USING THE

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

IN COUNSELING

A SUMMARY OF SELECTED NEW RESEARCH RESULTS

by

Dr. Paul E. Meehl
Assoc. Professor of Psychology & Psychiatry
University of Minnesota
and VA Consultant

Supplement on Subtle and Obvious Keys
by Dr. Daniel N. Wiener
Chief, Advisement & Guidance
VA Center, Fort Snelling

Supplement on Back and Brain Injury Keys
by Dr. William Hales
Chief Clinical Psychologist
VA Hospital, Minneapolis

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

Research Results for Counselors

With the permission of Dr. Paul E. Meehl, and Dr. Daniel N. Wiener, the following pages of selected new research results on the Minnesota Multiphasic Personality Inventory have been reproduced for use with an advanced course in Psychological Bases of Counseling, taught by Dr. Milton E. Hahn, Dean of Students and Professor of Psychology, University of California, Los Angeles.

These materials were based upon a lecture to the Minnesota V. A. Psychologists, March 10, 1950. The notes were edited in the Advisement Section, VRED, Fort Snelling, St. Paul, Minnesota, and originally reproduced by the St. Paul Department of Education, August, 1950.

Item Lists and Norms for new keys have been omitted from the present edition.

Credit for the materials in the manual is given to Dr. Paul E. Meehl, Associate Professor, Psychology & Psychiatry, University of Minnesota, and V. A. Consultant. The Supplement on Subtle and Obvious Keys is by Dr. Daniel N. Wiener, Chief, Advisement & Guidance, V. A. Center, Fort Snelling. The Supplement on Back and Brain Injury Keys is by Dr. William Hales, Chief Clinical Psychologist, V. A. Hospital, Minneapolis.

> Original copy was not proofread by Paul Meehl. The present version was lightly edited to correct typographical errors and improve punctuation.
> — Leslie J. Yonce-Meehl, March 2020.

CONTENTS

PREFACE

Despite past indifference and hostility of psychologists to it, personality testing seems to be here to stay. How so important an aspect of human behavior could for so long be overlooked in counseling may someday be analyzed in a book on folklore in psychology.

The most widespread present objection to the use of personality tests in vocational counseling programs seems to stem from their association with serious emotional disturbance. The assumption is that one should attempt to measure personality only when a serious nervous disorder is seen or suspected. This would seem to be as reasonable as giving aptitude or interest tests only when high degrees of ability or interest are seen or suspected—if it were not for the many inadequacies of the personality tests themselves. Generally speaking, the sparse validity data which are available on personality tests are devoted to distinguishing between relatively maladjusted or emotionally disturbed groups, and so-called normals.

The Minnesota Multiphasic Personality Inventory, which is the subject of this manual, has many things to commend its use in counseling. In the latest group form, it is easily administered and scored, it is based upon unesoteric terminology, and it is solidly constructed in terms of carefully defined samples and empirical differentiations. But most important, there is a growing body of data on its use with normal populations to which Drs. Hathaway and Meehl are unceasingly adding basic data, and contributing searching hypotheses and syntheses. Living constantly with the test, their facts and speculations are continually providing bases for both more specific and broader interpretations. Their tireless and fruitful work with the test, which Dr. Meehl's exposition well illustrates, provides nourishment insuring its growth in scope and correction of deficiencies. Exactly those deficiencies of the test most commonly cited—in use of psychiatric nosology, in methods of profile analysis, in application to normal groups and common behavior—are the major focus of Dr. Meehl's lecture.

It was extremely difficult to edit the lecture as transcribed from a recorder, to make relatively smooth reading, while attempting to preserve the rare flavor of Dr. Meehl's verbal behavior. This manual represents a compromise which is neither as well-organized as a work which would have been originally written

for publication, nor as zestful as the original speech. It is issued now to meet the widespread demand from our counseling staff for a reference for MMPI interpretation.

This is not a basic manual on the use and interpretation of the Multiphasic. It assumes a fairly sophisticated reader—one who has read the test manual, knows the general content and interpretation of the original scales, and has used the test extensively. It is sparse in describing much of the published litera-ture, and is heavily weighted with hypotheses, tentative research findings, and suggestions for future research. It ends in the mood of most counselors who are using it extensively now: hopeful, critical, and vigorous in the pursuit of more information.

We are grateful to Dr. Meehl for giving the lecture and permitting us to pub-lish this manual. He would undoubtedly have corrected certain, we hope minor, errors which have probably crept in, but was unable to review this edition. We are grateful, too, for Dr. Hales' contribution to the program and manual.

The publication of this lecture does not, of course, indicate an exclusive endorsement of the test in the field of personality measurement. The VA encourages the use of most of the widely-used personality tests, depending upon individual counselor preference, and organization and counselee needs. It happens that more data on use of the MMPI is available in this area, through various organizations and consultants, than for other tests. It is hoped that similar manuals can be issued on other personality tests to meet the growing request from counselors, whenever they use any personality measure, for more information in this most vital and complex area of behavior-interpretation.

—D. N. Wiener

INTRODUCTION

As you might possibly predict, even though my topic officially was supposed to be Recent Research in Personality Measurement, I want to talk about a certain test of about 550 items with which I have some connection. But although I am going to use the Minnesota Multiphasic Personality Inventory as a basis for talking, and tell you some things that have not been published about it, I want to use it to illustrate some pretty general questions and problems in the field of personality measurement. Many of the things I am going to say, I am sure you will see immediately, are not at all con–fined to the Multiphasic, but have a general application to all structured personality tests and to many projective personality tests.

PROBLEM OF CRITERION

General Approach

Before I go to any concrete data, I would like to say something about certain aspects of our general approach to Multiphasic research with which you may not be familiar. The first thing is that these days we are tending to start with the test, sort people on the basis of it, and then take a good look at the people to see what kind of people they are. This, of course, is different from the way in which the test was built, and different from the usual psychiatrist's notion of a test where you start with groups of people already sorted on some basis—for instance, by formal psychiatric diagnosis —and you try to build a test which will guess or predict or agree with that diagnosis or whatever you use as a criterion.

At the risk of exaggerating, I am going to be a little flamboyant and say that this conception of the function of the test seems to me to be rather foolish. I don't mean to talk down the importance of formal diagnosis in certain situations, but in general my attitude would be, allowing for some exaggeration, that if I want to know what the psychiatrist is going to call somebody I probably should simply ask him. The idea that the primary function of psychometrics is to permit me, sitting in my little psycholo-gist's cubicle, to prophesy what the psychiatrist is going to say about somebody is, while still widespread, not a very powerful way of looking at the function of an instrument such as the Multiphasic.

Making Test Better Than Criterion

You all are aware of the terrible unreliability and fallibility of formal psychiatric diagnoses. There have been some better studies of them recently published. The unre-liability of formal diagnosis, even by a fairly good staff, is rather impressive. You might say, well, if it is so unreliable, how come you're even fooling around with an instru-ment which is built on the basis of the formal diagnosis of psychiatrists in the first place. The essential notion here (which I have bored some of you with ad nauseam in

the past, I know) is the notion of statistics in item analysis enabling you somehow to lift yourself by your bootstraps so that you are better with the instrument after you get through fooling with it, than is the criterion that you started with.

Everybody recognizes this notion in the sense of intelligence testing. For instance, if you ask why people were impressed with Mr. Binet's work, it was because his tests, little samples of behavior, would predict the opinions of school teachers, whether as ratings, or in the form of grades. But these days if the school teacher says Johnny must be dumb, he can't seem to do very well in school and all the teachers agree that Johnny must be dumb, and you have a good psychologist under appropriate conditions of motivation and so on, giving intelligence tests, and Johnny's IQ turns out to be 125, nobody says "My, my, my, the poor Stanford-Binet missed again." Rather, they say, "Isn't that interesting; what is there about Johnny's behavior in other respects that makes the school teachers unable to perceive how intelligent Johnny really is."

The interesting question from the standpoint of methodology is "How do you tell when you've got that far?" We have no good set of rules for deciding when our instruments have transcended our original criterion—when we are doing a better job than whatever it was we started with.

Theoretically it is obvious why such things can happen. If you think in terms of the statistics of the situation you could classify a lot of people erroneously and the statistical character of item analysis will generate scales for you which measure whatever it is you are talking about better than the original criterion did. There is no theoretical problem involved here. The question is the practical, not the logical, one: "How do you tell when you have gotten that far?" Some of the suggestions I have will be apparent as I talk about some of our recent Multiphasic research; but you should all be very much alive to the problem of thinking about this, in order to make more explicit the rules by which we can make up our minds that our tests are doing better than whatever it was that gave us confidence in the first place.

Now, you can overdo that line, of course. It is not valid to feel at home with an instrument over a period of time to the point that you begin to make it the criterion without having documented the argument that it is better than the real criterion. That's very easy to do, of course, especially in the personality field where everybody recognizes that the things we're trying to get at are somewhat subtle, difficult to detect, hard to observe and so on. It is easy to say, well, of course on the surface this fellow doesn't show his basic introversion or his basic extroversion or his real psychopathy or his this or his that, but the tests show that he has it. Sometimes that's all right, sometimes it's not. The question is, "How do you tell when it is?"

You don't have to assume, in the case of the derivation of scales such as on the Multiphasic that there are disease entitles, as some people have claimed, or that schizophrenia is a mental disorder comparable to measles as a physical disorder, or whatever. (I have my own prejudices as to that point, but you don't have to take any of these prejudices.) All you really have to assume to make sense of the whole enterprise is something like this—I can't state it precisely, I don't know how to: People who are sorted together by psychiatrists are, in general, more like one another in some important ways than they are like people sorted in some other way by the psychiatrists.

If you don't want to agree with that, then of course you just threw the whole business out, which maybe we should do—I'm not sure. But if you are willing to accept that little bit—people who are sorted together, under one category name, by a reasonably sensitive psychiatric staff, are more like one another on the average, with all the fallibility and unreliability and so on, than they are like the people who are sorted together under some other heading by a skilled psychiatric staff—if you will admit that little bit, you've got your foot in the door. Then some statistics, theory, and common sense should enable you to fill out, with a set of items, those components or dimensions which the psychiatrist is dimly getting at when he makes his formal diagnosis. That is the conception we have of the Multiphasic, at least at the present time.

Let me give you just one little example (I'll go into the real research material later) of the kind of thing I mean when I say "lifting by the bootstraps" and what I mean when I say "starting with the test."

I walked into Hathaway's office several months ago, and there was a stack of 21 case summaries—from one to three or four single-spaced typewritten pages, Multiphasics omitted—and he said, "Leaf through these and write down what strikes you." So I sat down and read through them and made little notes, and I came out with about eight or ten comments that I had some faith in.

There was just one thing that knocked my eye out, by which I mean there was not just a majority, but in all but one out of twenty-one—21 isn't a big N, but after all 20 in 21 is a big percentage—namely, low heterosexual aggressiveness. Twenty out of 21 cases showed it in different ways. The summary would say, "The patient was an attractive and sociable girl but did not seem to care for boys," or, "He was active in the Boy Scouts and Hi-Y but he was a little shy with girls," or, "Although he was a good-looking man with good health, and had always made good money, he had remained a bachelor because, as he put it, women did not interest him very much," or, "He never found the right one," or something of that sort. It showed up in various forms, sometimes

adjusted, sometimes not adjusted; sometimes rationalized and sometimes not—but just an overwhelming trend.

I was kind of curious; the diagnoses varied—psychoneurosis, hysteria, obsessional neurosis, involutional melancholia, drug addiction, and so on. So when I got done I said to Hathaway, "What in the devil are these people, where did you get 'em?" Well, they were all of the cases in our records, either sex, who had very safe validity indicators (that is: ?, L, K, and F) and for whom Pd was the lowest score. That was the only defining property. Hathaway just went through the files and looked for all the codes that had Pd as the bottom score and pulled out the case summary, and this is the kind of thing that emerged.

Well, I don't think that you would be likely to notice that rather interesting little trend if you were concentrating on diagnosis, if you were always asking the question, "How can I guess the diagnosis from the test profile?" But if you start with the profiles, in this case the peculiar arbitrary choice of profiles with low Pd, regardless of anything else, and regardless of how high it is also, here you have something that is a very impressive trend, even with an N of 21. That is the kind of thing I mean in general when I say "starting with the test and then looking at the people," instead of starting with a big group of people and trying to guess to what subgroup somebody belongs using the test.

DIRECTIONS OF RECENT BASIC RESEARCH

Need for Pattern Analysis

The second general thing that has been emphasized in our research lately (and I'm sure most of you are familiar with this) is the emphasis upon patterns or configurations. Now I get embarrassed always beating a drum about this point, but it sort of irks me that sane people don't seem to be able to assimilate it very well. Even psychologists who are always talking about globality and configurations, etc., when they talk about anything else—the interview, the Rorschach, or the TAT or something—don't seem to be able to assimilate the idea in connection with structured tests. If you are congenitally averse to patterns or configurations, it would be better for you not to use the Multiphasic at all. If you are built so that you can't be talked out of lifting up a Multiphasic profile and looking at Sc at the top as the highest score, but with blinders for everything else, well, then in my opinion it is a mistake that you have the test in your hands.

Nobody can be blamed for being a little naive about this point because of the fact that our publication of these things (as everybody complains) is lagging far behind our own thinking about it. Nevertheless, there have been a series of articles by people like

Gough, Schmidt, and myself over a period of three or four years that it seems to me should indoctrinate anybody fairly well with the idea that we think in terms of patterns or configurations. We do not think in terms of the single scores.

The third shift in our research emphasis these days that is very important to you as counselors—more so than to clinical psychologists, for example—is that we have departed increasingly from the emphasis on the usual psychiatric variants, and we have moved in two directions.

Developing More Common Language

The first is in the direction of common language. Psychiatry has recourse to the ordinary adjectives that we use in describing human behavior, but recently we have been concerned to find out what words, in garden variety language, are applied to people who have certain kinds of Multiphasic scores. Some of you have cooperated with us in the research we did on that subject a little while back which is still unpublished. We have a big mess of data and I'll give you some samples of it later.

That is, employers would be much more likely to use terms like "careless," "friendly," "worrying," "unreliable," or "proud" or something like that, than they would be to use terms like "obsessive," "compulsive," or "schizoid" or whatever; and we have been interested in finding out what kinds of words in ordinary discourse tend to be applied to people who have this or that sort of profile. This is the so-called adjective study which I'll talk more about a little later.

Developing Non-psychiatric Scales

Secondly, we have been interested in getting at some more technical psychological dimensions but still of a non-psychiatric sort, such as dominant, submissive, feeling of responsibility in the social-ethical sense, extroversion, and psychological status (that is, the psychological correlate of socio-economic status—socio-economic status seen from the inside rather than as the sociologist might see it in terms of a person's social position).

Some of that material, the St (social status) key, for instance, has been published; but two very interesting scales—one for responsibility, and one for dominance—have been developed recently and are not even in press yet. We've been writing the manuscripts and I'll give you some of the data on them. Keys for them will be obtainable from Dr. Hathaway's office before very long, and I think in your work you might find both of them rather interesting and helpful.

Problems in Coding Profiles

We have been doing a lot of work on frequencies of profile forms on the basis of the coding system that Dr. Hathaway invented. The problem in developing a good coding procedure has several aspects. For instance, you've got to figure out some way of conveying information about both the height and the pattern of the profile. We don't

know at the present time the relative importance of those two things—which are important for which purposes. And it is a difficult thing to combine those two sorts of information, or to decide whether they even ought to be combined, instead of being indicated separately in some way.

Secondly, there is the problem of getting enough information into the code, but not too much. The local war between Welch and Hathaway has gone on for some time. Welch had invented a more extended coding system that put in all the Multiphasic scores. Hathaway stuck to the original one he invented where you didn't code scores within a half a sigma of the mean on either side. Obviously you don't want to put in too much of the data; you might as well sit and look at the whole profile then. The function of the code is to abstract in such away that the mind's eye, so to speak, can grasp what is being presented. It is hard to know how much you can get in and make it pay off in terms of the whole function of the code.

Another thing is, you want any coding configuration system to be sensitive, but not too much so; That is, you want the code to represent in some way the difference between a situation like this, and one like that. [Indicating profiles on a chalk board?] The code doesn't do that now if one of these happens to be above 70, and one around 60. On the other hand, if you start playing around with too slight differences, then you may have three coded scores that are very close to one another, with slight variations well within the standard error of measurement which from some unpublished data seems to be best estimated at anywhere from three to six or seven T-score points.

Don't forget that! Slight fluctuations within the standard area of measurement may change the code pronouncedly. If you are doing research when you have a code file as we do, where you pull out the 37's and look at them, you are likely to have a problem of marked displacements as to position because of the fluctuations which are actually random and do not represent anything about the patients at all.

FREQUENCIES OF CODED PROFILES

Need for
Large
Samples

We have determined the frequency of occurrence for various codes in some detail, but I'll give you some data only for the first two digits. These interpretations will be based on 618 normals, and 1763 patients who were in the hospital.

It's a little hard to make general statements as to what characterizes the people with a given code because the frequencies get so small. That's a big problem. We don't know how to deal with it. For example, the commonest 2-digit code among normal males is 94 (with a peak on Ma, and a secondary peak on Pd). But to say it is the commonest is not saying much. In terms of its actual occurrence, it occurs only 4.3

percent of the time. You see, there are so many possibilities. That means to have ten such normal males to scrutinize, you need a sample of 233 normal males, so that the problem of getting enough people in order to be able to say anything with any confidence, at the same time getting it fairly detailed as to the configuration, is a very difficult problem.

We want to make some kind of a compromise. We don't want to be too superficial, we want a relatively intensive knowledge of each individual we're going to say things about. If it gets too superficial, you might as well not fool with it. After all, the test should do a few things you can't do by looking at people, so we want to have fairly intensive information for each subject; and yet we want to have an extensive sampling because we have got to have a lot of subjects in order to believe what we say about the subjects that we study intensively. It is not just a technical problem: You just don't have enough people studied intensively by one staff with one instrument, is what it amounts to.

Uncoded
Profiles

The most common type of Hathaway code for the normals is an uncoded profile, that is to say, one in which none of the scores is outside the T-score range 46 to 54. All eight of the scale-scores (we left out Mf because so many of our records don't have it) lie within one-half standard deviation of the mean. This occurs in about one-fourth of the cases, and that one-fourth occurs in each sex among the normals. Less than 2% of abnormals have no score outside the range plus or minus one-half standard deviation. But at least it's a relief to find that you have 12 times as many uncoded profiles among the normals as the abnormals.

Sex
Differences

The sexes are very similar as to the relative incidence of various code types, except for D. There are twice as many Multiphasic codes among normal women, with Depression as the peak, as there are among normal men, even though there are separate norms for the T-scores for the two sexes.

The most common peak score among normals (and I'm sure you've all observed this in your work) is Ma. That is true for either sex. Eighteen percent of the males and thirteen percent of the females have the Ma score as the highest.

Abnormal
Profiles

You are commonly asked what is the frequency of abnormal profiles in the normal population. If by abnormal you mean at least one score above T-70, for our sample, it is approximately one-seventh of the persons in the normal population who have at least one score equal to or greater than 70. Now, only between 1/20th and 1/40th of the normals will have T-scores above 70 on any one key, but of course you have eight keys to get up on. They are not exclusive possibilities so you just can't multiply it by 8, but you can see it takes a considerable rise. That's a healthy datum to keep in mind:

that among people in general who were not locked up, were not under psychiatric scrutiny, and so on (not that they're all normal by any means, but at least they're staying out of the psychiatric ward and surviving in the community), one in seven of them will have at least one Multiphasic score of 70 or above.

This contrasts with three out of four of the abnormals, who have at least one score of 70 or above. Keep in mind, too, that one out of four people who are in the hospital with an NP [neuropsychiatric] diagnosis manage to keep all their Multiphasic scores under 70.

Abnormals
vs
Normals

There are certain code types or configurations that arise much more often among abnormals than normals, and hence are more worthy of attention if you see them. That is, they are indicative of pathology in the probability sense because they occur relatively frequently among abnormal people and not very often in the normal population. For instance, whereas Ma is the commonest normal peak, a combination of Ma and Sc, or Sc and Ma, as the top two scores is rare, and occurs much more frequently in abnormals. That is something that one should keep in mind. If you have somebody with an Ma of 65 or 72 even, this nice peak for the normals is probably just for a salesman or something (that kind of interpretation is overdone I might add, but I'll say more about that later); but to have somebody with an Ma and Sc as the two highest scores is much rarer among the relatively healthy people, and much commoner among the sick.

The combination of D and Pd, that is, a 24 or a 42, is much more common among the sick. That makes sense if you think about it awhile. It is a peculiar constellation of things to find in a normal person.

The 27 pattern, D and Pt, occurs twelve times as often in the sick as it does in the well. I'm not talking about elevations now, I'm just talking about the patterns, regardless of how high they are. Sometimes they seem more striking if you talk about elevations. Twenty-five times as frequently among the sick as the well, you have pattern 27, both of them equal to or greater than 70. The combination of Pd and Sc is more common among the sick, but Sc and Pd in that order is not particularly more common. The pattern 87, that is, with Sc and Pt, is eight times as frequent in the sick as in the well. As a matter of fact, with either of these scores equal to or greater than 70, the pattern did not occur once among 258 normals.

There are some curve types which appear more often among normals than abnormals, such as the code where only Hy is coded. The combination, interestingly enough, Pa Pd, 64, is more frequent among normals than abnormals.

The profile where only Ma gets coded is ten times as frequent among normals as it is among abnormals. About the only abnormals that can manage that of course are the manics, a few of your alcoholic seniles, and a few aberrant conduct disorders that you just wonder about diagnostically. As a matter of fact, even if the Ma key is 70 or above —and you should remember this so you don't go around over-interpreting your findings—if it is alone in the code it will be three times as frequently found among normal persons as it would among abnormals even if there were the same number of normals and abnormals in the population. The reason for this I won't go into, chiefly because I don't know the reason, but I have some speculations that are not very well documented, and Hathaway and I don't agree about it anyway. I think it is a purely statistical phenomena and he gives it a psychopathological interpretation. We're probably both a little bit right.

INTERPRETATIONS OF CODED PROFILES OF ABNORMALS

<div style="float:left">Using Code Terminology</div>

One more point: If you can, get into the habit of using the code to talk about curves, instead of talking about the psychiatric category names at the top of the profile sheet. It's healthier, we feel. It's worst to talk about the schizophrenia key; it's better to talk about the Sc key; it's best to talk about code 8. That is, of course, entirely in line with what we were saying about starting with the test and looking at the people, instead of trying to guess the diagnosis. When you are working chiefly with relatively normal individuals, as you as a group are doing, it is still more desirable to avoid the psychiatric implication.

It's all very well to say, "well, we won't talk about the psychiatry of it," but the mind is such that if you always talk about the schizophrenia key, you can't help thinking about it; whereas if you talk about the 87's or the 23's, then you can set up relatively fresh associations with the significance of those numbers. Therefore, it is desirable to cultivate the habit of talking always in terms of the code. We have really tried to practice it, and some of us, I guess, are getting fairly good at it. It's not too easy.

<div style="float:left">Basis of Interpretations</div>

Another thing which we've been doing with the codes besides finding how often they occur, is to study characteristics of code types in the hospital population. Cutting across diagnosis, you just take all the patients who have a 13 or 27 or whatever it might be, and ask what kind of stuff they show. Take the 12's and the 21's for instance; that is, those people whose Multiphasic shows an Hs D, or D Hs peak. We've classified the incidence by things that are found in the majority of these cases, things that are found in one-third to one-half, and things found in one-fifth to one-third.

The actual values of these fractions are practically meaningless. It is only the relative amounts of these symptoms that make any difference, because to say they are found in these people means that the junior medical student who wrote up the summary was struck by it and mentioned it, and hence it got checked by the staff member. It's of relative significance only. If it happens in a third of the cases on the chart, it might happen to two-thirds of them in actuality. I'll stick to the interpretation of majority results since time presses.

<div style="float:left">12 and 21
Interpretation
(Abnormals)</div>

The people with 12 and 21 have, in the majority of cases, pain; this is regardless of diagnosis now. Some of these people may be called schizophrenics, psychopaths even. Pain, depressed, irritable, shy and seclusive, and somatic concern. Somatic concern is distinguished from conversion. These people have anxiety and worry about the state of their bodies. Two-thirds of the patients with a 12 or 21 in our sample are diagnosed "psychoneurosis."

<div style="float:left">23 and 32
Interpretation
(Abnormals)</div>

For the 23's and the 32's we had so few males in our sample that we analyzed tallies only for females, and the only thing that shows up in the majority of them is depression. Showing up in a strong minority are weakness, apathy, and agitation or tenseness.

<div style="float:left">27 and 72
Interpretation
(Abnormals)</div>

27's or 72's, that is, those with a D and a Pt combination, in a majority have depression, or are described as tense or nervous; and for a third to a half, a strong minority, we have listed anxiety, insomnia, and sensitive.

<div style="float:left">28 and 82
Interpretation
(Abnormals)</div>

For the majority of 28's and 82's, we find listed depression, anxiety or agitation; and in a strong minority, hysterical tendencies, excluding pain however. There is a collection of hysterical phenomena here like conversion, paralyses, or blindness or something like that, not pain. There is a pro-illness personality described as unsociable. There is mental loss in the sense that the patient complained he couldn't concentrate or there was psychometric evidence of it, or he said he was confused, or others said he was getting inefficient in carrying on his activities. They are suspicious or sensitive, and hypochondriacal. Heredity is bad, defined here rather crudely simply as psychosis in siblings or parents.

<div style="float:left">31 and 13
Interpretation
(Abnormals)</div>

The 31 or 13 code has become known as the conversion or the hysteroid valley, with D being below the two somatic variables. In the majority, two things appear: pain and something about eating. It might mean actual anorexia, or hysterical vomiting, or that the person complained of discomfort after eating, or that he ate too much, or something along those general lines. I haven't tallied it here, but I recall that last year in the seminar we were discussing the fact that the 13's and the 31's tend to put the pain in different places from the 12's and the 21's. The 13's and 31's get pain in the

head and peripherally, arm, back, head very commonly, eyes hurt, and so on; whereas the 12's and 21's go a little bit more for innards, lower bowel and that kind of stuff. When these people go for innards, they have pain up just level with the innards, with heart and precordial pain, and pains in the chest, and think they've got tuberculosis and the like.

An interesting finding that fits in with theory for a change, thank goodness, is a sizeable minority described by the medical student as sociable and extroverted. That's your hysterical personality, which again you don't find in all people with a conversion, but which characterizes them relatively more than hypochondriacs and obsessive characters and the like. And also marked in a sizeable minority (I'll do a little more selection of the more interesting things) is that they objected to psychiatric study. They took a dim view of the enterprise. They came in to the Out-Patient Clinic for their sore back and now they're getting processed in this psychic business, and they don't like it a bit.

64 and 46
Interpretation
(Abnormals)

Nothing shows up to characterize the majority of 64's and 46's. Very surprising to me. But in a sizeable minority, we have things such as depression (surprising to me) irritable, nervous, introverted, suspicious, judgment defect, and alcoholic.

68 and 86
Interpretation
(Abnormals)

68's and 86's have, in the majority, only one characteristic—paranoid delusions. In a strong minority there are four characteristics: depression, apathy, irritability and withdrawal. The majority of these, as you might expect, are psychotic.

78 and 87
Interpretation
(Abnormals)

If you move to the 78's and the 87's, you have two things showing up in the majority: depression and introversion; and in a strong minority: withdrawn, apathy, nervous, a worrier as described by himself or an informant, and irritable. Incidentally, the 78's and 87's are significantly younger than the other code groups, at least in our hospital sample. The difference is significant at the 1% level.

I'm afraid the things I read emphasize similarity in certain types more than differences. As you move into the still smaller minority you begin to get greater patterning. The strong pilings up are of course, in a large part, a function of the disposition of the medical students who say certain things about all kinds of people if they're sick. Striking things. Most people who are in the hospital, except manics and some psychopaths, are depressed in varied amounts, no matter what the formal diagnosis.

49 and 94
Interpretation
(Abnormals)

Well, I'll only give you one more of these. I've picked the ones where we had a big enough sample so we get some stability. The 49's or the 94's come as a little bit of relief after these other characters, I think. In the majority, they are over-active; and in a strong minority, irritable, violent (that is, they assault somebody, break the dishes,

etc.), talkative, extroverted, ambitious, and energetic. This group of 94's or 49's showed the biggest piling up, incidentally, of formal diagnosis, as you might expect. Of the 22 cases here, 16 are psychotics, five psychopaths, and one neurotic of all things. If you're going to have neurosis and have a 49, you might as well have an anxiety neurosis as anything else. It doesn't fit too well as is.

INTERPRETATIONS OF CODED PROFILES OF NORMALS

<div style="float:left; border:1px solid #ccc; padding:4px;">Method</div>

Well, let me read you some stuff about normals. This adjective study, on which some of you cooperated with us, was one in which we had people rate their friends, or counselors rate their clients, on a two-page list of adjectives taken from Cattell's list of traits that cover the personality sphere more or less. The N's are so small that we had to depart from our own rules. We had to go by single scales because if you figure out what the sample of about a hundred for each sex means, you know you can't do any statistics if you take even two digits of the code.

So this is based upon a kind of a half-baked attempt at pattern analysis, namely, what adjectives turn out to discriminate between high and low scoring groups. For example, you get a list of adjectives that discriminate, say, high from low Hy, and then you look at those two lists and say what adjectives are common, and you hope that there isn't any bizarre patterning relationship here that will knock some of those out. What I am saying here is that when I say high D and Pt, it was not based on an analysis of 27's because you wouldn't find enough to do an objective analysis. It's based upon high versus low D, independently of that high versus low Pt, and then asking what adjectives show up in both.

<div style="float:left; border:1px solid #ccc; padding:4px;">27 and 72
Interpretation
(normals)</div>

Well, for both sexes, the way people with 27 curves (in this phony way) look to their friends or counselors or relatives, is indecisive. They're frank, self-dissatisfied (in your psychopathology views, you might say that occurs as a reaction formation against their hostility, of course), affectionate, dissatisfied in general (not only self-dissatisfied, but dissatisfied period), and individualistic; and, receiving a negative loading, so that D Pt people tend not to be described this way: cheerful.

<div style="float:left; border:1px solid #ccc; padding:4px;">12 and 21
Interpretation
(normals)</div>

For Hs and D, that is 12, three adjectives emerge in both sexes: high-strung, soft-hearted, generous. Negatively, there is not self-control. Then there is a pronounced trend in the female, but not showing up at all in males, to be described as frank (I'd like to know what this frank business does here with these women). And, very nice clinically, we see the female of this sort described as amorous. I don't know what it means, but it comes out.

For 23's and 32's, we read Hy adjectives. We have four of them: worrying, affectionate, high-strung, and individualistic. Negatively (not characterizing these people), for males: balanced. Men who have D and Hy are not considered "balanced" by their friends; and females are low (that is, negative) for facing life and acting orderly. Males, only, are pronouncedly described as generous.

Here's a beautiful list, I like this one. This is Ma-Pd, both sexes. As a matter of fact, there are a few things that are very hard to understand, theoretically. There's a big loading—of all things—on worrying. Now you've got me, how that gets in there. That shouldn't be in there but it is, clearly so, and for both men and the women. Whether that means the significance of Ma and Pd in the normal range is altered, or whether it means that there are some peculiar aspects of the phenotypic behavior seen by judges which they misconstrue as worrying, I just don't know.

The rest of the list, for the most part makes it look as if the Ma-Pd in the normal range is just a nice watered down clinical one, and for that reason I'm inclined to the second hypothesis, but I have no good evidence of it. They are high on both of these two combinations: talkative, self-dissatisfied (that's a little surprising); enthusiastic, sensitive.

And then there's a special type of "sociable" which Cattell seems to have listed separately, and I think very wisely, as "social, forward." For my money, the most striking single thing about the Pd in a normal range is what I call a lack of social fear. He doesn't even have the normal garden variety of social fear, and it's nice to see that the statistics support that in indicating sociability of the forward type as characteristic. Versatile, high-strung, impulsive, verbal, amorous (marvelous characters, these), likes drinking (hardly an adjective but Starke has got it in there, thank goodness), rebellious, and individualistic. And then there's a lovely set of four negatives: people with Pd-Ma in a normal range are definitely not described by their friends as practical, orderly, balanced, and mature.

All right, Pa and Sc, 68, 86, both sexes. Here comes "generous" again. My own hypothesis, incidentally, about this "generous" is that it comes in these people from the inability of the neurotic to show the minimal aggression involved in refusing things to people—refusing favors, refusing to lend money, refusing to help. I think that if you studied the behavior of the judges that you'd find that this generous business comes about because of this negative factor—not from a warm giving, the milk of human kindness generosity, but from the neuroticism which makes refusal a form of aggression, an unacceptable anxiety-producer. That's just my opinion, I don't have any

statistics. Self-dissatisfied, sensitive, sentimental, soft-hearted, frank, high-strung, emotional; and in the negative, balanced.

You've probably got writer's cramp, and are also a little bored, so I won't be reading you any more lists, at least for a while.

Now the importance of the adjective study as I see it is that it enables us to talk with a little bit of empirical data about what characterizes normals that have this Multiphasic profile. When we get all these adjectives published, people who are using them to test in the normal range will not be completely out on a limb basing themselves solely on their clinical impression, or from a watering down of the syndromes; but they will have this pitiful little bit of data so that they know at least how these individuals look to someone else.

NEW KEYS

Introduction

Now I'd like to tell you about some of these newer keys we have been working on recently. Two have been developed jointly by Dr. Gough, who is now in California sending us manuscripts hot off the griddle, Mr. McCloskey in the Department of Political Science, and myself, chiefly for some research in the field of political behavior; but I think these two keys will probably have some considerable usefulness in general.

Dominance Key

First, dominance. That's a variable we talk about a good deal, and in several kinds of activity it is important to have it judged. We had the 15 most and 15 least dominant members of a fraternity and of a sorority at the university. The judgments were made for the sexes separately, but we pooled the men and the women to do our item analysis because the N's were too small otherwise. Then, also, we had the 50 most and 50 least dominant students in a senior high school class in Minneapolis. In the case of both of those groups, the judgments were not made by a teacher, but were made by a semi-sociometric procedure. The fraternity members rated each other; and in the high school class, the students rated one another as to their dominance.

I'm sorry, I don't have with me a copy of the definition we used, but it was in general along the lines of a tendency to appear strong and to maintain the ascendant role in face-to-face situations. Then there was a list of some examples (such as salesmen and the like), and there was the specific warning for judging not to confuse it with intelligence or with special advantages from having more money than other people had; and that it did not mean domineering; that some dominant people conveyed this feeling of personal strength even though they didn't particularly want to run things, etc.; that it was behavior, and not the person's desire to be dominant that we were

interested in; and this is not the person's self-concept; rather, it is how the person behaves as others see him.

We wrote 150 items specifically to get at dominance (the three of us wrote these items), and then we had the Multiphasic item pool also administered to these criterion cases. The entire pool of the Multiphasic was given to the fraternity and sorority groups, and about a hundred promising looking items from that study were then given to the high school group, together with the special ones we wrote. We don't have a good cross-validation and you should keep that in mind; we're going to collect more data. But we do have at least a semi-cross-validation in the sense that these items work for both sexes in the fraternity and sorority. We did our item analysis on the whole bunch for significance (but we did look at the tallies for the men and women separately). Then we moved a hundred items from the Multiphasic into the second sample, so it isn't as dishonest a validity indication as you might think, but it still needs good cross-validation.

Sixty-items were retained on the basis of these various analyses: 32 from the pool written for that purpose, and 28 from the Multiphasic. That's a very interesting thing methodologically, you see. We sit down and make up 150 items to measure dominance, of which 32 worked; and then out of 550 general items in the Multiphasic, almost as many worked. This is something you run into constantly in this type of personality testing—the inability, thank heavens, to predict beforehand what kind of item content will correlate with the criterion.

Regarding reliability, the only thing we have is the Kuder-Richardson Formula 21, which gives an estimated reliability of .79. Validity—not cross-validation yet, but correlation with ratings by others—was .60 in the college group and .69 in the high school group. So even though we haven't cross-validation, the values here are pretty high, quite encouraging. Correlation with self-ratings was .52 in the college, .56 in the high school group. I don't attach much importance to that: that's one bunch of self-ratings against another. Biserial correlation with social service points (high school students got points for doing special things around the school) was .33.

The mean differences are extremely large, even considering the fact that this is the original group. For the college group, the mean difference between the most and least dominant is about 2½ standard deviations, and for the high school group it's approximately 3 standard deviations; so even allowing for a fair amount of shrinkage, I think you can see this key has a good deal of validity in it. Correlation is .35 with socio-economic status as measured by the home index. The younger are significantly less dominant, but that's younger just in the sense that they're in high school, and I don't

know what explains that [difference]. I mean, you do have a socio-economic selection in college, and various other variables, so I don't know what to say about that [significant difference].

Well, that's a key that I think you might find it worth getting hold of and seeing what value it has in your situation.

Responsibility—that's something we certainly talk about. Is this person responsible, or not? Well, the definition we gave to the raters was: "Willingness to accept the consequences of his own behavior," dependable, trustworthy, showing a sense of obligation to the group—as others tend to describe him. Others tend to describe the person as, "you can depend upon him," "he is a straight shooter," "he always does his part," and the like. This was all put in the paragraph we gave the judges. There was another set of warnings to avoid confusing responsibility with intelligence, dominance, and popularity.

Here we had more samples: still we don't have a real cross-validation. You might say, why don't you keep some of these cases strictly for cross-validation. Well, we argued about that, but the feeling was that it was more important not to have some bum items in there that were specific to a certain population. We could get cross-validation of a scale later, but if we get some bum items in in the first place, we'll just get lower validity later. That is the reason we don't have any cross-validation yet.

But here we have the 40 most versus the 40 least responsible students in a ninth grade class, rated by teachers. Then we had the 30 most versus the 30 least responsible in a senior class, selected by the principal. Then we had the 50 most versus the 50 least responsible in a city social science high school class, chosen by sociometric judgments of fellow students. And finally, the 50 most and 50 least responsible in a college fraternity and sorority, sexes rated separately. We analyzed these groups separately so we are fairly comfortable about the scale. It is harder to get at, incidentally, as you'd expect, than is dominance.

Well, out come 32 items from the Multiphasic. We have a 56 item key for responsibility, but only 32 are in the Mult. The Multiphasic sub-scale, however, correlates .84 and .88 in two samples with the total responsibility key. It correlates .47 with ratings in the fraternity and .53 with ratings in the high school sample. If you just say "let's separate the most from the least responsible by an arbitrary cutting score," you get 78% hits in the college group, and 87% hits in the high school group, in the expected direction, thank goodness. Well, that's another scale that will be available before long which I think might be refreshing to play around with a little bit.

Mr. Friedman is doing his thesis on cerebral localization on the Multiphasic, of all things—an enterprise in which I took very dim view initially, but to which I am gradually getting converted. I always believed that the mind had something to do with the brain, but I don't know why I was so unwilling to accept this sort of stuff. Now he has some beautiful profile forms for the difference between people with focal lesions of the frontal versus people with focal lesions of the parietal lobe. It's just unbelievable as you go through the curves. I'm still not entirely convinced, but I can't see anything wrong with it.

He has also developed a key specifically by item analyzing the frontal versus parietal cases, and on the original group—now remember, this may shrink some—he gets 93% hits separating frontal from parietal lobe lesions. This is really hot stuff and better than the EEG or anything else we have around these days, so we can allow for some shrinkage and still have a very impressive result there. Judges looking at the profiles and sorting them are able to do 75% accurate separation. Probably the satistization of the profile pattern will do better than the judges, but will probably not do as well as the scale based on item analysis.

Well, there are some other things that are not so useful to you but testify to the inherent power of the item pool anyway and make us feel a little bit optimistic about what we can do. Dr. Hanvik, who took his degree here last year in the VA training program, has a low back pain scale that does a nice job separating organic and functional cases—people with slipped disc versus those who have functional back pain. Then we have some stuff on epilepsy which I'll let Dr. Hales tell you about while I get a breath, since I don't know much about that research. He's had more to do with it.

(Supplement by Dr. William Hales)

I'd like to tell them first, Dr. Meehl, about the backache scale. They might be interested in that. I don't know how many of you see people with backaches. There is a key, already mentioned, available for it. Here are the respective curves, with the lower curve for people with organic backaches. These are primarily disc cases, as contrasted with functional backaches. Notice that you get what amounts to a typical Hy-D that Dr. Meehl mentioned, in the case of the functional backaches. Research on the test is taking many devious paths, and it's really amazing that you take what was assumed by everyone to be a personality test and find that it can be used in separating out well-defined organic conditions.

We have gotten away, too, from the feeling (something I'd like to emphasize) that we can only apply one kind of a test to a specific type of condition. I'd like to mention in this connection, about the organic type of conditions, that we are getting more and more away from the idea that we will be able to find any sort of a scale which will measure or discriminate organic brain damage in its entirety, as an entity in itself. What we are probably going to come up with in the future, the way the trends look like now, will be many different types of scales or other measuring devices for attacking this problem of so-called organic or intra-cranial damage.

As you all know, for a long time the research was devoted to trying to find a single instrument which would pick out all persons with any kind of intra-cranial damage. Now effort is being directed to breaking up this idea of a totality and attacking it from various sides—for example, the parietal versus frontal, the dominance of one area over the other, laterality, and so on.

As an example, I have one of the first rough keys that was made for the parietal-frontal scale which I assume that Mr. Friedman will make generally available. We are at present using it to advantage, and as Dr. Meehl mentioned, we are getting a very significant number of hits. We can say that even now it looks very favorable.

(Return to Dr. Meehl)

I don't know whether the previously published St key that Dr. Gough developed is of any use to you, but you might take a look at it some day. Its items, scoring, and other data are in a series of articles in the American Journal of Sociology for the last two or three years. If you look for the word "status" in connection with the name "Gough" you'll find them. They're interesting because he isn't looking upon the key as a way to get at somebody's economic level—you see that's just like guessing a diagnosis, only worse. But he has, for example, asked a principal of a high school to describe a student who comes of a poor home and has a high St-score, the principal not knowing why the student is being picked; and conversely, a kid from a good family by usual standards, who has a low St score.

In these kids of the high status family with the low scores, you get the principal describing them as: "I don't know what's the matter with Filbert, his father is a pillar of the church and an influential man in the community, etc., but Filbert doesn't seem to take his responsibility seriously," and so on. Whereas a girl from a very poor home who has a high St score is described as: "She seems very ambitious, she spends a lot of time baby-siting to save money so she can go to music school, become a nurse," or whatever it may be. "She is one of the better students in our class and she works very

hard and she seems to have some trouble adjusting to her family," and the like. What-ever the important psychological properties of the St key are, I don't know, but it certainly is an interesting kind of thing.

NEW APPROACHES TO PROFILE ANALYSIS

<div style="float:left; border:1px solid #000; text-align:center;">Measuring
Profile
Similarity</div>

Another thing that we have been fooling with lately that we hope will mature into something useful (but hasn't so far) is the problem of quantifying the extent of similarity in profiles. Now a lot of people have worried about that—Bordin studied it, Molish did a paper on it some years ago, DuMas of Denver did a study on it.

We find the DuMas index quite unsatisfactory because it's based only upon slope, not upon magnitude; and secondly, more importantly, the index is an arbitrary function of the particular order that the scales of the test happen to be put down in, and that certainly is disadvantageous. We have found that the code is a little bit too fluctuant, especially for normal curves; also that it does not convey indication of such spike effect as I was talking about before. Ranks are, like the code, defective in this respect. You might think, well, we can do a regular rank order correlation between any two profiles. This certainly tells you something, but we have tried it, and have gotten to the conviction that it doesn't quite do all we would like to have it do.

<div style="float:left; border:1px solid #000; text-align:center;">Meehl Index
of Profile
Similarity</div>

Recently I invented an index of profile similarity which I would be grateful for any information about. It is based upon representing anybody's profile by first, the height—the elevation of the median of the top three scores (don't ask me why I chose that, I just wanted to get some elevation value in there and couldn't think of any other one; it took me about two days before I realized it's the same as taking the next to the highest score, believe it or not).

Taking the next to the highest of the top three scores as your reference line, you then obtain the difference (keeping algebraic sign) between each of the other scores and that value in T-score points. Then you record those differences. That's a set of differences for the first profile. Now those numbers convey all the information about the configuration. In some sense they have all the information because you can repro-duce the profile exactly from those numbers. If the second highest score is D, then I know the set of differences (algebraic T-score differences) for the other 8 scores (I used seven because our old records don't have Mf) and could reproduce the profile.

That set of differences would then obtain for the first curve. The same procedure is done with the second curve, not taking the same base line. If in the second curve the highest point is Pa, and the next highest is Sc, then the Sc is taken as the base—say 72

is the T-score—and we get a set of differences from it. And then we take the difference between the differences. Now don't get traumatized, this isn't as much trouble as it looks; you can do it in less than 30 seconds once you've done a few.

   Then if you think of a person's configuration, neglecting elevation, as being repre- sented in a hyperspace of seven dimensions, you say "What's Meehl's Multiphasic form," not "How high is it." Well, what is its form? Meehl could be located up here in the 7th space in terms of the coordinates representing the magnitude of these deltas (or differences). And then here's Hales' Multiphasic configuration over here. What's the difference between Meehl and Hales? It's the distance between them in the space, is one way of looking at it, so you take the old familiar geometrical principle of the square root of the sum of the squares. The discrepancies of these coordinates is how far apart two points are in the space, or how far apart Meehl and Hales are on their profile configuration.

So what we do is take the differences of the differences, square them; but before we square them, we round them off (this really terrifies everybody, but it correlates beautifully with what happens if you don't) to the nearest ten. That is, here's a difference of a difference of twelve, and here's a difference of a difference of 17, and there's a difference of a difference of five. You round the first delta off to one, another off to two, and the other off to zero. You just forget everything else, you see, and it horrifies your compulsiveness when you first do it. Then you can square up in your head, 1 plus 4 plus 0 plus 1, and the square root of whatever you get is the difference index. It tells how far apart Meehl and Hales are in the hyperspace of seven dimensions that characterize the profile.

You can't neglect the sign when you're subtracting, but you can neglect it when you write down the result. The convention is in case of .5's, to drop back. (You have to make a convention to get norms, so you just arbitrarily choose a way.) One squared plus 1 squared is 2, and so this difference between Meehl's and Hales' profiles is 1.414.

Now, of course, it doesn't have any absolute meaning. We still have to get a distri- bution of the values of that index for normals, abnormals, and so on. We have to find out what the degrees of difference represent. The rounded index correlates .95 with the unrounded index, so Hathaway tells me over the telephone, so don't let this bother you. We'll get a bigger sample and try it again, but it seems to be pretty close and will save a lot of clerical work.

We have correlations with three judges, that is, judges looking at Multiphasic pro- files asking simply how similar the pattern. The correlation of the index is .68 with one

judge, .65 with another judge (that one a non-psychologist), and .91 with Hathaway. Now, we don't have any way of knowing whether the difference between the index and the judge's idea of similarity in any given case is valid or invalid for the index. The eye together with the brain does some of this matching sort of stuff very well and does some of it very badly; and one of the great clinical problems is what kinds of things does the brain do better than the calculating machine. Well, why talk all day on that, so I won't. But that's a little thing you might fool around with.

We are thinking of trying to establish by various external criteria, not by just agreeing with judges, the similarities of people who have similar profiles versus dissimilar profiles. We don't know anything about the sampling distribution of this thing, we don't know anything about the effect of unreliability. It's just the first step, but we think it has some possibilities.

Certainly it's preferable to the one that Cattell published recently which involves a very unrealistic set of assumptions. His article had profiles that simply did not go with the indices. It's obvious that something is rotten in Denmark, and that you can <u>never</u> use an index that indicates the kind of relationships that his does. I think that the trouble is that we are being too compulsive about the mathematics of the situation. We ought to quit worrying about the factors and the components and the independent assumptions and all that, and say, "Let's try to make up some half-baked way of combining the numbers that will do at least reasonable justice to what the brain does to the profile form." I think that's the most profitable way to start.

Cross Product Approach to Profile Analysis

Well, another line of our research, so far as patterning procedure is concerned, has been going on for so long that everybody kids me about it; it's the so-called Meehl–Dahlstrom data. These data are about two years old. We seem to have an inhibition about working with them. Dahlstrom meanwhile has left town, which makes things worse, but the general approach I still think is promising. I have been working on it again recently and I hope over the vacation now to finish it off.

Essentially, there is still another way of looking at the problem of profile configurations and what is involved. We have taken as our problem, to distinguish neurosis from psychosis. Leave aside now the importance of that, that's a different question. But suppose you want, for some reason, to distinguish neurosis from psychosis—how can you do it? What we have worried about is the possibility—the likelihood, I would say—that the importance of the elevation of a given score depends upon the size of some other score, or upon the difference between some other pair of scores. That means that no kind of regression equation or discriminate function that just adds numbers, or

even takes squares and adds them, will do justice to the relationship. It involves, for example, cross products of some type.

I'm going to give you one example of that. Suppose you decide that a spike, when the whole profile is high, suggests involutional psychosis; and a spike, when the whole profile is low, suggests an anxiety neurosis or reactive depression; and the whole thing high, with not so much spike, suggests a manic-depressive psychosis. Let's just say that; so that if you want to consider that spike effect on the Hs, D, and Hy, you want to have some way of describing the differences.

Well, one obvious way to think of doing that is to say that what's crucial here is the difference between D and Hy, and also, between D and Hs, but that the difference is important to the extent that the whole profile is markedly up. So that would mean something like taking some weight of D, times the difference D minus Hy, plus another weight of D times the difference D minus Hs. Well, if you sit down and figure out how many cross-product terms, squared terms, and first degree terms you have (I forget what it comes to, 50 or something) for a 9 variable equation, the problem of working it out plus the problem of needing a tremendous sampling here…. Well, we said, "the heck with that," and got sort of graphical which is always a good procedure if you can't think through something.

What we have been saying is that we will assign weights to certain discrepancies, but we let those weights vary, depending upon the values of other discrepancies. [Referring to chalkboard.] For example (this kills me, but it isn't so hard to do, actually, after you've done a few), we have distributed Pt minus Hs along the abscissa, algebraic sign covered. Along here, we've got Sc minus D. You see the kind of game we're playing. We're trying to quantify the usual patterning remarks about the psychotic and neurotic band. Then we've got bands 1, 2, 3, 4, 5, so you get Pt minus Hs over here, and then that person has a profile where Sc minus D is such and such a value, and you locate him up here. All right, this band has a set of rules. There's a different set of rules for the second band, and the third band. So you look there and this delta—which is a function of Pt, Hs, Sc, D, Pa, Pd and Hy—the difference of this sum and this sum if this delta is greater than 15, call it psychosis; if it is less than 15, call it neurosis, unless schiz is over 80, in which case you call it psychosis.

It's the kind of thing that recently appeared in connection with the Rorschach work by Buhler, and it is an attempt to do something that ideally would be done with continuous functioning. You see, instead of multiplying and so on, having continuous changes, all you do is break up a certain difference into discrete steps of five, and the different rules within each band correspond to the different weight that you'd have if

you were using continuous variables. That's what it is; it's a half-baked graphical method of trying to do something in discrete form that is impractical to do with a continuous variable.

Now our aim in this is not primarily to help solve the administrative problem of "Who is a psychotic," although that is not entirely trivial—every clinician knows that some decisions are made a function of that—but we are primarily interested here in lifting ourselves by the bootstraps.

We are taking the patients who are diagnosed "psychoneurosis," who have a clearly psychotic profile, and who were diagnosed back in 1942 and 1943. We have a follow-up of several hundred cases whom social workers ran all over the state of Minnesota following up on: had they gone to state or private hospitals and for how long, or were they in jail, or had they committed suicide, or what was the situation. Naturally you can see what our hope is. We hope, and I believe myself, from looking into the case studies that go with these things, that a large number of the cases called neurosis but who had psychotic curves will, in terms of their subsequent histories, prove to have been in some sense really psychotic at the time they were diagnosed; and that there are some very important prognostic and other differences as a function of these configurations.

PROPOSAL FOR CONFIGURAL SCORING

Advantages
of Configural
Scoring

Another thing we've been interested in is what I call configural scoring. Instead of talking about the configuration of the profile, you go right back to the original behavior, namely, that of responding to single items. Now I'll have to bore some of you who have heard me on this topic before. I'd like to introduce this by a paradoxical example which would never exist in the real world, but it's fun to talk about.

Suppose you are interested in separating schizophrenics from normals, and you had a yes-no item which had zero validity for that purpose, and the validity of 0 didn't depend upon a particular method of item analysis. There was just no percent of people in the normal or in the schizophrenic group who hit the item true or false as the case may be, or it was at exactly the same, say 50%, difficulty level in both populations. So whether you do a chi-square or a phi-coefficient or a tetrachoric, no matter how you do it, the item has no validity, literally none, and I'm not talking about sampling here. Similarly we have a second item which also has zero validity strictly in the supply.

I'm fond of asking on PhD prelims, "Is it possible that these two items could have some validity for the criterion?" Almost to the man, the student will think (he thinks there's something rotten in Denmark, of course), but end up by saying "No, they

couldn't." He thinks that at least one of the items would have to have some little validity if the other one was to act, say, as a suppressor. But the interesting thing is that, algebraically, it's quite possible to have an item of zero validity, another item of zero validity, and the two items jointly having perfect validity. As I say, it would never happen in real life, but the fact that it is algebraically possible is important because it means that we should be looking for the approximations to it in the real world.

<div style="margin-left:2em; border:1px solid; padding:0.3em; width:8em;">Method of Configural Scoring</div>

Suppose, for example, that we consider item 1 which can be answered plus or zero; and item 2 which can be answered plus or zero. There are four ways to behave with respect to the two items: You can hit both of them, you can hit neither of them, you can hit one and not the other, you can hit the other and not the one.

Supposing we have 50 normal individuals who give you the response plus, plus; the other 50 normals give the response zero, zero. Now the difficulty of item #1 in a normal population will be 50% and for item #2 will be 50%. But now if you considered 50 schizophrenics and another 50 schizophrenics, then if anybody gives you the pattern plus-plus, he is one of these people and hence he's a normal. If he gives you the pattern zero-zero, he's one of these people and hence he is also normal. If he gives you the pattern plus-zero, then he's a schizophrenic; and if he gives you the pattern zero-plus, he's a schizophrenic.

So you look at those two items jointly, and don't fool with them by adding them up. You've got to look at them, so to speak simultaneously, in your scoring key. You've got to have a scoring key to give item number so and so, and item so and so, with some kind of a red line between them, or some other device. You look at this item, and you run your finger down here. If it says the same thing here as it says here that's OK— whether they both say true or both say false doesn't matter—but if they don't say the same thing, it's not OK, and you get one point for being a schiz. Now in any traditional item analysis these items would just fall through the hopper, wouldn't they? You would not discover them; they wouldn't seem to have any intrinsic validity.

Now it can be shown that you don't have to take such extreme cases as this, that if we consider a phi-coefficient in one population, say the schizophrenics, and the phi-coefficients in the other populations, say normals, that the configural validity will be one-half of the algebraic difference between those phi-coefficients when the symmetry is around zero. When the symmetry gets off from zero, it's a little bit higher than that, actually. So you have two items, both with zero validity for schizophrenia. But among schizophrenics, the first item had a phi-coefficient with the second of +.50, and among the normals, the two items had a phi-coefficient of –.50; and every time you got a TF or FT pattern you would say normal, and wherever you got a TT or FF you would say

schizophrenic, and the phi-coefficients of the validity of that judgment of normality or schizophrenia would be minus 50, or 100/2, which would be 50.

You get that obvious case when you have symmetry about zero. You can even work it when both of them have the same sign. If one of them has a phi-coefficient of .90 among the normals and only .20 among the schizophrenics, half the difference is going to be .35, configural validity is slightly over .35. As you get away from zero, you get the curve going up, positively accelerated, for the configural validity. Again, we know nothing about the sample distribution of that statistic; we have no information as to when configural validity pays off, if it ever pays off.

I have a couple of students doing some research on this problem with the Pa key, which as you may know is one of the feebler Multiphasic keys. When it works it works nicely, but it doesn't work worth a hoot on a lot of sharp paranoids; and the idea is that you could catch some of these cagey paranoids by a configural scoring system. It's a little harder, you know, to figure out how you have to answer two things simultaneously.

If you want to get psychotic and grandiose about this you can extend the reasoning a little bit. You could say, why couldn't we score by triads, or pentads, and really go to town. After all, that's what you do in a clinical interview. The patient says this and then you start out, and the patient says this and he has this slip of the tongue, and he has this dream, and so on, and so on; and the interpretations given are functional through the whole configuration.

Well, you can imagine some super-duper scoring key like this, you know—if the person says true to this, then you score him 5 points for schizophrenia if these two are in disagreement. So, in terms of the pattern, you would write this one: T F T or F T T are scored, but all of the other eight combinations remain unscored. You get socked for schizophrenia if you don't answer these two the same way if, and only if, you say true to that one. You can complicate scoring so it would really be unbeatable. A genius psychologist would have a hard time beating such a test, and it would contain, of course, many possibilities for dynamic interactions which our structured tests have ordinarily not contained.

* * * * * *

Well, that ends what I have to say about our local activities. How would you rather go on from here? I have page after page of research summarization, or we could have a discussion; whichever you think fruitful. I have recently reviewed all of the Multiphasic data, all 195 articles—I'm amazed that I can get up any drive to talk about the

Multiphasic at all, but I always seem to be able to manage once I get going. Whatever you think would be best, Dan. [DNW: I wonder how the group would feel? We can present these two alternatives to a vote: one, to have Dr. Meehl continue with his discussion, as long as we have the time; and the other, to open it up for general questions and answers now. (The vote showed preference for Dr. Meehl to continue his lecture.)]

* * * * * *

MMPI RESULTS FOR SPECIAL GROUPS AND BEHAVIOR

Let's talk about the research then. This is in no rational order; it's in the order that it was in a folder that I have, strictly random. Some of this research I am sure is familiar to you; some of it probably isn't so much so.

Characteristics of the Blind

We have some data on the blind. There seems to be very little difference between blind and sighted females on a Braille form of the Multiphasic. There is a tendency for blind males to be a little more abnormal than sighted males, especially on the Mf key. But if you look at means and not at the percent of scores over 70, the differences between blind and sighted are not significant. In other words, they are in the border-line range, apparently.

In terms of applicability of the test to the blind, it seems to be fairly good in the sense that when a group of presumably sentient judges, including some that did a good deal of work with the blind, pick items that would seem to be inappropriate for administration to blind individuals, these items tend to show up as not significantly different in the blind and the sighted. So it seems that it is fairly safe to use the Braille Multiphasic, of which there are several sets around the country and locally, for work with the blind provided that some allowance is made in the case of males as regards the interpretation of Mf. Whether you should say that this latter is invalidity or whether it indicates a feminization of males who are blind, it would be difficult to say. Again, as in most cases, my guess would be some of both.

MMPI & Interests and Attitudes

The Multiphasic does not seem to be appreciably intimately related to Strong scores nor to vocational choice in some studies. There are some significant correlations, but on the whole they aren't particularly impressive, from studies covering various domains like education, engineering, medicine, law, and journalism. There is a study indicating that students in nursing, music, liberal arts, and teacher training do not differ appreciably in their Multiphasic profiles.

There is a slight relationship between proneness to homesickness in college women and Hs, D, and Pt. There is evidence that Hs, D, and Sc are significantly related to radicalism in a college sample, and Hy to conservatism—very interesting set of findings. There is evidence that college students with a wide range of recreational activities show lower scores on F, Hs, D, Pt and Sc than those with narrow recreational interests. Students with high social introversion have been shown to be significantly less participant in college activities.

MMPI & Delinquency

There is evidence that the MF scale may be very easily faked by manifest homosexuals. We know that delinquent girls and also delinquent boys have an elevated Pd score. You should be aware that that widely quoted study on delinquent girls fuses three curved types and all you see is the mean curve which shows low neurotic triad, high Pd, and secondary psychotic elevations, Pt, Sc, Pa, Ma. But that if you look at the actual records, the curves break into a sizeable majority of psychopathic curves, relatively pure; a smaller number of psychotic curves; and a very small number, but still distinct, of neurotic curves, so that the means of Capwell's data are a little bit misleading.

MMPI & Alcoholism

Multiphasics of persons sober and slightly drunk show marked similarity in form, even when the elevation changes, and even when individual items change a lot. That is one of the most interesting findings we have. Give the Multiphasic, then give somebody two or three good strong cocktails and let him take the Multiphasic over again, and he gets a very similar pattern though he may have as little as 45% overlap in his significant item responses. The theoretical significance of that is great, but I don't have time to talk much about it.

Validity Indicators

Maximum separation of presumably valid from presumably faked curves is achieved if you call it fake when the result of F minus K raw scores is equal to or greater than 9. That is the latest dope based upon a study of 1800 or so presumably authentic records, and over 300 faked records from several different experiments. F minus K raw scores equal to or greater then 9 will catch three-quarters of faked records at the expense of 3% of the valid ones. This is faking bad or faking sick, not faking good.

There are two or three investigations indicating strongly that F scores over 70 are much more likely to be valid indicators of psychotic mentation or multiple neurotic complaints than they are of validity per se. When the F gets up to 70 or 75, you should not conclude that the test is invalid; in the majority of cases it will be valid. It will be an indication of the person's sickness. Even F raw scores over 16 are rather frequently valid although you naturally get suspicious when they get way up there.

But so far as deciding about validity is concerned, it seems safest to use the F minus K index jointly with questioning the patient and doing a retest. After all, if you manage to get a very similar profile, you are almost certainly getting some attention to the test. You may have some peculiar somatic distortions, but that's a form of psycho-pathology anyway—at least you're not getting the cards thrown at random. So it is preferable to question the patient about some items, having the retest and looking at the profile form itself, together with the F minus K index, instead of just looking at L, and looking at F, and looking at K.

MMPI
Results for
Saleswomen

There is some evidence that female saleswomen and factory workers show few significant differences from controls, but the impression I have from that study is chiefly that the differences are slight, even though they are significant. I am more impressed in general with differences being slight when I find them significant than I am with the mere fact that they are statistically stable.

Results for
Insurance
Salesmen

Insurance salesmen are significantly above the standard group on Hy, Mf, Pa and Ma; and those with high Kuder Persuasive scores have lower Hyperchondriasis and Depression scores than salesmen with lower Persuasive scores.

Results for
Social Workers

Social workers with lower Social Service score on the Kuder have significantly higher Sc on the Multiphasic.

Rehabilitation
Trainees

Unsuccessful rehabilitation trainees have somewhat higher Multiphasic scores than do successful trainees for four occupational groups studied.

Actors

In an unpublished study of male actors, professional actors on Broadway scored significantly above controls of similar age and IQ, on Pd, Mf, Pa, and Ma. The mean Mf of these male actors is 77. Very interesting, and what you would expect theoretically, of course.

Financial
Aspiration

Students from upper class homes but with low St scores on the Gough St key have the property I mentioned before, and also tend to anticipate lower income. Just ask a bunch of kids in a class how much money they expect to be making at a certain age. The kids with a low St score from upper class homes expect lower incomes than kids with high St scores from homes of a similar economic level.

Group vs.
Individual
Forms

Wiener showed that the mean scores of the group versus individual form based on testing odd and even cases are not significantly different although 7 of the 9 show a higher score on the individual. Standard deviations are also very similar, although they were not tested. If you test the same subjects on the card and group forms rather than taking cases randomly as Wiener did, the correlations range from .56 in the case of Pa,

to .91 in the case of Mf, with a median of .60, which values are fairly close to the best estimates we have on the reliability.

Femininity

If you throw together the Mf scales of the Strong, Kuder, Multiphasic and some other test, you find that the Multiphasic correlates more with the other three than they tend to with one another. Whatever they are all doing as a mass, the Multiphasic is doing a little more, as far as femininity goes.

Neurosis and Medical Treatment

Patients in a general medical out-patient clinic were routinely tested. Everybody coming in was given the Multiphasic regardless of his complaints. Those who have two or more neurotic scores on the neurotic triad above 70, if they are contrasted with those that have no triad score above 60, have three times as many heavy charts— charts weighing 6 oz. or more. Hathaway just took the hospital charts and weighed them. Very interesting operational way of getting at validity, you see. These people have neurotic scores coming in, multiple diagnostic procedures are carried out, they come in again and again, they get sent from one clinic to the next, and so on. I love that study.

Predicting Outcome of Therapy

If the criteria of outcome are reliable, the Multiphasic can be used to predict, better than chance, the results of in-patient and intensive out-patient therapy along short-term psychoanalytic lines. Adverse signs for such therapy are high Pd, Sc, Pa, or in general having other scores higher than Depression, or having Pd or Sc above Pt. The study was done in California.

Insecurity

Multiphasic Pt correlates .52 in men, and .45 in women with a rather carefully validated measure of insecurity. This makes theoretical sense and also fits the face value of the character of the items in this case.

MMPI & Religion

Seminary students have elevated Mf scores, a mean T-score of 63, which certainly you would expect. Local clergy, by the way, have elevated Mf, Hy, and a little L, which you can interpret in several ways. Religious students at the University of Minnesota are reliably less depressed than non-religious, significant at the 1% level, and over a sigma different, which is interesting, even when age, sex, and total membership in organizations are held constant. The non-religious students are more abnormal in all but the Ma scores, in varying degrees.

Business Students

Non-successful students in a school of business have higher profiles than successful ones; in the case of Ma and Sc, the difference exceeds 5 T-score points.

CAUTIONS AND CONCLUSIONS RE TEST INTERPRETATION

<div style="float:left">Major Errors in Interpretation</div>

I have a few general comments that I can spend the next several minutes on. The two most important errors I would say are being too cautious, and in not being cautious enough. I would like to warn you that it is not safe simply to water down clinical syndromes to decide what people in the normal range are like; and, in general, I'm suspicious of making vocational extrapolations only on theoretical grounds. I am not suspicious, naturally, when the data are there, and I don't see any reason why a person should not allow himself to think about these possibilities. But I do believe it is not wise to make straight out vocational extrapolations from theoretical considerations, because the theory, so far as I'm concerned, of vocational interest as related to psycho-dynamics is not sufficiently well-elaborated at the present time. Remember that the odds are for normality just on plain garden-variety bases. Knowing nothing else, a person is probably not very sick, and therefore you should be careful in interpreting, especially minimal deviations.

<div style="float:left">Blind Analysis</div>

The instrument, like any instrument, ought to be used with all the data. I am in favor of blind analysis for certain research purposes, and in some clinical, settings, even I, as opposed to most clinicians, would defend blind analysis, because I believe the kind of corroboration given blindly has a higher weight for cases where the interpretation is subjective. But on the whole, you can look upon blind procedure, as Murray says, as a nice parlor trick, and not the optimum way to use an instrument in a clinic setting.

<div style="float:left">Reverse Discrimination</div>

Keep in mind the possibility that the validity of an item may actually undergo a reversal in the normal range. I don't have time to really develop the statistics of that and I don't understand all of them anyway, but there are some items on the test which are discriminating backwards as long as we are working outside of the abnormal population.

It may well be that if you consider "normal" non-suspicious people, "normal" paranoid suspicious people, and full-blown diagnosed psychotic paranoids, some of the items which discriminate this latter group from the other two lumped together, may not be functioning statistically to discriminate between these two groups. I have in mind particularly items like Pa-subtle. Dan and I carry on mild warfare at all times about the meaning of the subtle scales, and I don't have any better hypothesis than he does, and no worse, I think.

But it is possible, at least, that if we take these items like "Some people are so bossy that you feel like doing the opposite to what they say even when they are right,"

that we should score it for Pa. Answered "false"—in other words, "No people aren't that bossy"—it may represent the rigid pseudo-objectivity that the full-blown paranoid shows clinically so often, a phony rationality about the wrong things. When you move into the normal population, that response may be characteristic of the more healthy individual and to say "true" to it doesn't give you a Pa score, but in the normal range it may be indicative of a little bit of that garden-variety suspiciousness that normal people have. And I think particularly on all of these scales which have sizeable components of subtle items, that there is a serious problem of interpretation which needs to be independently worked out by such things as the adjective study.

Need for Special Keys

Further, I believe that it is dangerous to play too fast and loose with extrapolations from the hospitalized population. As always, I advocate building scales for specific purposes. I realize there are objections to the tailor-made approach. It requires effort and so on. It's nice to have a domain already covered. But if a certain specific purpose is constantly coming up and is very important, it would seem to be better to make a set of items for it rather than taking whatever intrinsic but obscured validity for that purpose is buried in various places for the other groups.

It doesn't always work. You can't do it because there are complicated patterning and suppressive functions that you can't get at so readily that way. We've tried to make a straight psychotic key, for example, for neurosis versus psychosis, and it didn't work worth a hoot. It should have, but it didn't. It doesn't work nearly as well in our present group, at least as the set of differences does.

Over-cautious Use of Test

On the other hand, I think some people are too nervous and too anxious about the test. I hear people saying, "Gee, maybe I shouldn't be using this." Now it hardly hurts you to have a little more data, as Hathaway always points out. As long as you don't go off the deep end, what harm is there in having an instrument that can contribute something, as long as you're reasonably careful? I don't think anybody should confirm or disconfirm the test score by his own judgment, and I don't think he should do the converse either. It's a matter of raising the probabilities a little bit. Sometimes the test is seeing something you are not seeing, sometimes you see something that the test doesn't see.

Problem of Vocabulary

I think everybody who uses this test in the normal range—and even in the abnormal range—should practice the development of a harmless-sounding vocabulary for psychopathology. You should not suggest to people that they are schizophrenics or homosexuals, or whatever. You all know that, I realize, but it's a little hard to develop the skill of talking in a harmless way about something. It depends upon the kind of counseling you do whether you get into this box at all, but if you get into such

situations, you should practice a set of words. There are for most of these keys, even for the more sick and malignant ones like Sc, a few nice little words that tend to antagonize or frighten people, and yet, on the other hand, don't make you look like a liar in your own eyes. Some compromise between those two aims has to be achieved, of course.

<div style="float:left; border:1px solid; padding:2px;">Vocational
Implications</div>

So far as work with normals is concerned, there's one other point I have. I think it is less dangerous to extrapolate when you are talking about general personality characteristics than when you are talking specifically about vocational things. In other words, I am less nervous, even if I didn't have the adjective study to go on, about saying of somebody with a high Pd, "I think this guy is a little on the irresponsible side," something like that, than to say (take an absurd extreme), "I think this guy ought to become a watchmaker because he has so much Pt, and it's kind of compulsive fussy business making watches."

This latter kind of extrapolation partakes of large elements of danger. The specific vocational and interest aspects I think remain to be established, although some very interesting leads have been presented by Harmon and Wiener and others. I'm leaning over backwards in caution here, whereas there are some obvious extrapolations.

In almost any kind of work, you don't want some borderline psychopath, and it's more serious if he's going to be an airplane inspector than if he is going to be a janitor.

Well, I have one final set of stuff on how to improve the test and make a better Multiphasic some day, but that would take the rest of the day. I wouldn't want to run through it without doing it justice, so I think I'll just stop at this point.

* * * * * * * * * * * * * * *

(Observations by Dr. Wiener)

<div style="float:left; border:1px solid; padding:2px;">Local
Counseling
Problems</div>

I think we can agree that this has been a very solid and interesting presentation. I have made a few notes, rather than trying to summarize this very extensive lecture, which might point up some of the problems that we've faced as we've done our work, and which tie in with some of the things that Dr. Meehl has said.

He raised a question at the beginning which was extremely significant, although he didn't try to answer it. The question was whether the Multiphasic is now good enough so that we can do with it what we do with the Binet when we put results at variance with the usual criterion: we often wonder what special factors in the usual criterion may be defective rather than immediately dismissing the test as invalid

Also especially significant was Dr. Meehl's stressing that we should start with people and then look at the test, rather than vice versa. This approach probably more than any other accounts for the substantial contributions of the group under Dr. Hathaway and Dr. Meehl in the field of personality testing.

The importance of a new language has been one of the more serious problems that we have faced in using the Multiphasic. We've certainly tended to develop the bad habit of speaking of schizophrenia, paranoia and so on, in referring to the test results, before we had personally absorbed enough data to use them in the proper clinical sense. Rather than stigmatizing our clients with serious diagnoses, however, it seems that the strength of the terminology has become considerably vitiated in its frequent application to relatively normal people.

Persons in the "normal" population who have T-scores above 70 have frequently posed a problem of interpretation, and it occurred to me as Paul was speaking that these individuals might often have high K scores reflecting subtle scores. This may be a rather desirable feature in a normal population, as Dr. Meehl pointed out toward the end of his lecture, so that the misses in a sense may be misses because of the original validation of the test on the hospital population. [Meehl: With no evidence at all, I'd bet that that would be the case.]

For the benefit of those of you here who may not know it, I gathered that the descriptions on the scale will be out in a Multiphasic atlas soon, Paul? [Meehl: Yes] Do you have any idea when that will be out? [Meehl: Well, the first volume is supposed to appear in the summer or the beginning of the fall; the second volume probably after Christmas sometime, I'm afraid. I'm not 100% sure what will be in which. A big mess of case studies will be in the first one. How much of the adjective stuff will be in the first one I don't know. We haven't decided on that yet. I'm trying to get started with making a little mimeographed list of the differentiating adjectives so that we can circulate that among people who use the test before the atlas comes out. I'll keep twisting Hathaway's arm; I think we can manage that.] We'll be looking forward to publication of the atlas; it will certainly be a valuable addition for personality interpretation.

In analyzing profiles, the further we break results down, the more we run into the problem of the extent to which an individual counselor can comprehend the configuration of the various patterns. Any breaking-down attempt runs along with attempts toward synthesis, toward simplifying the basic areas in which human behavior is interpreted. Dr. Meehl's talk today illustrates well the attempt to get at the complex differences in human behavior, while Rogers seems to view diagnostic categories as a

simple continuum of seriousness of problems generally, at least for therapeutic purposes.

THE SUBTLE AND OBVIOUS KEYS (Dr. Wiener)

<div style="float:left">Rationale for Subtle–Obvious Keys</div>

Meehl and Hathaway have described the need of a scale to indicate the effect of test-taking attitudes upon scores obtained on structured personality inventories. Their conclusions were that there is a conscious or unconscious tendency for subjects to present a picture of themselves that has a considerable influence upon their personality test scores; that this tendency might be to place them in an overly favorable or unfavorable light; and that the present "validity" scales of the Minnesota Multiphasic Personality Inventory did not seem to be sufficiently subtle to detect this test-taking attitude.

The concept of relatively subtle and obvious keys for the scales of the Minnesota Multiphasic Personality Inventory developed to meet the problem raised above. It was felt that the development of such keys on individual scales of the MMPI would yield more information and be of more practical usefulness than an overall validity scale.

The main problem of the counselor working with a relatively normal population, as differentiated from the clinical psychologist working with a seriously disturbed group, is to distinguish non-disabling personality factors that characterize counselees like aptitude and interest test results do. The seriously disturbed group can probably be distinguished by a test consisting of items obviously indicating deviate personality characteristics, since the extreme deviates are mainly unaware of the significance of their symptoms. Screening devices developed for the military services and for private industry which consist largely and obviously of deviate items, probably owe their success to this fact.

To help the counselor working with a normal population, however, a much more subtle test is required which will both distinguish the extreme deviates, and also the characteristics of "normals." These two services of a personality test would appear to be served by developing subtle and obvious keys.

<div style="float:left">Method of Developing S–O Keys</div>

To develop subtle and obvious keys, we divided all items of the MMPI into two groups—those easy to detect for indicating emotional disturbance (obvious) and those relatively difficult to detect (subtle). Using several criteria, all of the items for each scale were sorted into these two categories. No attempt was made to equalize the number of items in each group, and more "O" than "S" items resulted. The keys thus developed were used to re-score the test sheets of a representative sampling of

100 cases of the original male norm group for the MMPI, and T-scores were developed for S and O.

Raw score tabulations for the subtle and obvious keys (hereafter referred to as S and O) indicated positive skews for most of the O-item distributions of the norm group. Relatively few individuals in the normal population answered the obvious items in a significant direction. The S items, on the other hand, were distributed in a relatively normal manner: For 139 normal males, on all five scales the S items were answered in a significant direction approximately twice or more as frequently as the O item. In addition, for 65 of the 110 items in the S keys for these five scales, the "significant" direction for scoring reversed the expectation of Hathaway and McKinley when they included these items in the MMPI, whereas only 8 out of the 146 O-items scored in a reverse direction from the original authors' expectation.

The attempt was originally made to develop S and O keys for all scales of the MMPI, but the results for Sc, Pt, Mf, and Hs, were almost uniformly negative. Hs consisted almost entirely of obvious items. Pt and Sc by definition also consisted of extremely deviate items which were therefore obvious, and Mf probably has too low validity as a scale to yield positive results.

| Correlations Among S&O Keys and With K |

Intercorrelations were computed among the S and O keys, including Hs which is really an O scale. In general, these intercorrelations showed the O keys highly correlated positively with each other, and uncorrelated with the S keys; while the S keys show low positive correlations with each other. As a rough index of relationship, the average intercorrelation among the O keys is +.60, the average correlation of the O with the S keys is –.15, and the average intercorrelation among the S keys +.21.

The uniformly high negative correlations that exist between O minus S scores and the K scale strongly suggest the appearance of a test-taking attitude in each of the five scales, and a close relationship between S responses and the K score.

| S&O Relationships With Other Factors |

A group with high scores on the "Lie" scale (9 items or more) was higher on the S keys of all five scales than on the O keys, and was also, on four of the scales, higher on the 3 keys than was the low "Lie" (0 and 1 items) scale group. For the group with low "Lie" scores, the O scores for all scales were approximately equal to or higher than the S scores.

Individuals of high ability (intelligence T-score above 60, and some college work) have approximately equal O and S scores, whereas individuals of low ability (T-score below 40, and less than 9th grade education) have generally higher O scores than S, and higher O scores than the high ability group.

The Multiphasics of a psychologically sophisticated group showed S much higher than O whether the group was giving "honest" results or was attempting to "fake good." With this group it appeared to make little difference whether the test was taken "honestly" or "faked good." In either case, O items were successfully avoided, whereas S items yielded average and above average T-scores. Very few O items were answered in a significant direction.

Generalizing from a very limited number of cases, there is a possibility that individuals without neuropsychiatric diagnoses but with high MMPI profiles (one or more T-scores above 70) are somewhat higher than on the S keys and somewhat lower on the O keys than a group with neuropsychiatric diagnoses and with high MMPI profiles.

S–O Scores of Successful & Unsuccessful Trainees

The S and O keys were applied to the MMPI answer sheets of successful and unsuccessful veterans who had taken school or on-the-job training. All the O keys in all cases showed a significantly higher score for the unsuccessful compared with the successful group, the S keys indicated insignificantly higher results for the successful group, and the total scales showed differences somewhere between these results of the O and S keys, insignificantly (with one exception) favoring emotional stability in the successful group.

In most counseling situations where work is done with a relatively normal population, it is probable that the present MMPI total scale scores will frequently fail to differentiate between the successful and unsuccessful because the total score represents a compromise between two fairly well differentiated aspects of each scale. Apparently elevation on the S keys tends to indicate, or at least does not contraindicate, success in school or on-the-job training. Conversely, elevation on the O keys apparently tends to predict failure. Results of total scale scores alone obscure this difference in function of S and O items.

The Control Factor in Social Adjustment

The phrase "in control of himself" is often used to designate the person able to direct his own activities, to adapt to present social demands, to plan for the future. "Out of control" is a term often used to describe the individual who seems at the mercy of immediate environmental stimuli.

In terms of S&O scores, previous studies had indicated that those individuals whose actual social adjustment, or potentialities for adjustment, seemed the least, tended to have obvious scores higher than the subtle. On the other hand, the more successful individuals tended to have subtle scores equal to or higher than their obvious scores.

The tentative hypothesis which was derived may be stated as follows: successful adjustment in society requires knowledge of socially acceptable ways of behavior, and the desire and ability to act in these ways. The socially acceptable way to behave on the personality test, as well as more overtly, seems generally to be to avoid deviate behavior. On the MMPI, the most deviate items are the obvious items, "deviate" because they are seldom answered in a significant direction by a normal population. The socially successful person may have the ability to recognize and to avoid making scores on personality test items which obviously indicate maladjustment, while the socially unsuccessful person may be unable to recognize or to heed signs of deviate behavior on a personality test.

Successful adjustment may show itself in many different configurations of personality test factors. A "control" factor may be postulated which affects the various scales of a personality test in different ways.

<div style="float:left; border:1px solid #ccc; padding:4px; text-align:center">Testing the
Control
Hypothesis</div>

With the very active and helpful cooperation of Dr. Hales, it was possible partially to test the hypothesis. Most veterans discharged from the service with neuropsychiatric diagnoses have now had several years in which to adjust to civilian society. The nature of their adjustment may be dichotomized, simply, by saying that one group is now hospitalized in mental institutions, while the other is not. If two such groups can be matched in background, an analysis of their differences on a personality test may throw some light on a test "control" factor which may improve the accuracy of prognosis of breakdown.

Two groups of veterans with diagnoses indicating schizophrenia were obtained. One group consisted of 100 cases in a single mental hospital, while the other group, of 52 cases, was composed of men not in the hospital at the time of case selection. Differences in test results are not maximum because of overlap in care selection; certain hospitalized cases were on the verge of being discharged, while some non-hospitalized cases had been and would be hospitalized. Education and age of the two groups were not significantly different.

<div style="float:left; border:1px solid #ccc; padding:4px; text-align:center">Control
Factor in NP
Hospital-
ization</div>

With the exceptions of K, Hy, Mf, and Ma, the mean T-scores for the hospitalized group are higher than for the non-hospitalized. The chief characteristics of the regular scale profile of hospitalized group are the high elevations in Pt and Sc—evidence of the validity of these scales. The profile for the non-hospitalized group shows no such outstanding elevations: the low mean profile, combined with the fact of non-hospitalization, suggests the possible invalidity of some of the present psychiatric diagnoses of schizophrenia. However, the diagnoses of schizophrenia have been subject in almost all cases to from two to four psychiatric examinations.

Results from the S and O keys tend to confirm the original hypothesis. The relatively successful groups (the non-hospitalized here) had S scores higher than O, while the unsuccessful (the hospitalized here) had O scores much higher than their S scores. On all five scales which have S and O keys, the non-hospitalized had higher subtle T-scores than the hospitalized. On the O keys, the hospitalized had significantly higher scores than the non-hospitalized on all scales except for Hy-O.

The sharpest differences between the hospitalized and non-hospitalized groups were on the O keys. If Pt and Sc, being almost entirely "obvious," are considered together with the O keys, the impression is strengthened that it is getting scores on obviously deviate items that differentiates the relatively "successful" from relatively "unsuccessful." Successful vocational trainees are even lower than the non-hospitalized schizophrenic group in O scores.

There may also be significance to the amount of gap existing between O and S T-scores. Whether this relationship is exclusively one of getting high O scores, or whether it is one of dynamic relationship between S and O, is a moot question. The "control" explanation which postulates a dynamic relationship is preferred here because the subtle items apparently do not contribute, in a "normal" population, to the validity of the total MMPI scale scores. That is, there appears to be a slight tendency in this study, as well as in previous ones, for successful groups actually to obtain somewhat higher scores on the subtle items, than the unsuccessful.

A simple subtle-obvious index was developed to summarize differences between the S and O T-scores on the MMPI of a single case. A plus one is given to an individual each time his O score is ten or more T-scores above his S score; and a minus one when the O T-score is equal to or less than S. Thus the possible range of scores for an individual MMPI profile is plus five to minus five (there are S-O keys for only five of the MMPI scales).

Using both the S–O index and an Sc score of 70 and above, 78% of the hospitalized cases ware selected, while only 40% of the non-hospitalized group were similarly selected. This combination of Sc with the S–O index both increased the number of hospitalized cases selected, and widened the difference between the hospitalized and non-hospitalized group in numbers of cases with significant signs.

Conclusions re Control Factor

The hypothesis presented and studied here has been that recognition and avoidance of behavior which is socially deviate, marking of test items which subtly indicate maladjustment, and being "adjusted" or "successful," tend to go together. Similarly,

sensitivity to or avoidance of unusual behavior, marking of test items which indicate obviously deviate behavior, and lack of success in society apparently tend to go together.

It is the tendency of subtle items to have negative discriminating power which suggests a dynamic relationship justifying use of the term "control." With the socially adjusted or successful, apparently the person tends to check test items subtly symptomatic of emotional disturbance and to avoid the obvious symptoms.