persons are perpetual "plusgetters" on personality tests, and in fact an interest in them was, as has been pointed out before, the starting point for the present study. When the present investigation began, the Rosanoff-Humm-Wadsworth hypothesis had structured the field so that the failure of these high-scoring "normals" to be in a mental hospital or at least incapacitated for their normal functions was envisaged in terms of Category II instead of Category I. On the present evidence it seems more fruitful to conceive of these persons as simply mismeasured cases, and the

practical problem as one of improving the instrument so as to suppress the verbal-personality components that lead to the mismeasurement.

It is not denied in the present hypothesis that "personality" factors are involved in these components, whatever they may turn out to be. That they are allied to whatever the psychasthenia scale measures is clear from the correlation with Pt, as is also the case for Sc. But the difference between this and Category II-B-1 ("other traits inhibit the trait in question") is that the other traits involved do not inhibit the measured trait but rather prevent it from being accurately measured. Thus, a high N score does not indicate that the personality component measured by N or Pt is acting to "restrain" a man's depression component—an interpretation that is far fetched to begin with in the light of what psychasthenia is like subjectively—but rather that a man's depression component as well as his psychasthenia component are not so strong as the scores on D and Pt would indicate, because he has an excessive amount of whatever it is that is measured by N and produces high scores on psychiatric scales.

The extent to which N involves factors closely allied to psychasthenia proper as contrasted with more superficial and semantic trends leading to high scores on Pt and Sc alike will not even be estimated here because there is no evidence on the point. The correlation with Sc remains mysterious except for the minor point that there is an element of semantic confusion and unrealistic thinking, which is common to these pseudo-psychasthenic persons and to schizoid persons alike, but which in the latter cases assumes a much more malignant form. "I wish I could be as happy as other people seem

to be" may represent the distress of an acutely miserable person with a genuine depression. But it may *also* represent the superficial complaint of ruminative, unrealistic, semantically dissociated person who, to be sure may be feeling blue but has no real conception of what it means to be "depressed" as is a manic-depressive. It is something akin to this kind of characteristic that I have in mind when I use the phrase "plus-getting" as descriptive of N.

The high positive correlation of N with Pt and Sc, the high negative correlation of N with L_o, the high positive correlation with C_n, which is simply a "correction scale" (suppression variable) for hypochondriasis introduced because the old H scale was often elevated in psychiatric patients free of hypochondriasis, the high negative correlation of N with the "psychiatric denial" items of the Hy scale, all of these findings are readily understood in the light of the hypothesis that N measures "plus-getting" tendencies. That such a trait should show up correlated positively with almost all of the MMPI scales for abnormal components is to be expected. That it should differentiate normals with deviant profiles from abnormal with deviant profiles is understandable since it means that the normals are not deviant as their profiles suggest. On the other hand, that normals in general should not differ much from abnormal in general (unless the latter achieve normal profiles) is also understandable, since the N scale is not measuring a personality trait of intrinsic psychiatric significance but one which shows up as significant only because a test has been taken. The fact that a low raw score on N with a deviant profile indicates clinical abnormality means merely that if a person

does not have much of this plus-getting tendency, the "plus" responses he gave are likely to indicate a genuine psychiatric deviation. On the other hand, a *high* raw score in an abnormal person is not difficult to understand, since there is in this interpretation no reason to deny a lot of the plus-getting tendency to an abnormal person any more than to a normal one.

It must be admitted that the present hypothesis is hardly detailed enough to be theoretically satisfying, but it would be illegitimate to do more than hint at its nature on such flimsy evidence as now exists. To design experiments having this hypothesis in mind is difficult, for it is not possible to use as criteria of "genuine" psychiatric deviation any of the indicators—responses to test items, self-descriptions in interviews—which are presumably greatly affected by the factor hypothesized. One possible approach would be an objective (say, multiple-choice) test based upon those items on N which are answered in the scored direction, in an effort to find out whether the differences in interpretation could be elicited through verbal means alone. Thus, to say "I am afraid of fire" may be a sign of a phobic reaction in a clinically compulsive-obsessive patient. But this response also scores a point on N, in which case it may merely reflect a difference in interpretation on an almost purely semantic basis. There is, for example, an "obvious" sense in which everyone is afraid of fire, in that everyone will have anxiety if trapped by fire, everyone will avoid putting his hand in the flames, and so on. Yet the item frequencies for the normal population show that to interpret the item in this "obvious" way is a rare thing. The majority of "normal" persons do not say they are

afraid of fire, presumably because they interpret the question in a different way. A criterion "normal" who says "I often dream about things that are best kept to myself" gets scored a point for depression. If he says many things of this sort, he gets a high T score on depression. The unusual semantics, self-deprecation, verbal pessimism, or "psychiatric hypochondriasis" which leads him to respond in this way may also, however, lead him to say "I am afraid of fire" and by so doing raise his score on N. How often does "often" mean when he says he often dreams about things best kept to himself? To remove the ambiguity of that "often" would lower the discrimination of many MMPI items, by reducing them to mere self-rating devices. But when we leave the ambiguity in, we leave it in at a price, due to the existence of other variables besides those we are testing by that item. If N is loaded also by some of these other variables, and not heavily loaded with depression itself, then we can try to suppress the non-depression components of the D-scale by the use of N. An intensive study of deviant-scoring "normals" with high and low raw scores on N using a multiple-choice method might shed some light on this problem. When a subject says "I am afraid of fire," he could indicate later whether he interpreted the item to mean "Thinking of fire gives me anxiety," "I become rather nervous when near a large fire," "I would probably be frightened if caught in a burning house," "It is a fact that fire can be dangerous to a person," and so forth.

Of the various facts collected so far regarding the N-scale, the only one which does not actively support the present interpretation is the deviation of college groups. At least the present hypo-

thesis has the merit that, unlike the others, it does not come into any real conflict with this finding, although it can certainly not be said to predict it as a logical consequence. Whether college merely selects persons of sufficiently high verbal intelligence to minimize the probabilities of such semantic distortions as have been suggested remains to be verified or refuted by further experimentation. But that college educated persons should have less of a tendency than people generally to interpret personality test items in such a way as to put themselves in a "bad" light psychiatrically—possibly also as a result of greater verbal test sophistication—is at least a good a priori possibility compared with the implications of the alternative hypotheses regarding N.

In conclusion it must be admitted that the psychological nature of the chief factors contributing to the N score remains essentially undetermined on the present evidence. The hypothesis tentatively supported here leads to interesting and important possibilities for investigation of the whole problem of improving question-answer type personality inventories. As was suggested in Chapter II, it is not inconceivable that the future development of this type of personality measuring device will be aided in large part by the systematic analysis of personality-test-taking as a behavioral process. Such study will certainly increase our ability to construct more valid personality tests over what would be possible with the more traditional approach, which conceives of such tests as simple self-rating scales in which "error" is accepted as inevitable because of the basic untrustworthiness of human ratings and particularly those ratings demanding "objective" judgments concerning the self.

CHAPTER VII

SUMMARY AND CONCLUSIONS

THE ORIGINAL aim of the present investigation was the isolation of a scale of items on the Minnesota Multiphasic Personality Inventory which would quantify the hypothetical factor of general "normality" or "control" as described by Rosanoff and purportedly measured previously in the Humm-Wadsworth Temperament Scale. The procedure was one of empirically isolating items from the total item pool of the MMPI on the basis of their discrimination between a criterion group of apparently normal persons showing marked deviations on the three scales of the "neurotic triad," and a matched group of clinically abnormal persons with similar profiles. The scale of 78 items resulting from this analysis was called *N* for "normal" although this term here indicates its differentiating power only and does not imply any particular dynamic interpretation of its operation.

The *N* scale was found to have a fairly satisfactory odd-even and test-retest reliability, and was shown to continue discrimination when applied to a new "test" group of deviant-scoring normals. There was also evidence of some degree of generality in that it showed statistically significant differentiation between "test" normals and abnormals deviating on other components than the neurotic triad on which the derivation was based. This latter discrimination, however, was not nearly so good in terms of overlap and could hardly be of any practical utility. The overall finding from the study of the neurotic triad seemed to be that the typical deviant-scoring normal scores in the neighborhood of one standard deviation (raw score) above the mean of the

general population, and that the lower such a raw score becomes, the greater are the probabilities that the person is abnormal clinically. If a person shows a deviant profile and his *N*-score is as small as one standard deviation below the normal population mean, he is practically certain to be psychiatrically involved to the extent of actually being under psychiatric care. Exceptions to this rule are almost completely confined to persons who are in college or have been graduated from college; this group deviates on the average a full standard deviation from the mean of the general population. It was suggested that further investigation may warrant the construction of special norms for the college group.

It was shown that unselected abnormals tend to score about the same as the general population, with a slight tendency in the case of female abnormals to show lower raw scores than normals. In the case of abnormals, there was evidence that *N* functions as an "inverted lie" scale to some extent, in that hospitalized abnormals of heterogeneous diagnosis but with normal-appearing MMPI profiles tended to show low raw scores on *N*. A minor experiment in "blind" diagnosis indicated that the use of *N* by an unskilled person in a purely mechanical way would enable him to differentiate the profiles of normal from those of abnormal persons as well as or better than could be done by the judgment of persons highly familiar with MMPI and skilled in the interpretation of profiles.

It was discovered that *N* was correlated in varying degrees with the other scales of MMPI, in almost all cases significantly positive. The highest correlations were

with the scales for psychasthenia and schizophrenia among the personality components proper, and with a discarded correction scale for the hypochondriasis component. Negative correlations were found with scales indicating a tendency to put oneself in a favorable light, and with a sub-set of items from the hysteria scale measuring a tendency to deny that one has social or psychiatric maladjustments. There was also found a positive correlation of N with chronological age.

Study of the items themselves shows that they are in the very great majority of instances answered by the "normals" who have abnormal appearing profiles in the direction which is statistically abnormal in the unselected general population sample, and which furthermore is the direction that ordinary judgment would consider the unhealthy, unhappy, or psychiatrically undesirable direction. Those items on N which are common to both it and the "normal" component of the Humm-Wadsworth are scored in the opposite direction in every case on the two keys. This paradoxical finding combined with differential and correctional data leads to a definite rejection of any interpretation of N in terms of Rosanoff's active "normalizing" component as originally described.

Hypotheses as to the psychological nature of N are considered briefly. The interpretation of N as being simply a denial of the most severe and pathognomonic symptoms is immediately rejected because of its failure to distinguish normals from abnormals generally to any great extent, combined with the actual content of the items. An interpretation of N in terms of the greater "insight" shown by the deviant-scoring normals in admitting to the failings scored on N is admitted as more plausible but

finally rejected because of the college findings, the correlation with other scales, the lack of good differentiation between normals and abnormals in general, and the actual content of the most potent items.

The provisional hypothesis offered is that the N scale is loaded with a "plus-getting" tendency as regards personality tests, possibly allied to a kind of spurious psychasthenia shown by certain persons who often achieve extremely pathological scores on such tests as MMPI with a minimum of genuine suffering and complaint. The relative contribution of personality factors versus non-personological verbal and semantic factors is unknown at present. It is suggested that the markedly deviant neurotic scores shown by those normals who also have high raw scores on N reflect a distortion of response attributable to a component the influence of which must be artificially "suppressed" by the use of scales such as N in order to reduce test misses. In other words, abnormal profiles may arise from a genuine abnormality as indicated by the scale; but they may also arise from "deviations" which, while possibly related to various aspects of personality, are not the deviations for which the scale was built. To correct or partial out the influence of these other components requires that we first obtain measures of them, and the present N-scale is tentatively offered as an example of such measurement. It is suggested that a direct attack upon this hypothesis concerning N might be made via the use of multiple-choice situations applied to those persons who do deviate in regard to their N-scores, in order to find out whether the semantic-verbal distortions hypothesized really occur.

Finally even though the differentiating

powers of the present N-scale are not nearly so great as could be desired for practical reasons, the mere fact that significant differentiation exists with the items being of such a nature and scored in such a direction as they are, indicates that a pursuit of this lead in the field of personality testing may be well rewarded.

Once we have progressed to the point of sophistication which abandons the questionnaire as a simple self-rating situation and infers the significance of responses from the groups differentiated rather than the converse, new and important advances may be expected with confidence.

BIBLIOGRAPHY

1. ALLPORT, G. W. *Personality*. New York: Henry Holt, 1937.
2. ARNOLD, D. C. The clinical validity of the Humm-Wadsworth Temperament Scale in psychiatric diagnosis. Unpublished Ph. D. thesis. Univ. Minn., 1942.
3. BENTON, A. L. The interpretation of questionnaire items in a personality schedule. *Arch. Psychol.*, N.Y., 1935, No. 190.
4. BROGDEN, H. E. A factor analysis of forty character tests. *Psychol. Monogr.*, 1940, 52, 39-55.
5. DORCUS, R. M. A brief study of the Humm-Wadsworth Temperament Scale and the Guilford-Martin Personnel Inventory in an industrial situation. *J. Appl. Psychol.*, 1944, 28, 302-307.
6. DYSINGER, D. W. A critique of the Humm-Wadsworth Temperament Scale. *J. Abnorm. soc. psychol.*, 1939, 34, 73-83.
7. FISHER, R. A. *Statistical methods for research workers*. London: Oliver and Boyd, 1941.
8. FREEMAN, F. L. Toward a psychiatric plimsoll mark: physiological recovery quotients in experimentally induced frustration. *J. Psychol.*, 1939, 8, 247-252.
9. HATHAWAY, S. R., and MCKINLEY, J. C. A multiphasic personality schedule: I. Construction of the schedule. *J. Psychol.*, 1940, 10, 249-254.
10. MCKINLEY, J. C., and HATHAWAY, S. R. A multiphasic personality schedule: II. A differential study of hypochondriasis. *J. Psychol.*, 1940, 10, 255-268.
11. HATHAWAY, S. R., and MCKINLEY, J. C. A multiphasic personality schedule: III. The measurement of symptomatic depression. *J. Psychol.*, 1942, 14, 73-74.
12. MCKINLEY, J. C., and HATHAWAY, S. R. A multiphasic personality schedule: IV. Psychasthenia. *J. appl. Psychol.*, 1942, 26, 614-624.
13. HATHAWAY, S. R., and MCKINLEY, J. C. *Manual for the Minnesota multiphasic personality inventory*. Minneapolis: Univ. of Minnesota Press, 1943.
14. MCKINLEY, J. C., and HATHAWAY, S. R. The Minnesota Multiphasic Personality Inventory: V. Hysteria, hypomania, and psychopathic deviate. *J. appl. Psychol.*, 1944, 28, 153-174.
15. HOWELLS, T. R. An experimental study of persistence. *J. abnorm. soc. Psychol.*, 1933, 28, 14-28.
16. HORST, PAUL. *The prediction of personal adjustment*. New York: Social Science Research Council, Bulletin No. 48, 1941.
17. HUMM, D. G., and WADSWORTH, G. W. The Humm-Wadsworth Temperament Scale. *Amer. J. Psychiat.*, 1935, 92, 163-200.
18. HUMM, D. G. Discussion of "A statistical analysis of the Humm-Wadsworth Temperament Scale." *J. appl. Psychol.*, 1939, 23, 525-526.
19. HUMM, D. G. Dysinger's critique of the Humm-Wadsworth Temperament Scale. *J. abnorm. soc. Psychol.*, 1939, 34, 402-403.
20. HUMM, D. G., STORMENT, R. C., and IORNS, M. E. Combination scores for the Humm-Wadsworth Temperament Scale. *J. Psychol.*, 1939, 7, 227-253.
21. HUMM, D. G. Personality and adjustment. *J. Psychol.*, 1942, 13, 109-134.
22. HUMM, D. G. Discussion of Dorcus' study of the Humm-Wadsworth Temperament Scale. *J. appl. Psychol.*, 1944, 28, 527-529.
23. KRUGER, BARBARA H. A statistical analysis of the Humm-Wadsworth Temperament Scale. *J. appl. Psychol.*, 1938, 22, 641-652.
24. LEVERENZ, MAJOR C. W. Minnesota Multiphasic Personality Inventory: An evaluation of its usefulness in the psychiatric service of a station hospital. *War Med.*, 1943, 4, 618-629.
25. MALLER, J. B. General and specific factors in character. *J. soc. Psychol.*, 1934, 5, 97-102.
26. MALLER, J. B. Personality tests. In J. McV. Hunt, (Ed.), *Personality and the behavior*

- disorders*. New York: Ronald Press, 1944. Pp. 170-213.
27. MCQUITY, L. L. An approach to the nature and measurement of personality integration. *J. soc. Psychol.*, 1941, 13, 3-14.
28. MISES, R. VON. *Probability, statistics, and truth*. New York: Macmillan, 1939.
29. REICHENBACH, H. *Experience and prediction*. Chicago: Univ. of Chicago Press, 1938.
30. ROSANOFF, A. J. *Manual of psychiatry* (7th Ed.). New York: John Wiley, 1938.
31. ROSENZWEIG, SAUL. Frustration as an experimental problem: VI. A general outline of frustration. *Character and Pers.*, 1938, 7, 151-160.
32. ROSENZWEIG, SAUL. A dynamic interpretation of psychotherapy oriented toward research. In S. Tomkins, (Ed.), *Contemporary psychopathology*. Cambridge: Harvard Univ. Press, 1943.
33. ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt, (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944.
34. ZUBIN, J. Nomographs for determining the significance of the differences between the frequencies of events in two contrasted groups. *J. Amer. statist. Ass.*, 1939, 34, 539-545.