

Recaptured-Item Technique (RIT): A Method for Reducing Somewhat the Subjective Element in Factor Naming¹

PAUL E. MEEHL, DAVID T. LYKKEN, WILLIAM SCHOFIELD,
AND AUKE TELLEGEN

University of Minnesota

Patients or clients ($N = 791$) were rated by their psychotherapists ($n = 248$) on 329 phenotypic items from the Minnesota-Hartford Personality Assay (MHPA). The interitem correlation matrix was factor analyzed (Varimax rotation) and 40 factors emerged. The four authors, working first independently and then in conference, interpreted and named each factor by examining half the items showing high loadings on it. These 40 factor names were then presented to 10 skilled clinical judges in two batches of 20, together with sets of the other half of high-loading items per factor that had not been scrutinized in the factor-interpreting stage. The judges' task was to do a 20×20 matching of factor names with item sets. Success in "recapturing" items from factor names was almost perfect, indicating that the factor interpretation was communicating valid intersubjective knowledge. It is suggested that this recaptured-item technique (RIT) be used as one means of reducing the *ad hoc*, subjective character of factor interpretation; and that study of RIT rates in a variety of substantive domains may be helpful in evaluating the "psychological appropriateness" of competing analytical solutions to the rotation problem.

Two-thirds of a century after Spearman set forth the basic algebra for discerning a single general factor underlying a set of correlations (Spearman, 1904) and a full generation since Flanagan, in his doctoral dissertation (Flanagan, 1935), first applied Thurstone's methods to a structured personality test (the Bernreuter), it is a remarkable phenomenon, distressing but unblinkable, that hardly a single psychological "factor" (alleged in the research literature on personality to have been first identified by means of factor analysis) is even given passing mention, let alone made the focus of diagnostic interest, in *working*, patient-oriented clinical case conferences. As clinicians who are only too conscious of the frailties of clinicians, we do not propose to get undue methodological mileage out of that historical observation. But surely this social fact about the negligible impact of a powerful mathematical tool upon daily clinical practice is worthy of attention. The reasons for this strange state of affairs are several, and this is not the appropriate place to examine them (see Lykken, 1971). Practicing clinicians and factor-analytic personologists would not agree on the reasons, or at least would differ as to their relative importance. However, there is one reason which both clinical practitioners unfamiliar with the

¹ This research was supported in part by the Ford Foundation and by the National Institute of Mental Health (Grant M-4465). The senior author's work on developing the method and drafting this paper was carried on during a summer appointment to the Minnesota Center for Philosophy of Science, under subsidy from the Carnegie Corporation.

² Computer capacity limitations necessitated a multistage factorization of the 329 variables, which the editor and authors agreed would not be usefully detailed herein. Suffice it to say that the stepwise procedure was entirely "mechanical," not relying upon item content. Pending publication of the Manual reporting fully on construction and validation of MHPA, readers desiring a summary description of the factor-analytic procedure may write to Paul E.

mathematical underpinning of factor analysis, as well as research psychologists who rely upon this tool, would agree on. That reason is the fact that there is a large element of subjective judgment, intuition, and theoretical arbitrariness operative in the process of factor *interpretation*. Setting aside the well-known difficulties connected with the rotation problem—which the computerization of analytical solutions (e.g., Varimax) has of course left *conceptually* unsolved—it is generally admitted that two psychologists could have agreed reasonably well about the proper rotation, and hence could be dealing with a similar or identical set of factor loadings, and could still get into a lively debate about what the factor should be called. One does not need to have a bias, or an ignorance, concerning factor analysis as a statistical method to experience scientific discomfort in reading the typical book or article in which statistically identified factors are psychologically interpreted, whether these interpretations consist of a simple “christening” with a familiar name, the inventing of a neologism, or a brief paragraph of dynamic, structural, or etiological characterization of the factor. We have no wish to depreciate the methodological suggestions that have been made and, to a considerably lesser extent, actually followed, as means of reducing this subjectivity. Thus we know that students trained by Thurstone came to recognize that he, despite his great mathematical contributions, tended to look upon the computational procedures as deserving a much smaller amount of the investigator’s time and energy than thinking about the tests, introspecting and asking for others’ introspections about what mental processes went on in dealing with a certain kind of item, modifying the item content or format in various ways to see what would happen to the factor loadings, trying out a subset of the test variables in a new context (i.e., a different set of other variables) to see whether an alleged factor continues to “emerge clearly,” and the like. There is Cattell’s well-known emphasis (Cattell, 1946, 1950, 1957, 1959) upon being able to identify a postulated factor in different methodological contexts, e.g., ratings versus objective tests versus verbal self-report inventories, or contexts defined by whether the source of variation is individual differences (R-technique) versus temporal changes in a single organism (P-technique). Then we have Eysenck’s combination of factorial methods and empirical group discrimination in which a major consideration in solving the rotation problem is the theoretically “appropriate” behavior of a factor in relationship to a criterion classification of individuals (Eysenck, 1950; for a nice application to behavior genetics see Jones, 1971). It must be admitted, with all due respect to these important contributions, that they have not as yet, singly or collectively, dispelled the skepticism (not to say cynicism?) that many psychologists feel about the psychological interpretation of statistical factors. The present paper makes no grandiose claim to solve this problem. We simply offer an additional procedure for inclusion in the total armamentarium of methods by which the psychological meaning of factors may be sought.

The clinical practitioner or personality theorist who asks himself, while reading the report of a factor-analytic investigation in the personality domain, “Why do I have this uneasy feeling about the fellow’s interpretation?”, does not, we think, have to go very deep or introspect very long to find the answer. The plain answer is that factor naming is an after-the-fact, *ad hoc* enterprise, in which the interpreter can usually render fairly plausible why some of the tests should have moderate-to-high loadings on a factor christened “X,” has to work a little harder to explain some of the others, and then finds himself either expressing complete bafflement or indulging in one or another kind of special pleading by way of explaining away a couple of embarrassing factor loadings. He struggles with variables which don’t fit the interpretation and others which, in terms of the item content, fit it very well indeed, but unfortunately fail to be

loaded as high as other variables that do not appear clearly to do so.

It is strange and interesting that two kinds of “analysis” widely perceived as being at opposite ends of the continuum of scientific versus prescientific thinking, namely *factor* analysis and *psychoanalysis*, have this methodologically unfortunate property in common. A factor analyst can make it appear more or less plausible—depending upon the skepticism and substantive theoretical biases of his reader—that the proper interpretation of the second largest factor emerging from a correlation matrix is that it should be called “perceptual speed” or “ego-strength” or whatever. Similarly, the psychoanalyst can (again depending upon the predilections of his listener) make it more or less plausible that the manifest content of a dream, and the patient’s free associations to it, deal with the theme of “penis-envy” or “anal rage” or “positive transference.” The fact that Cattell relies upon matrix algebra whereas Freud did not, should not mislead us into supposing that the one method is *ipso facto* more scientific and objective than the other. Fundamentally, of course, psychoanalysis and factor analysis, like all other methods of investigating behavior structure, organization and etiology, rely upon the basic empirical fact of *covariation*, i.e., that “something goes with something else,” either in the sense of individual differences or in the sense of change over time. And in both instances, the theoretical task is to infer (postulate, construct) the “hidden variable,” whether that is a constellation in the historic past (e.g., schizophrenogenic mother), a structural feature of the society (e.g., social class), or a structural–dynamic entity within the individual (e.g., schizo-gene, death-wish) inferred to be causally responsible for generating this covariation of observable dispositions.

Basically, we are skeptical of a factor analysis as we are of a psychoanalytic interpretation, not because we think that the factorist or clinician is a crook, or because we doubt his ability to perceive the raw facts of the behavior emitted, but primarily because we are properly conscious of the slippery ingenuity of the human mind when it applies itself to “making sense out of” a state of affairs in which the rules for what is “making sense” are incapable of being rigidly drawn. Thus, we know that Freud’s analogy to the jigsaw puzzle is misleading because there are straightforward criteria for deciding when a jigsaw puzzle has been properly assembled and when it has not; whereas no such criteria exist for evaluating the material of a psychoanalytic hour (Meehl, 1970a). We should also admit that, despite the mathematics (partly arbitrary) which precedes the interpretative stage, no fully satisfactory criteria exist for evaluating factor interpretation either. Anyone who has engaged in research involving the psychological construal of factors, or who has tried to make up his mind whether the research of different investigators has identified the “same factor” in spite of a difference in factor names, is surely aware of the ubiquity and recalcitrance of this problem.

We shall not examine one all-too-easy resolution, that of adopting a superpositivistic line and saying that the problem is illusory because “a factor-analysis is nothing more or less than a convenient, economical expression of certain correlational facts.” This is not the place to discuss the philosophical aspects of that approach, so we shall content ourselves with observing that there is nothing very convenient or intellectually economical about expressing some correlation coefficients in terms of a set of numbers which enables the reader to get back (almost) to the correlation coefficients by carrying out the multiplication of a matrix by its transpose! Furthermore, so far as we are aware, there is no factor analyst who does not concern himself, one way or another, with the rotation problem; whereas given the above superpositivistic view of a factor as nothing but an arbitrary reference axis in a hyper-space, making no claim to what Allport (1937, pp. 287-289) would have called “biophysical trait reality,” there is little point in

even discussing the rotation problem. One solution is just as “good” as another, all of them being mathematically capable of generating the given correlational matrix.

However, a more sophisticated variant of this “operational” line is deserving of respect. It might be argued that psychological factors (at least in the personality domain on present knowledge) should be dealt with, so far as “identification” is concerned, rather like the way in which the nutritionist or biochemist dealt with vitamins in the early days of that field. Thus, for example, before the chemistry of ascorbic acid had been worked out, we only knew that there was some kind of substance, called noncommittally “vitamin C,” which was (to use Tolman-Hull language) “anchored” on the input side by listing those foods in which it was characteristically found, such as citrus fruits; and, on the output side, by listing some of the clinical phenomena likely to be produced by its deficiency. This noncommittal christening of a nutritional “factor” by the neutral letter “C” did not involve any naive positivism, but on the contrary left open for detailed investigations the *actual* chemistry of the situation. But in the meantime it was possible to talk about the mysterious substance, and the phrase “vitamin C” was defined “implicitly” or “contextually” by reference to intake foodstuffs on the one hand and characteristic vitamin deficiency symptoms on the other. There is a good deal to be said for this approach in the factor analysis of personality variables, and if someone argues that the easiest way to liquidate the subjectivity of factor naming is to avoid it, we shall not complain. The present paper is intended for readers who, while respecting this approach, are not prepared consistently to adopt it.

One of the main reasons, especially important to the working clinician, for making a factor interpretation (whether by applying a familiar name, inventing a neologism, or writing a short paragraph of psychological construal) is the pragmatic context of clinical utilization of tests or behavior ratings, where what Cronbach (1960, pp. 602-604) calls “band-width” is desired even at the expense of some loss in “fidelity.” To take the obvious example of intelligence testing, no practitioner is intrinsically interested in whether a patient can arrange colored blocks so as to duplicate a design presented to him by the psychometrist. The domain of behavior-dispositions which it is clinically important to infer from this artificial psychometric task is a domain which we can only characterize roughly in either common or technical language, and whose conceptual boundaries (i.e., the defining stimulus- and response-class properties of the dispositions appreciably saturated with the same factor or factors) are, in the nature of the case, fuzzy and fluctuating. A more detailed methodological consideration of this matter can be found in Cronbach and Meehl (1955; see also, on “open concepts” generally, Meehl 1959, 1971 [published 1972], and references cited therein). In the ideal state of Utopian psychometrics and behavior theory, one would be in a position to *derive*, in the strict literal sense of that term, the degree of dependency of a specified empirical variable, such as Kohs Block Design performance, on a given factor. The conceptual difficulties involved in carrying out this Utopian task are frightening, inasmuch as it would involve a complete specification of *all* relevant dimensions of the stimulus side and *all* correlated aspects of the response side, with due account taken of their configural relationships as well. Short of something like this, “derivation” as that term is used in theoretical physics, cannot take place. What we have instead, as is usual also in other fields of psychology than psychometrics (and, let it be remembered, in many branches of the biological sciences as well), is a somewhat looser chain of probable inference in which we move from a postulated state of affairs allegedly characteristic of certain theoretical entities to a plausible resultant at the “operational” level. (The cluster of related methodological problems denoted by the phrases “quasi-derivation,” “probabilistic deduction,” “explanation-sketch” exists in many

scientific fields, and these difficult issues are currently in dispute among logicians and philosophers of science. See, e.g., Hempel, 1965, especially Chapters 10 and 12 and references cited therein; Dietz, 1970; Scriven, 1959, 1962; Brodbeck, 1962.)

To the extent that characterizations of the theoretical entity on the one hand and the vaguely delimited domain of operational dispositions on the other rely upon the human investigator's own conceptual grasp mediated by nonmathematical language, we cannot at present reduce the process of "quasi-derivation" to operations performable by a computer or a clerk. After examining a set of factor loadings and introspecting about my mental operations in the tasks showing high loadings versus those showing low loadings, I may emerge saying, "Factor F seems to be a factor involving perception of relations in three-dimensional space, particularly if that apprehension would be facilitated by one's ability to imagine an object moving; perhaps both visual and kinesthetic elements are involved." It is surely a mistake to dismiss such a characterization of Factor F as "merely verbal" or "purely intuitive" on the grounds that it is expressed in words rather than functors. (Of course even in the exact sciences, contrary to what some psychologists assert, the meaning of a theoretical construct is not *exhausted* by the mathematical formalism. There is always an embedding text which contributes importantly to the meaning and, in fact, is typically required to legitimate some of the derivations.) Yet it must be admitted that this kind of factor characterization may be accepted by two psychologists who would not, as a result of their conceptual agreement, necessarily arrive at exactly the same inferences as to the arrangement of factor loadings for a set of new tests or test items. We make here the familiar point that whenever tight derivation chains are lacking, the scientist permits himself loose quasi-derivations. The kinds and sources of derivational "looseness," beyond the scope of this paper, are currently under study by logicians, statisticians, and historians of science. Without prejudice to the issue of a technical controversy outside our expertise, we may merely list as examples such "looseness" contributors as: mathematical approximation (e.g., dropping terms), *ceteris paribus* clauses, idealized physical entities, auxiliary hypotheses as problematic as the substantive theory of interest, partial analogies (one sense of "model"—not the rigorous one), statistical deduction when the prior probabilities are unknown (the Bayesian problem), and all derivation chains involving "fuzzy sets" (extremely open concepts). As mentioned above, these sources of derivational looseness are found in all sciences, even those we customarily designate "exact." (A psychologist who doubts this should have a look at any intermediate-level treatise on astronomy, physics, chemistry, geology, etc. Or skimming the *Encyclopedia Britannica* articles on special topics in the physical sciences, e.g., "valence," "acid," "thermodynamics," "quantum mechanics," "friction," "fluid mechanics," "gravity," suffices to make the point.) But in the social sciences, as Max Weber and others asserted before the turn of the century, we confront an additional source of derivational looseness, to wit, the *intentionality* of mental processes. Stones do not think or have purposes, but persons (and chimpanzees, and rats?) do. "Reasons for acting," "means-end relations," "valid argument," "cognitive structure," "rule-governed behavior," "goal-equivalence" are among the varieties of intentionality the psychologist has to deal with. As Cattell explained a quarter century ago, in the course of one of the most sophisticated methodological analyses ever written, the psychologist finds it convenient to invoke teleological causation in his discourse even though he may believe that the ultimate analysis will reduce it to efficient causation (Cattell, 1946, p. 111). Analyzing behavior at the molar level, one cannot, for example, make much sense of a "subsidiation sequence" (Cattell, 1946, pp. 111-115, 1950, pp. 155-157, 1957, pp. 505-512; Murray, 1938, pp. 86-88) without relying on concepts like *means-*

end appropriateness and *class-inclusion*. We cannot even discern that a certain behavior-disposition is correctly *subsumable* under a trait-rubric, or that a realized state of affairs would “count” as goal attainment, without implicit reliance on intentional relations. (See, in this connection, Hempel, 1965; Martin, 1969; Nagel, 1961; Popper, 1966; and reply by Meehl, 1970b.)

It is interesting that criticism of factor analysis on grounds of its being simplistic is almost always directed at the formalism (e.g., linear functions, negligible interaction terms, unsuitable metric, adoption of an arbitrary “parsimony” principle to resolve the rotation problem) rather than at the danger of substantive psychological oversimplification. It may be that psychologists have taken the “connections” of psychological *content* as being simpler, shorter, and easier to discern and formulate than they are. Clinical experience would suggest that the thematic relations between a factor and its phenotypic indicators should be expected to be subtle and complicated in many instances. It could be that the psychologist’s apprehension of the intensional relations in a “dynamic lattice” of overlapping, cross-connected subsidiation sequences (Cattell, 1957, p. 508) sometimes requires that he possess an abundance of what Spengler (1928 *passim*) calls “physiognomic flair,” at least a modest order of the cognitive virtuosity at theme tracing and analogy attributed to the master players of Castalia’s Glass Bead Game in Hesse’s wonderful novel (Hesse, 1969).

But the crucial role of the human thinker leads directly to an element of subjectivity, and hence to our skepticism concerning factor namings. So we have a dilemma: How do we accept the human mind’s currently unavoidable role in making quasi-derivations from factor interpretations to the pattern of factor loadings on indicator variables, but reduce as much as possible the *ad hoc* subjectivistic element arising from the fact that human ingenuity (plus a little semantic “trimming”) will usually permit something plausible to be said by way of interpretation, even if it has very little objective merit?

From a “Popperian” point of view, the distinguishing mark of a scientific theory, as contrasted with metaphysical and theological doctrines, is its “forward-looking” character, i.e., its ability to predict consequences other than those which entered into its formulation (Popper, 1959, 1962; Bunge, 1964; but see Feyerabend, 1965, 1970a,b; Lakatos, 1970; Mackie, 1969, especially pp. 30-32). Of course “predict” in this context does not intend to emphasize the date of a fact’s objective occurrence, but rather its epistemic relation to the theorist’s cognitive activity, i.e., whether it was available to him in making his theoretical formulation. Thus a theory about some historical event can be said to “predict,” in Popper’s sense, if it leads us to expect certain archeological or paleographic findings; as a matter of fact these may have already been “found” (or even be in the historical literature) and still count as Popperian predictions, so long as they did not enter into the formulation of the theory. One need not be a strict Popperian (none of us is) to recognize the importance of exerting methodological control over *ad hoc* theorizing, especially in fields where open concepts are so “open” as to be highly seductive. We want to allow the psychologist plenty of room for ingenuity, but we want also to control it (at some stage) by imposing a *predictive* task upon his work-product.

This way of minimizing the malignant influence of human ingenuity in concocting *ad hoc* concepts, tailor-made to fit a finite (and often rather restricted) set of facts, leads to a suggestion for the interpretative phase of factor analysis. Suppose a mathematically identified Factor F shows up with strong loadings on operational variables $x_1, x_2, x_3, \dots, x_m$. Our methodological skepticism about the factor interpretation arises from our realization that human ingenuity, plus a

little *ad hoc* corner-cutting and equivocating, will normally make it possible to concoct an after-the-fact characterization of what kind of psychological, social, or physiological entity (the three possibilities!) might be imagined that would exert a sizable influence on each of these indicator variables. Short of constructing some new measuring devices after the manner of Thurstone—the best predictive check when it “works”—is there any way we can introduce a Popperian “predictive” element into this situation? Obviously there is. Since it is not the world date of the event or measuring operation, but the epistemic state of the interpreter that is relevant, the straightforward procedure is to delete a portion of the information from that available to the interpreter and to see whether the resulting interpretation, based upon this portion only, will suffice to mediate prediction of the remainder. *Procedure:* We order the high-loading indicator variables x_1, x_2, \dots, x_m by the size of their factor loadings on the to-be-interpreted Factor F; we then delete, say, the second, fourth, sixth, etc., items in this rank order from the list of indicator variables; and we present our factor interpreter with the reduced set, consisting of the odd-numbered indicator variables $x_1, x_3, x_5, \dots, x_m$. He formulates his psychological interpretation of Factor F solely on the basis of this odd-numbered half of the indicator variables. The question then becomes whether this interpretation permits the even-numbered indicators to be “recaptured.”

A contaminative difficulty immediately arises in that the psychological interpreter who christens or construes the factor from the odd-numbered indicator set cannot erase this item content from his memory when he looks at the even-numbered set; and, therefore, strictly speaking, we have no way of knowing that the *interpretative* word, phrase, or paragraph is actually mediating a correct identification of the second set. (The same is true, although less obvious, for the more laborious procedure of concocting new indicator variables in reliance on the provisional interpretation, since we know both the factor name and the indicators-cum-loadings as we go about the new item-building job.) From this it follows that we must use at least two psychological interpreters, the first one having the task of christening the factor on the basis of half the items with heavy loadings, and the second having the task of matching the factor name (generated by the first judge) with the remaining set of indicators. The first psychologist sees the odd-numbered items only, and on this basis comes up with a factor name; the second psychologist does not see the items that elicited this factor name from the first judge, but only sees the factor name itself. His task is to match this name with the remaining set of items, those which were not seen by the first judge. So that even though the “derivation” of an operational disposition is only a quasi-derivation mediated by some mixture of sophisticated psychological know-how and common sense, not to say intuition, on the part of the human mind, nevertheless, we at least know that the theoretical conceptualization achieved by the first psychological judge was successful to the extent that his verbal expression of it, which obviously cannot have been *ad hoc* with respect to the half-set of items he did not see, somehow managed to convey the psychological essence of the postulated dimension to a second judge, whose matching can be checked against statistical reality.

Ideally, it would probably be preferable for the second judge (or set of judges, since more than one will usually be desirable) to evaluate the to-be-recaptured items singly with respect to each factor in a system. However, with a limited number of skilled clinical judges (there is no point in doing this with laymen) and a large number of factors and variables, the sheer logistics of this task make it unbearably tedious. It appears that, at least under some circumstances, one can arrive at reasonable confidence in the intersubjectivity or communicability of factor interpre-

tations by a less onerous approach, namely, allowing the second judge to have knowledge of which *sets* of indicator variables go together factorially. His task is not a rating task on single indicators but rather a matching task, i.e., to match up factor names with the sets of indicator variables that have moderate-to-high loadings on the same factor. Considering the logical and psychological processes of inducing the nature of a factor, and of eliminating all but one from a set of rather plausible construals, there is a good deal to be said for the matching technique, in spite of its well known disadvantages. We present the judge with a list of factor names, which include Factor F, and a batch of item sets, so that the psychology of the matching process tends to involve the second judge in a double process of “Would this be consistent?” together with “But would that be inconsistent?” Admittedly the matching procedure provides the second judge with information that we would perhaps rather withhold from him so as to make his task even more difficult. But the ungrouped, item-by-item procedure is, in a way, artificial and unnatural from the standpoint of scientific theorizing, where we do not simply “add up facts” but carry out mixed confirming-and-refuting thought operations upon the relation between data and hypotheses.

The proposed method may be illustrated by its application to a problem of interpreting factors in the rated behavior of psychiatric patients, a domain in which one has become accustomed to a distressingly large element of subjectivity and malcommunication between different investigators. For present purposes, where we wish merely to illustrate the proposed method in the hope of inducing others to examine its possibilities on various kinds of factorized data, only a bare minimum description of the empirical data collection will be necessary. The details of the entire study await publication elsewhere (Glueck, Meehl, Schofield, and Clyde, unpublished). As part of a long-term program on the skilled clinician’s assessment of personality (Meehl *et al.*, 1962; Glueck, Meehl, Schofield, & Clyde, 1964) a sample of patients ($N = 791$), ranging in severity of disturbance from hospitalized psychotics to essentially “normal” persons (seen in college counselling settings for educational and vocational counselling), were rated, following a minimum of 10 hr of interview contact, by their respective counselors or psychotherapists ($n = 248$) on a pool of 329 phenotypic items from the Minnesota-Hartford Personality Assay (Glueck *et al.*; Boelhouwer, Henry, & Glueck, 1968; Glueck & Stroebel, 1969; Melrose, Stroebel, & Glueck, 1970; Mirabile, Houck, & Glueck, 1971; Hedberg, Houck, & Glueck, 1971). The clinical material was drawn from a variety of settings all over the United States. The 248 clinical judges were paid for their time (more than their hourly psychotherapy rate), included professionals from psychology, psychiatry, and social work, and ran the gamut of theoretical orientations (e.g., Freudian, neo-Freudian, Horneyan, Adlerian, eclectic, Rogerian, rational-emotive, Sullivanian). It is worth mentioning that the overwhelming preponderance of these clinicians were neither familiar with, nor sympathetic to, the factor-analytic approach to personality, which makes it all the more interesting that such a “nonclinical” factor as Cattell’s *Surgency* emerges clearly from the analysis. (It is probable that 95% of the clinicians making these ratings never think about “surgency” in dealing with patients; our educated guess would be that the majority of them have never even heard of it.) The ratings were made on the basis of detailed instructions intended to minimize high-order inference. Each of the 329 indicator variables was rated on an 11-step scale, imposing a semi-forced distribution that allowed considerable freedom to the rater. With this item pool the characteristic shape of the distribution of unforced item placements is considerably more platykurtic than normal, and sometimes is not far from rectangular.

Factor analysis was carried out by the Varimax method and 40 factors were extracted.² Readers of the draft have expressed surprise at both the large number of interpretable factors and the size of the loadings. We like to think that (in addition to the huge N) our “good” results reflect (a) the care with which the item pool was constructed and screened, (b) the detailed rater instructions, (c) the clinical skill of our raters, (d) their motivation to do a creditable job for a professional fee, and (e) the unusual nature of the rater-patient contact. There is no point in applying mathematical power to qualitatively feeble data, but factor analysts have often done so with predictably depressing results. It should be pointed out that in the long history that went into the construction of the original phenotypic item pool, much attention was paid to diversifying the qualitative domain so as to avoid the stereotyped concentration of certain kinds of material in the tradition of clinical ratings, and face-valid subdomains had been employed at one stage in an effort to eliminate items which were factorially very mixed in composition (see Meehl *et al.*, 1962, pp. 1-2).

After rotation the strongest factor accounted for 8.3% of the variance and the 40th factor accounted for 0.7% of the variance. As might be expected, even with a relatively large but heterogeneous item pool, the small factors were represented by very few items. Although two of the present authors (PEM and WS) had been involved in the original item construction several years earlier, the factor analysis was conducted in another city (by Dr. Dean Clyde, one of the coinvestigators in the long-term program), and we successfully resisted the strong temptation to so much as glance at the results until after the completion of the present study. Neither of us had occasion to read through the entire batch of 329 items for a couple of years prior to this investigation. The items loaded 0.25 or higher (in one case only 0.23) on a given factor were arranged (by a clerk) in order of their factor loadings. Then the items in each list were split by her into a “naming” set and a “recapture” set, simply by assigning every alternate item in the loading-ordered list to the naming set and the next one to the recapture set. The naming-set items for each factor were then typed out on a separate sheet, being merely labeled as Factor 1, Factor 2, and so forth. Each of the four authors received 40 such sheets, each sheet listing the “naming” half of the items representing a factor with their factor loadings. He first attempted to choose one best single short factor name, whether technical or from common language; secondly, he could (and usually did) list several alternatives to this single best factor name; and thirdly—especially if he was dissatisfied with all of the short names—he tried to characterize the essential psychological nature of the factor in a short paragraph. After these 40 factor interpretations had been independently made by each of us we circulated each others’ characterizations and studied them at leisure. Then we met as a committee of four, debated at length, and hammered out a Quaker-unanimity characterization of each factor. It turned out, somewhat to our surprise, that we were able to reach agreement on a set of 40 short factor names, although of course our group confidence in some of them was greater than in others.

In designing the item-recapture stage (Phase II), it was felt that a 40×40 matching would present too much of a psychological hurdle to the judge. Furthermore, we were, for purposes of the main research project, more concerned to assess the objectivity of naming for the large

² Computer capacity limitations necessitated a multistage factorization of the 329 variables, which the editor and authors agreed would not be usefully detailed herein. Suffice it to say that the stepwise procedure was entirely “mechanical,” not relying upon item content. ~~Pending publication of the Manual reporting fully on construction and validation of MHPA, readers desiring a summary description of the factor analytic procedure may write to Paul E. Meehl, Psychiatry Research Unit, Box 392 Mayo, University of Minnesota, Minneapolis, Minn. 55455.~~

factors than for the small ones, and we did not wish the possible lack of intersubjectivity in interpretation for the small factors, represented by only one or two indicator variables each, to contribute to inaccuracy in matching the big ones. Therefore preparatory to Phase II the 40 factors were divided into two batches of 20 factors each, the task of an item-recapture judge being thereby reduced to two separate matching tasks, each of order 20×20 . Thus, what each Phase II judge had as materials in the first 20×20 matching was a “factor” page, listing the Phase I-derived *factor names* (previously randomized as to order, of course, since otherwise there might have been some unconscious contaminating influence, the big factors corresponding to larger item sets); and a batch of 20 “item-set” pages, each of which had on it the (Phase I-unused) set of items having significant loadings on a given factor. His task was to match the 20 factor names with the 20 item sets. He then carried out a similar 20×20 matching on the second batch of factors. We present a few examples by way of illustration.

As one might expect in research on psychiatric patients, with a clinically oriented emphasis in the item pool, one of the most important dimensions which emerges from the analysis (second strongest factor) is that known to clinical psychiatry as “thought-disorder” and in psychoanalysis as “ego-weakness.” Because of the variation in clinicians as to whether they prefer to use psychodynamic language or the less inferential language of descriptive psychiatry, and also because perceptual distortions are not literally *thought* disorders, this factor was christened “cognitive slippage” (Meehl, 1962, 1964) with the familiar synonyms being indicated parenthetically as “(thought-disorder, ego-weakness, reality-distortion) We present below the items (with their factor loadings) which were available in Phase I to the team of four factor namers in christening the factor; and below that we show the alternate items (and their loadings) which were presented to the clinical judges in Phase II for recapture by matching.

*First Factor: Items Available to Factor-Naming
Team (PEM, DTL, WS, AT)*

<u>Item content</u>	<u>Factor loading</u>
Manifests atypical, strange, or bizarre mentation.	.86
Dereistic thinking present; his mental activity lacks accordance with reality, logic, or experience; fantasy, rumination, and attempts at problem-solving or understanding his situation tend to be autistic.	.81
Has a tendency to experience perceptual distortions. (Extreme high rating means hallucinations.)	.79
Has feelings of depersonalization.	.78
His thought processes are confused.	.77
Manifests strange or deviant verbalizations.	.75
Is unsure of his own grasp of reality.	.75
Reports peculiar feelings whose exact quality seems hard to get across in words.	.69
Has inappropriate affect.	.64
Experiences body-image disturbances (e.g., it sometimes seems to him that all or part of his body changes size, or is distorted, or that a limb is some how	.61

“disconnected,” or that an external object is connected with his body, or has in some sense become momentarily fused with it or belongs to it).	
Experiences the feeling that people are concerned about him, observing him, and talking about him, without presenting adequate evidence for these notions. (Extreme high rating means delusion of reference.)	.56
His thought processes are slowed down. (Rate in terms of objective slowing down, apart from whether patient complains of this when it is absent or is unaware of it when present.)	.52
Has persecutory trends: experiences the feeling that others are unfair to him, critical of him, against him, or attempting to harm him, without presenting adequate evidence for such notions. (Extreme high rating means delusion of persecution.)	.48
Has retrospective distortion; in giving an account of past situations or experiences omits significant details, or relates events which did not in fact occur.	.44
Sensorium is clear; is oriented for time, person, and place.	-.49

Factor Name Chosen: “Cognitive slippage”
(thought-disorder, ego-weakness, reality-distortion)

Item Set to be matched by second group
of judges in Phase II

<u>Item content</u>	<u>Factor loading</u>
Experiences fleeting episodes of actual cognitive distortion. He can snap back to reality (or be snapped back by the therapist). Yet, during the episode, the pathological idea seems to have more the character of a belief than of an obsessive notion whose subjectivity the patient recognizes.	.82
Has short-lived psychotic-like states (micropsychoses) in which hypochondriacal ideas, ideas of reference, and feelings of depersonalization occur interlocked and in practically delusional form.	.80
Experiences at times a feeling of strangeness, unreality, or unfamiliarity with regard to ordinary objects of his physical environment; e.g., things seem somehow “not quite right,” “changed,” “at a distance,” “look different,” “don’t appear real.”	.78
The associative linkages in his discourse exhibit a defect in cognitive control.	.77
Reports occurrence of ideas which are objectively unreasonable and which it is difficult to distinguish clearly from delusional thinking, even though they are mere <i>ideas</i> and not, according to him, actual <i>beliefs</i> that he holds.	.75
Has a thinking disturbance which has resulted in inefficiency of his adaptation to life situations.	.75
Experiences blocking of his thought processes in which his ideation actually stops for a period.	.70
Reports periods of fantasy or preoccupation, lasting for several minutes or more, from which he emerges with a feeling of “coming to” and cannot recall clearly just what he was thinking about during the interval.	.65

Is frightened at times by consciously experienced inability to direct the course of his own thoughts.	.62
Has experienced the idea or feeling that his thoughts are somehow capable of bringing about objective consequences in the external world without intermediate action on his part.	.61
Has a conscious fear of insanity; expresses concern over the possibility that he may be losing his mind.	.55
Preoccupied; his attention to externals is weakened by his attention to inner ideation.	.50
Feels that others regard him as strange, odd, peculiar, "different."	.46
Vacillates between criticizing himself and criticizing others for his difficulties.	.29

A second example is the factor we labeled 'sociopathoid immaturity' represented by a much smaller number of phenotypic facets, as follows:

<u>Item content</u> (available to authors in Phase I)	<u>Factor loading</u>
Sophomoric; has a callow, immature, adolescent quality.	.51
Tells lies when there seems to be little or no point in doing so.	.44
Has a tendency when confronted with a reality problem to think in terms of unrealistic schemes, improbable windfalls, the getting of the "big idea," or similar unlikely and immature solutions.	.41
Self-indulgent; rarely denies himself much in the way of momentary ease, pleasure, and gratification of appetites.	.39

Factor Name Chosen: "Sociopathoid immaturity"

<u>Item content</u> (available to Phase II judges)	<u>Factor loading</u>
Boastful: talks about his own merits, connections, possessions, or attainments; speaks pridefully regarding himself or things pertaining to himself; bragging, daunting, vainglorious.	.46
Pursues momentary satisfaction; demands immediate reward; lives from day to day.	.44
Showoff: makes ostentatious display of his possessions or accomplishments; a pretentious attention-seeker.	.40

An example of a factor whose naming problem presented difficulties because of the different levels of inference at which the four Phase I interpreters were willing to operate (the reader can imagine the strength of the temptation to take a peek at the unread items during such debates!)

was the factor which finally ended up unanimously christened “repression,” although two of the four interpreters were still dragging their feet a little. Although this was one of the few factors in which so much as a single error was made in item recapture (by one judge, who interchanged it in his matching with a factor named “negative therapy attitude”) the designation “repression” seemed to work pretty well, since 9 of the 10 Phase II judges matched it correctly. The item content and loading are as follows:

<u>Item content</u> (available to authors in Phase I)	<u>Factor loading</u>
Has little conscious recollection of childhood events; has forgotten most of his early years.	.77
In therapy sessions he has difficulty finding things to talk about (other than his symptoms) which are significantly related to his psychological conflicts.	.67
Is consciously withholding material from the therapeutic sessions.	.58
Is resistant to the idea that his symptoms are related to or due to emotional maladjustments or psychological conflicts.	.51
Is psychologically oriented and sophisticated; evaluates the motivation of others in interpreting situations.	-.49
Describes his symptoms and experiences in detail.	-.62

Factor Name Chosen: “Repression”

<u>Item content</u> (available to Phase II judges)	<u>Factor loading</u>
Tends to be reticent and uncommunicative about his history.	.73
Considering his intelligence and verbal ability, shows a marked inability to produce a flow of material during the therapeutic sessions.	.64
Has difficulty in verbalizing cause and effect relationships in his own behavior.	.52
Communicates his feelings freely, with little emotional inhibition; inner feelings are not concealed but exposed to public appraisal.	-.46
Fluent: words come easily, the flow of speech is effortless.	-.51

A refreshing example of identifying a factor unexpectedly was the case of *Surgency*. This factor was first found by Cattell in 1932, and he considers it to be one of the best-established personality components in research by himself and others over the years since then; but one almost never hears it mentioned in a clinical case conference (or, for that matter, in the theoretical writings of clinical practitioners). As stated above, it is probably safe to assume that only a negligible minority, if any, of the 248 clinical rates participating in the raw data collection had any such factor in mind, either consciously or unconsciously, in recording their clinical judgments on the items of the phenotypic pool. Nor did the Ford project investigators have *Surgency* in mind during the early stages of the pool’s construction, none of us being a Cattell

disciple. Nevertheless it emerges as the strongest factor (first in order of percentage variance accounted for after the Varimax rotation), with loadings as follows:

<u>Item content</u> (available to authors in Phase I)	<u>Factor loading</u>
Jocular: tends to make jests and jokes, and in a cheery, jolly manner.	.82
Cheerful: in good spirits; light-hearted, ungrumbling; his mood and manner have a positive, undisturbed, enlivening, happy quality.	.79
Enthusiastic: tends to become intensely and pleurably excited about activities, pursuits, plans, ideas; readily develops strong, lively interests; can be seized and energized by involvement in something.	.77
Is a stimulating personality; tends to liven up the atmosphere; is a scintillating, "colorful" person.	.75
Spontaneously reports pleasure experiences; describes occurrences which "felt wonderful," "were lots of fun," "gave me a big kick," "provided real satisfaction," "enjoyed a lot."	.72
Smiles often.	.70
Meets people easily.	.69
Sees the humorous in himself; can poke fun at himself, view some of his behavior or circumstances with a degree of detached amusement; can laugh when he has made a fool of himself or looked comical from another's point of view.	.65
Enjoys and expresses the full range of emotional outlet in degrees appropriate to situations in which he finds himself.	.63
Does not take important matters seriously.	.63
Assured: his manner conveys the impression of confidence, self-acceptance, freedom from self-doubts, social timidity, or inferiority.	.62
Attempts to pass off frustrations and conflicts as a joke; tries to rob situations of their threat by "laughing it off."	.57
Has savoir faire: shows an immediate, smooth, and natural knowledge of how to act in interpersonal situations; tactful, socially experienced, poised, sophisticated.	.53
Is able to have emotional rapport and react empathically to the emotional states of others.	.51

The initial characterizations of this factor by the four authors (preparing independent memoranda prior to discussion) reflected considerable disagreement, intra-judge conflict, and puzzlement as to its nature. Possibilities mentioned included extraversion-introversion, Eysenck's dysthymic-hysteroid dimension, affective tone (elation-depression), and low-confidence suggestions by two of the naming team as to the possibility of its being a somewhat contaminated indicator of *anhedonia* (Rado, 1956, 1960, 1964; Meehl, 1962, 1964; Stein & Wise, 1971). When we conferred there was debate about the relative importance of the social-

extraversion aspect in contrast with the affective or mood aspect, e.g., perhaps mood should be conceived as the primary *real* (biophysical) interpretation, the “social impact” flavor (as perceived by the clinical interviewers who provided the raw ratings) being understood mainly as a sort of indirect outcome in personal interchange arising from the patient’s depressed or anhedonic condition. We were also bothered by a hysteroid element in the items reflecting use of humor as a defense. Only two of the four authors spontaneously mentioned Cattell’s term “surgency,” one of them having produced as his final prediscussion choice the hybrid factor name “surgent hedonia,” striving to have the best of both worlds! One of the four wrote the following: “Euphoria–dysphoria (or euphrasia–dysphrasia), perhaps one of Cattell’s terms fits. Low end is depression, not just sadness. Doubt that high end is mania—probably healthy surgency, etc., well-being, confidence, enthusiasm.” Heated discussion ensued. Of the two naming-team members who had not thought of Cattell’s Surgency as a possibility, one was soon converted by the arguments of the two who had, but the other was unfavorably disposed to this label. His opposition was partly on the grounds that this factor, let alone the received terminology for it in Cattell’s nomenclature, would be so unfamiliar to clinicians (especially psychiatrists) that its presence on the factor profile would not be clinically useful.

Without wishing to argue *ad verecundiam*, but mainly for the sake of getting an additional informed opinion “right from the horse’s mouth,” we wrote to Professor Cattell, presenting him with the half-list of alternate high-loading items of Phase I and an accompanying open-ended query, “Which, if any, of your factors would you identify this one with?” We gave him only the additional information that the raw data were behavior ratings by therapists, that the factor analysis was conducted with a Varimax rotation, and that Meehl, Lykken, Schofield, and Tellegen had some disagreement as to its proper interpretation. He promptly replied, “The set of variables you present here is very close indeed to that which best expresses the F Factor *Surgency*.... There is some hint of contamination with H Factor, *parmia*, ... as you anticipate, I am going to argue that this is due to that astigmatic Varimax rotation But really, it is a pretty beautiful specimen of an F Factor pattern.” This strong statement by the initial identifier and christener of the Surgency Factor led us to settle, although still with some misgivings, upon that as the most acceptable interpretation.

However, in the recapture phase, since we knew that hardly any of our 10 clinical judges were familiar with Cattell’s system, let alone in the habit of employing “surgency” in their clinical characterization of patient’s personalities; and since most of them would not have ready access during the matching task to all of Cattell’s books, we added to the designation “Surgency (Cattell),” the definition of this factor that appears in *English and English* (1958). This expansion of the mere factor name, while it seemed necessary given the fact that most of our Phase II judges would not have been able to give even a rough definition had they been presented with the mere word “surgency,” admittedly loads the dice in favor of a good item recapture. The definition in the *English and English Dictionary* is not really a *theoretical* specification of a hypothetical underlying dimension so much as a mentioning of the phenotypic facets that “define” [= contextually specify] it. Furthermore, because of the naming team’s concern about the anhedonia continuum, we supplemented the dictionary definition by the phrase, “with a positive hedonic tone.” Thus our characterization of the factor for purposes of the item-recapture task suffered from the kind of pseudotheoretical quality that is present in the early stages of vitamin research, when we have to say that “Vitamin C is found in lemons, oranges, tomatoes, etc.,” or on the output side, “The dietary absence of this substance gives rise to symptoms of

weakness, irritability, aching joints, acne, bleeding gums, poor healing of wounds, etc.” We found in this and other examples that the concrete application of the item-recapture method highlighted for us the dilemma of the factor analyst, by insistently posing the question, “How interpretative (theoretical) does one dare to be in factor naming, without losing a decent likelihood that what is communicated by the mere *name* will still be closely enough linked conceptually to the phenotypic facets in the to-be-recaptured list available to the Phase II judges?” The compromise between these countervailing considerations is an educational experience which we recommend heartily to the reader. The methodological bite of this issue can perhaps be conveyed by listing the item set to be recaptured from the Surgency definition by the Phase II judges:

*Second Factor: Item set to be recaptured
as representing “Surgency”*

<u>Item content</u>	<u>Factor loading</u>
Lively: his speech, gesture, and posture manifest a combination of quickness, fluidity, zest, and high energy output; manner and style are animated, spirited, vivacious.	.79
Genial: friendly in a warm, cheery manner; affable, amiable, cordial.	.78
Elated: has a happy mood, in which ego-inflation, success-feeling, expanded self-image are important components.	.76
Frivolous; given to trifling; manifests levity when not fully appropriate.	.75
Playful: tends to indulge in nonserious conversation, have “fun;” sportive, roguish, frolicsome, mischievous.	.72
Witty: quick or ready to express amusing congruities or incongruities; cleverly facetious.	.69
Seeks to have fun, enjoy himself, find a good time.	.65
His behavior is socially participant (contrasted with spectator or solitary behavior).	.64
Becomes enthusiastic, zestful, zealous about the activities involved in his work.	.63
His behavior has a general, hearty noisiness; the shout, the explosive laugh, the sharp “pistol-shot” cough are characteristic.	.63
Overtly affectionate; experiences emotional closeness which he expresses openly and directly by words, gestures, or physical contact.	.60
Characteristically recovers quickly from adverse experiences; bounces back well from threats, failures, disappointments, and emotional upsets.	.55
Energetic: puts a lot of energy and effort into his activities; works hard at what he does; industrious, vigorous, active, forceful, strenuous.	.52
Makes use of gesture, facial expression, and postural adjustments in communicating; employs “expressive movements” as an aid to conveying thoughts and feelings.	.50

These examples perhaps suffice to give the flavor of the problem and data.

Ten clinical psychologists, carefully chosen for brains, clinical experience, and probable cooperativeness on a gratis basis, were asked to serve as judges in the second (recapture) phase.

Although all of them have some kind of academic appointment in connection with the University of Minnesota Clinical Psychology Training Program, only four are full-time University faculty, the others being employed full-time in clinical installations and engaged largely in patient care. The four full-time University faculty are also currently engaged in clinical practice either as part of their appointment (e.g., Medical School), or in part-time private practice. Seven of the 10 had received the Ph.D. degree from the University of Minnesota in clinical psychology, and one each from Stanford University, the University of Iowa, and the University of North Carolina. Their theoretical orientations vary widely (e.g., Freudian, Sullivanian, Rogerian, social-learning, Skinnerian, “eclectic”). It is perhaps worth mentioning that none of them would consider himself a factor analyst, and none of them relies to any appreciable extent upon the psychological “factors” allegedly identified in the psychological research literature on personality. Subsequent conversation or explanatory letters spontaneously sent with their completed matchings indicate all of them being “research-oriented” clinicians.³

RESULTS

In Table 1 we present the number of correct matchings of factor names with the to-be-recaptured item sets for each of these 10 clinicians in Phase II of the study.

TABLE 1
Correct Matchings by Clinical Judges (Two
Batches of 20×20 Matchings of Factor Names
with Item Sets) in Phase II

Clinician	Hits, Batch I	Hits, Batch II
A	20	18
B	20	18
C	20	20
D	20	20
E	20	18
F	18	20
G	20	18
H	20	18
I	20	18
J	18	18

³ We express our great indebtedness to these 10 clinicians who gave freely of their time and energy and performed such a beautiful job: Dr. James N. Butcher, Dr. Norman Garnezy, and Dr. Robert Wirt (University of Minnesota Psychology Department); Dr. Harold Gilberstadt and Dr. Gail K. Lumry (Clinical Psychology Service, Veteran’s Administration Hospital, Minneapolis); Dr. Thomas Kiresuk and Dr. Zigfrids Stelmachers (Clinical Psychology Service, Hennepin County General Hospital); Dr. Sherman E. Nelson (private practice, Minneapolis Clinic of Psychiatry and Neurology); Dr. Alan Roberts and Dr. Lloyd K. Sines (Clinical Psychology Division, University of Minnesota Hospitals).

To carry out a formal significance test of these results would be a work of supererogation, and we were frankly somewhat nonplussed by the impressive performance of our 10 clinicians when presented with what we had thought might be a rather difficult task. Since two errors is the least possible short of perfect matching, the results are almost as close to perfection as one could get. It is perhaps worth noting that the reversals on identification in Batch II were in every instance based upon a confusion between the same pair of factors in that batch, each of which was represented only by a single item in the matching task. The naming committee had very little confidence in the labels we finally settled on for these two factors, and even after noting this systematic reversal by half of our judges, we were not prepared to cook up a plausible *ad hoc* “excuse” (nor a plausible reinterpretation after having looked at the remaining badly matched phenotypic items).

In the matching of Batch I, 8 of the 10 judges performed without error. One judge interchanged the factors named “repression” and “negative therapy attitude,” and the other judge interchanged the factors named “self-criticism” and “dominance-*submission*.” The remarkably high success in matching is consistent with the judges’ introspective reports that they found the task easier and faster than they had anticipated when first presented with the materials, and that any major conflict of decision was experienced with respect to a small minority of the elements in each batch.

Another way of looking at this information is in terms of the number of Phase II judges who correctly matched each factor with its to-be-recaptured item set. Table 2 shows the list of 40 factor names, the number of items in the to-be-recaptured set (either equal to, or one less than, the number of items representing the factor during the Phase I naming process) and the number of judges who correctly identified it.

TABLE 2
Factor Names, Number of Items to Be Recaptured and
Number of 10 Phase II Judges Successfully Recapturing

Factor name	Number of items in recapture set	Number of judges successfully recapturing
1. Cognitive slippage	14	All 10
2. Surgency	14	All 10
3. Manifest hostility	11	All 10
4. Altruism (nurturant dependability)	10	All 10
5. Accomplishment-motivation	8	All 10
6. Heterosexual drive	6	All 10
7. Dominance-submission	6	9
8. Interpersonal insecurity	6	All 10
9. Self-criticism	6	9
10. Negative therapy attitude	5	9
11. Repression	5	9
12. (Low) affective reactivity	5	All 10
13. Super-ego integration	5	All 10
14. Anxiety proneness	5	All 10

15.	Sociopathoid immaturity	3	All 10
16.	Heterosexual inhibition	3	All 10
17.	Somatic complaints	3	All 10
18.	Marital maladjustment	3	All 10
19.	Mother-valence	3	All 10
20.	Compulsive orderliness	2	All 10
21.	Decorous propriety (attitude, appearance, manner)	2	All 10
22.	Intellectual posture	2	All 10
23.	Therapist-parentification	2	All 10
24.	Social status orientation	2	All 10
25.	Vocational maladjustment	2	All 10
26.	Father-valence	2	All 10
27.	Religious-moral problem	2	All 10
28.	Power orientation	2	All 10
29.	Autonomous nonsuggestibility	1	All 10
30.	Depressed mood	1	3
31.	Interpersonal manipulation	1	All 10
32.	Interpersonal involvement-avoidance	1	All 10
33.	Mother-seeking	1	All 10
34.	Family-bound	1	All 10
35.	Alcohol consumption	1	All 10
36.	Dependency-anxiety	1	All 10
37.	Solitude-need	1	All 10
38.	Verbosity	1	All 10
39.	(Physical?) self-injury	1	3
40.	Effeminacy	1	All 10

Since the present paper is primarily methodological in intent, we shall not enter into substantive discussion about the merits of the factor names which emerged. (As clinicians we were pleased by some of the results, distressed—and puzzled—by others, e.g., how does it happen that such a clinically important factor as *depression* fares so badly?) Taking our findings as an illustration of the proposed technique for reducing subjectivity in factor interpretation, three cautionary comments should be made by way of guarding against undue optimism, to which one might be tempted by the unexpectedly “positive” results in item recapture. First, since the raw data generating the empirical correlation matrix are therapist ratings (rather than, say, objective test scores or literal behavior samples), some unknown but presumably nonnegligible source of the phenotypic covariations found must be attributed to rater stereotypes, conscious or unconscious. The factor loadings are from one point of view gratifyingly high, and the Varimax solution does not seem to have done any gross violence to the data, it being easy to find items that are heavily loaded on one factor and show only extremely small loadings on all the others. This outcome is partly an expected consequence of the thorough preliminary work which went into the construction and reduction of the Minnesota-Hartford Pool, briefly summarized in a preliminary communication (Meehl *et al.*, 1962, pp. 1-2). Nevertheless it is impossible, by a combination of item formulation, preliminary item analysis, and the most explicit instructional warnings, to eliminate completely the pervasive influence of rater stereotypes upon correlations.

Secondly, and in part flowing from the preceding, the method is here being applied to a situation having unusually high loadings and relative “factorial purity,” making the task easier for the naming team in Phase I and for the matching judges in Phase II. The amount of “contradictory” cognitive inputs to the judges was relatively slight, compared to what it would have been had the item clusters to be matched been characterized by greater item overlap. (As our colleague Professor Merrill Roff commented in looking at the results, “If everybody got loadings like that, there wouldn’t be such a difficult problem of factor interpretation in the first place.”) Finally, the material was such that many of the finally selected factor names have a minimal “theoretical content,” being rather too close—from the standpoint of ultimate theoretical interest—to the level of mere “summarizing abstractions” roughly characterizing the content of a set of rather obviously related behavior dispositions (*cf.* Cronbach & Meehl, 1955, pp. 292-293). In particular, recapturing an item whose content mentions alcohol from a factor name mentioning alcohol is a trivial exercise. We felt an exclusion of the smaller factors from the matching task would, however, be methodologically suspect and that two 20×20 matching tasks covering the whole range of factor contributions would be the least arbitrary joint protection against triviality and bias. We do not anticipate that application of RIT to other data will, in general, yield such impressive success in recapture. If that pessimistic expectation turns out to be correct, some rational nonarbitrary standard for the goodness of matching would need to be developed.

It may be desirable to work out a less “global” technique than the method of correct matchings, which, as is well known, has the disadvantage that it does not tell us the source of even systematic errors, nor enable us to distinguish between a near miss and a gross error. One possibility which preserves the essential method, that of testing whether a factor name communicates to a second psychologist so as to enable him to identify phenotypic manifestations without knowing those which gave rise to it in the “context of discovery” (Reichenbach, 1938, pp. 6-7), but which should yield more graded information about what is going on, would be to require that the individual items in Phase II be rated quantitatively with respect to the factors, without the items having already been grouped together for presentation to the Phase II judge. On the other hand, that procedure has the disadvantage of being so onerous that one would probably begin to suffer the disadvantages of inattention, boredom, and irritation on the part of the judges when working with such a sizeable set of items and factors as we were concerned with here; and, perhaps more serious, one might begin to lose, by such an “atomistic” approach, part of what is generally considered the distinct advantage of the method of correct matchings, to wit, that the judge can engage in mental trial-and-error, checking and rejecting competing hypotheses as to the nature of a factor, utilizing the information that one item (which doesn’t fit a possible interpretation) has almost as high a loading as do other items which appear to fit it.

With due allowance for these cautions and criticisms, the method appears to have sufficient merit to be worth investigating in a diversity of substantive domains. One interesting possibility might be a comparison of the RIT outcomes achieved by the several currently competing analytical solutions of the rotation problem, none of which can command universal assent as a mathematical implementation of any overarching psychological principle (see Lykken, 1971). We are not, of course, suggesting that RIT success should be a “criterion” for judging solutions. Since we reject simple structure, criterion analysis, etc., as *definitive criteria* of an optimal solution, we have no wish to lay down another indefensible touchstone rule of our own! But it is arguable that one consideration worthy of attention in evaluating a proposed analytical solution to the rotation problem is its long-run tendency, over a substantial number of studies and diverse

behavior domains, to facilitate the interpretative process in the direction of increased intersubjective communicability, this being one of the general demands that science makes upon any research procedure.

SUMMARY

Ratings by 248 counselors and psychotherapists of diverse theoretical persuasions on 791 counselees or psychiatric patients, ranging widely in kind and degree of psychopathology, were obtained on a quasi-descriptive (“phenotypic”) pool of 329 items belonging to the Minnesota-Hartford Personality Assay (MHPA). The intercorrelation matrix was factor analyzed (Varimax solution) and a set of 40 factors emerged. Arranging the items showing moderate to high loadings on a given factor, alternate items were removed from each factor’s item set, and the remaining half sets were examined by the authors, first individually and then (following mutual exchange of written memoranda proposing interpretations) in a group meeting which eventuated in the selection of a factor name for each of the 40 factors. These 40 factor names were then presented to a second group of clinical psychologists, together with the 40 half-sets of items, grouped by factors, which had been deleted before constructing the first item sets for factor naming. These 10 judges, working independently, were asked to match the factor names with the item sets in two batches of 20×20 matchings, to ascertain whether the item sets with moderate to high loadings could be “recaptured.” None of the judges committed more than one interchanging error in any of the 20×20 matchings, and in 11 of the matching operations all 20 matchings were correct. Thirty-four of the factors were correctly recaptured by all 10 judges. These findings are offered as indicating that something was being successfully distilled from the first set and intersubjectively communicated to the judges, permitting them to employ the factor interpretation so as to identify the items in the to-be-recaptured set. This recaptured-item technique (RIT) is suggested as a possibly useful procedure for helping to reduce the subjective element and the *ad hoc* tendency in the interpretative stage of factor analysis.

REFERENCES

- Allport, G.W. (1937). *Personality, a psychological interpretation*. New York: Holt.
- Boelhouwer, C., Henry, C. E., & Glueck, B.C., Jr. (1968). Positive spiking: A double-blind control study on its significance in behavior disorders, both diagnostically and therapeutically. *American Journal of Psychiatry*, 125, 473-481.
- Brodbeck, M. (1962). Explanation, prediction, and “imperfect” knowledge. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in philosophy of science, Vol. II* (pp. 231-272). Minneapolis: Univ. of Minnesota Press.
- Bunge, M. (Ed.) (1964). *The critical approach: Essays in honor of Karl R. Popper*. New York: Free Press.
- Cattell, R.B. (1946). *Description and measurement of personality*. Yonkers-on-Hudson, NY: World Book.
- Cattell, R.B. (1950). *Personality*. New York: McGraw-Hill.
- Cattell, R.B. (1957). *Personality and motivation structure and measurement*. Yonkers-on-Hudson, NY: World Book.
- Cattell, R.B. (1959). Personality theory growing from multivariate quantitative research. In S. Koch (Ed.), *Psychology: A study of a science, Vol. 3* (pp. 257-327). New York: McGraw-Hill.
- Cronbach, L.J. (1960). *Essentials of psychological testing*. (2nd ed.) New York: Harper & Row.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dietz, S.M. (1970). Discussion: A remark on Hempel’s replies to his critics. *Philosophy of Science*, 37, 614-617.

- English, H.B., & English, A.C. (1958). *A comprehensive dictionary of psychological and psychoanalytical terms*. New York: Longmans, Green.
- Eysenck, H.J. (1950). Criterion analysis—an application of the hypotheticodeduction method to factor analysis. *Psychological Review*, 57, 38-53.
- Feyerabend, P.K. (1965). Problems of empiricism, Part I. In R.G. Colodney (Ed.), *Beyond the edge of certainty* (pp. 145-260). New York: Prentice-Hall.
- Feyerabend, P.K. (1970a). Against method: Outline of an anarchistic theory of knowledge. In M. Radner & S. Winokur (Eds.), *Minnesota studies in philosophy of science, Vol. IV* (pp. 17-130). Minneapolis: Univ. of Minnesota Press.
- Feyerabend, P.K. (1970b). Problems of empiricism, Part II. In R.G. Colodney (Ed.), *The nature and function of scientific theories*. Pittsburgh, PA: Univ. of Pittsburgh Press
- Flanagan, J.C. (1935). *Factor analysis in the study of personality*. Stanford, CA: Stanford Univ. Press.
- Glueck, B.C., Meehl, P.E., Schofield, W., & Clyde, D.J. (1964). The quantitative assessment of personality. *Comprehensive Psychiatry*, 5, 1523.
- Glueck, B.C., Meehl, P.E., Schofield, W., & Clyde, D.J. *Minnesota-Hartford Personality Assay: Forty factors*. Hartford, CT: Institute of Living.
- Glueck, B.C., Meehl, P.E., Schofield, W., & Clyde, D.J. *Minnesota-Hartford Personality Assay: Doctor's sub-set*. Hartford, CT: Institute of Living.
- Glueck, B.C., Jr., & Stroebel, C.F. (1969 Suppl.). The computer and the clinical decision process: II. *American Journal of Psychiatry*, 125, 2-7.
- Hedberg, D.L., Houck, J.H., & Glueck, B.C., Jr. (1971). Tranylcypramine-trifluoperazine combination in the treatment of schizophrenia. *American Journal of Psychiatry*, 127, 1141-1146.
- Hempel, C.G. (1965). *Aspects of scientific explanation*. New York: Free Press.
- Hesse, H. (1969). *The glass bead game (Magister Ludi)*. (Trans. Richard and Clara Winston.) New York: Holt, Rinehart & Winston.
- Jones, M.B. (1971). Heritability as a criterion in the construction of psychological tests. *Psychological Bulletin*, 75, 92-96.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). Cambridge: Cambridge Univ. Press.
- Lykken, D.T. (1971). Multiple factor analysis and personality research. *Journal of Experimental Research in Personality*, 5, 161-170.
- Mackie, J.L. (1969). The relevance criterion of confirmation. *British Journal for the Philosophy of Science*, 20, 27-40.
- Martin, J.R. (1969). Another look at the doctrine of Verstehen. *British Journal for the Philosophy of Science*, 20, 53-67.
- Meehl, P.E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology*, 13, 102-128.
- Meehl, P.E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, 17, 827-838.
- Meehl, P.E. (1964). *Manual for use with checklist of schizotypic signs*. Minneapolis: Psychiatry Research Unit, Univ. of Minnesota Medical School.
- Meehl, P.E. (1970a). Some methodological reflections on the difficulties of psychoanalytic research. In M. Radner & S. Winokur (Eds.), *Minnesota studies in philosophy of science, Vol. IV* (pp. 403-416). Minneapolis: Univ. of Minnesota Press.
- Meehl, P.E. (1970b). Psychological determinism and human rationality: A psychologist's reactions to Professor Karl Popper's "Of clouds and clocks." In M. Radner & S. Winokur (Eds.), *Minnesota studies in philosophy of science, Vol. IV* (pp. 310-372). Minneapolis: Univ. of Minnesota Press.
- Meehl, P.E. (1972). Specific genetic etiology, psychodynamics, and therapeutic nihilism. *International Journal of Mental Health*, 1, 10-27. [in press at time of this publication]

- Meehl, P.E., Schofield, W., Glueck, B.C., Jr., Studdiford, W.B., Hastings, D.W., Hathaway, S.R., & Clyde, D.J. (1962). Minnesota-Ford pool of phenotypic personality items. (August 1962 ed.) Minneapolis: Univ. of Minnesota Press.
- Melrose, J.P., Stroebel, C.F., & Glueck, B.C., Jr. (1970). Diagnosis of psychopathology using stepwise multiple discriminant analysis. *Comprehensive Psychiatry*, 11, 43-50.
- Mirabile, C.S., Houck, J.H., & Glueck, B.C., Jr. (1971). Computer prediction of treatment success. *Comprehensive Psychiatry*, 12, 48-53.
- Murray, H.A. (1938). Explorations in personality. New York: Oxford Univ. Press.
- Nagel, E. (Ed.) (1961). Methodological problems of the social sciences. In *The structure of science* (pp. 447-502). New York: Harcourt, Brace & World.
- Popper, K.R. (1959). *The logic of scientific discovery*. New York: Basic Books.
- Popper, K.R. (1962). *Conjectures and refutations*. New York: Basic Books.
- Popper, K.R. (1966). *Of clouds and clocks: An approach to the problem of rationality and the freedom of man*. St. Louis: Washington University.
- Rado, S. (1956). *Psychoanalysis of behavior*. New York: Grune & Stratton.
- Rado, S. (1960). Theory and therapy: The theory of schizotypal organization and its application to the treatment of decompensated schizotypal behavior. In S.C. Scher & H.R. Davis, (Eds.), *The outpatient treatment of schizophrenia* (pp. 87-101). New York: Grune & Stratton.
- Rado, S. (1964). Hedonic self-regulation of the organism. In R. G. Heath (Ed.), *The role of pleasure in behavior* (pp. 257-264). New York: Harper & Row.
- Reichenbach, H. (1938). *Experience and prediction*. Chicago, IL: Univ. of Chicago Press.
- Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, 130, 477-482.
- Scriven, M. (1962). Explanation, predictions, and laws. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in philosophy of science, Vol. II* (pp. 170-230). Minneapolis: Univ. of Minnesota Press.
- Spearman, C. (1904) "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-292.
- Spengler, O. (1927). *The decline of the West*. (Transl. C.F. Atkinson.) New York: Knopf.
- Stein, L., & Wise, C.D. (1971). Possible etiology of schizophrenia: Progressive damage to the noradrenergic reward system by 6-hydroxydopamine. *Science*, 171, 1032-1036.