THE SCREE TEST FOR THE NUMBER OF FACTORS[1]

RAYMOND B. CATTELL
University of Illinois

Conception of the Problem

A brief, easily applicable test for determining the number of factors to extract in factor analytic experiments has long been in demand. Unfortunately, as pointed out elsewhere in a general theoretical examination of the number of factors problem (Cattell, 1966 b), a test does not exist—even a long or complex one—which is both mathematically precise and logically satisfying. The present paper does not touch the theoretical problem, except explicitly to state acceptance of the position there reached, but concerns itself with an empirical procedure for reaching decisions. That theoretical position, insofar as it can be briefly stated in familiar terms, is:

(1) That the "true" number of factors (considered either as the number of substantive determining influences in the physical world, or as these plus the number of factors from correlated error) must in general be greater than the number of variables used in the experiment. Consequently, the decision in choosing a point at which to cut off extraction must aim merely at encompassing what may be called the "NCV" (or non-trivial common variance, i.e., covariance). The decision as to what shall be considered trivial must depend on circumstances,[2] but in the sense of any fixed percentage of the total substantive variance it can be made only *after* rotation, when substantive and error common factors have been separated by rotation. Consequently the tentative decision *at* extraction is made only so as to permit final decision *after* rotation to be made most reliably. If "trivial" is defined in the sense of, say, 1% of substantive (non-error) factor variance, then it means 1% of the non-error portion of *that variance which would be extracted if we had the correct communality for the* $> n$ factors needing to be extracted. (This variance can be very closely approximated in most studies by taking the uniquely determined communalities for $n/2$ factors; though special cases will arise where $n/2$ does not cover the number of substantial substantive and error factors.)

(2) That common (experimental) error factor variance—notably the large error factors which may be larger than the smaller substantive factors—should also be considered part of the NCV

1. This investigation was supported by Public Health Service Research Grant No. 10274-01 from the National Institute of Mental Health. The writer wishes to acknowledge his indebtedness to Malcolm Coulter, John Horn, and John Nesselroade for helpful discussions on this problem.

2. By "circumstances" is meant: (1) whether the experiment is an exploratory one (hypothesis producing) or hypothesis testing; (2) the size of sample and measurement reliability of tests; (3) whether the factor correlations (through fixing hyperplanes) need to be exact, for a higher order analysis to follow; (4) whether one has chosen the tests and the sample explicitly to pursue the smallest factors in the domain, and so on.

needing to be extracted and rotated. (Note that until the discussion entitled "theoretical basis," below, we must use "real" and "substantive," on the one hand, and "error" or "transient" on the other, without pausing for the tighter definition they ultimately receive.)

(3) That it should be left to rotation to separate substantive and error of measurement factors, for as stated, what is trivial has its meaning finally in relation to the rotated substantive factors. In other words, the trivial rotated factors and the trivial unrotated factors are two different things.

(4) That rotation must be conducted with precautions to avoid the degenerative process of "factor fission."

(5) That when the non-trivial variance extracted is fixed at a percentage which encompasses most of the total (n/2 iterated) variance, and the error variance is then rotated into common error factors, the true factor hyperplanes will appear almost unblurred. (They would show a line—in a two space plot—*exactly* at zero only if we could hit the right communalities for the presumably $> n$ factors and rotate in that space.) That is to say, when the smaller error of observation factors have been properly included one can take perhaps $\pm$ .02 instead of $\pm$ .10 or $\pm$ .15 as "within the hyperplane", which will permit highly accurate rotation of factors.

If the above logic is sound a good deal of previous work (Burt, 1950; Coombs, 1941; Lawley, 1956; McNemar, 1941) attempts (by either rank of matrix or statistical tests) to answer a fictional question: "What is the correct (or most probable) number of factors?" For in reality there normally are more factors, either (a) "real," i.e., substantive, or (b) from correlated error, than there are variables. That is, the true number is $n+$, though most beyond $n/2$ will usually be extremely small. Factor influences beyond n can, of course, leave no impression on the rank of the correlation matrix, but they can leave evidence in the form of more than n oblique hyperplanes. (These can be recognized and separated, however, only in less than n space.)

The only theoretically correct question is: "How many of these factors is it worthwhile to take out?" and the corresponding practical decision we can call the WSF judgment (for "When shall we stop the factoring?"). The question has meaning only in relation to the finally rotated factors corresponding to scientific determiners, and our main problem is to translate the answer from a decision on the rotated substantives into a best WSF on the unrotated factors. This decision at the unrotated stage, where it *has* to be made, rests, however, not simply upon the anticipated percentage at which triviality will be fixed on the rotated substantive factors, but also upon three other conditions: (1) ensuring the successful recognition of error factors; (2) avoiding fission in rotations; and (3) the need for precision in seeking higher order factors through obtaining sharp hyperplanes. For example, the methods of Burt (1950), Lawley-Rao (Lawley, 1956;

Rao, 1955), and McNemar (1941) stop extraction at some arbitrary (though traditional) $p < .05$ or $p < .01$ level. This leads to rejection of considerable substantive factor variance which is needed in the subsequent rotations, even though it avoids inclusion of common error variance. Such sensitivity to a Type 1 error of statistical conclusion means considerable error in the rotated substantive factor patterns because the pieces of the jig-saw puzzle which should have been rotated into these patterns are simply not there. If, conversely, we shy off too sensitively from a Type 2 error we shall include a higher percentage of error variance. Our argument here and later is actually that inclusion of error variance is better, because we believe good rotation can handle the separation into distinct substantive and common error factors. At the same time our experience points to the existence of a new and real peril if one goes excessively in the direction of taking out trivial error and trivial substantive factors, namely, the risk of a degenerative solution (Cattell, 1966 b). Such factor fission in rotation is not mechanically avoidable and is something which only the most experienced rotators seem able to perceive and handle.

Thus our philosophy, in the situation of essential indeterminateness of the true number of factors—an indeterminateness analogous to that of the exact distance of the horizon from any fixed height in rolling country—is that we should aim at a definite percentage of the total substantive variance appropriate to our purpose and material,—say about 95 to 99% in most circumstances. The problem then becomes that of cutting off the unrotated extraction at a value likely to yield this in the rotated substantive factors—apart from the common experimental error factors. In this approach, incidentally, as with any other, we need to be alert to the phenomenon which has been pointed out elsewhere (Cattell, 1952; 1966 b) as "fluctuation of visibility."[3]

Test of the Scree Test on Plasmodes

If the final precision of conclusion on the number and nature of the real factors is relegated to the rotational process, our aim in the initial extraction is a WSF decision which ensures:

(1) That enough of the mixed real and error variance factors in the latent root extraction process are picked up to include all *real* (substantive) common factors to the desired NCV level. Most thorough investigators, except in rough exploratory studies, will

3. By this is meant that in any series of factors a through k, being extracted from a set of variables, the hyperplanes discovered will in one experiment be, at the tail end of the distribution, $k-2$, k, and $k+1$, i.e., missing $k-1$, and in another experiment, perhaps $k-3$, $k-1$, k, and $k+1$, i.e., missing $k-2$. Thus two or three experiments may be needed to locate the hyperplanes of all the smaller factors, and a fixed percentage of the total substantive variance will not always include *all* identical factors from research to research, i.e., not all non-matching factors are error factors. This fluctuation in the set of hyperplanes seen or chosen implies, as stated above, that there will normally be more hyperplanes than there are factor dimensions.

consider factors too trivial only when they cease to contribute say 1 or 2% to the common substantive variance.[4]

(2) That enough substantive variance is extracted (even at the cost of including common error variance, as stated above) to ensure that the hyperplanes eventually reached will not be blurred.[5] For exact determination of the nature of the primaries, and, especially, that precise determination of angles necessary for discovering the secondaries, requires that these hyperplanes be maximally sharp.

(3) That the number of non-trivial factors be not so greatly exceeded that the alertness required to avoid constant risk of factor fission in the subsequent rotation process, and to keep the trivial variance in the superfluous factors, becomes excessive.

The general theoretical position on the number of factors problem stated here is thus very different from that of the usual search for an exact mathematical solution or a boundary statistical position. It asserts that the former is chimerical and the latter beside the point (Cattell, 1958; 1966b). It states that a decision has to be made, defining some percentage value[6] appropriate to the stage

4. Parenthetically (see below) the Kaiser-Guttman test (of stopping beyond a latent root of unity), despite other arguments for its use, essentially is doing nothing other than apply such a test of triviality. For it implies that the investigator is not interested in a factor which contributes no more to the variance than does the average single variable in the study. Unfortunately, this is a shifting standard, rather than a fixed percentage as envisaged here, because with many variables it cuts off very far down and with few variables it stops factoring too soon. The discrepancies from the scree in all eleven researches in Figures 1, 2, and 3 are in this respect worthy of study. See also footnotes 7 and 8.

5. It will be evident, and is illustrated elsewhere (Cattell, 1958; 1966a) that if one takes out too little (or, in case of bad choice of communalities—or use of unities instead of communalities—too much) variance even the best rotational search for hyperplanes will end with the hyperplane points which should ideally be at zero being actually dispersed over a ± 10 or ± 15 or wider band. But in technically good factor analyses (i.e., where apt factor number and communalities are used and error covariance is rotated into error factors) razor sharp hyperplanes (virtually all intended 0's within say ±.02 of zero) can always be found. *For the principle should be noted that it is strictly only experimental error, and not sampling error* (provided the linear model holds), *that is responsible for a truly zero loading showing departure from zero.* Consequently, if the measurement error can be rotated into common error factors, and the specific error left in the non-common space, the variables in the hyperplanes of the real, substantive factors will lie *exactly* at zero, when rotated with thoroughness, *regardless of size of sample.*

6. A certain confusion persists in the concept of "percentage of the total common variance." As reported in the usual computer output it means, of course, the percentage contained in the given factor or factors, *of what the experimenter has chosen to use* as the total variance. This latter may be a wildly false estimate of the true total variance—and is nearly always an appreciable underestimate. Instead, we are speaking here of what percentage this total extracted variance is of the ideally estimated correct total variance. The basis for the latter which can usually be taken as very close, is to iterate for the unique communalities for n/2 factors. This we will call the "brief estimate." However, a closer approximation is to iterate for n factors. And since, admittedly the communalities cannot then be unique, the basis for this

and area of research, and dependent also on the post rotational distribution of natural factors in the given data. Only such global considerations can define what is to be called trivial variance. Extraction is then to be stopped at the point of trivial variance (which can be defined in percentage terms finally) but this still leaves the rotation process to decide what substantive factors are to be *finally* dropped as trivial (Cattell, 1966 a).

At this point, and in this framework of theory, a clue to a practical basis for decision has suggested itself to the writer from experience of a hundred or more factor analyses carried out over thirty years. For a certain uniformity has appeared which is not confined to any one (possibly idiosyncratic) area, but covers psychological, socio-cultural, physiological and physical data. As an exploratory step, before proceeding to the judgment on communalities and factor number in any research, it had been the custom of the laboratory to insert unities in the R matrix diagonal and plot the size of latent roots down to n or nearly n (the number of variables) factors, as in Figure 1, page 250 ff. below. As everyone knows such a plot falls first in a steep curve but then straightens out in a line which runs with only trivial and irregular deviations from straightness to the nth factor, as shown ideally in Figure 3 (b), page 257 ff. This straight end portion we began calling the *scree*—from the straight line of rubble and boulders which forms at the pitch of sliding stability at the foot of a mountain. The initial implication was that this scree represents a "rubbish" of small error factors.

Before long, in half a dozen studies, within that small fraction where ulterior evidence gave firm knowledge of the true number of substantive factors involved (and where this number happened to be definitely restricted), it was noticed that this scree *invariably* began at the kth latent root when k was the true number of factors, Accordingly this relation was put to a deliberate test by constructing plasmodes.[7] The results for two apt plasmodes under three conditions of one of them, are shown in Figure 1.

"longer estimate" (which is still not entirely adequate, since k, the true number of factors may exceed n) could rationally be a mean of some agreed number, say 6, such iterations carried out with n factors, beginning from a variety of initial rough estimation positions. The best means of making a communality estimate for n factors deserves intensive inquiry.

7. *Plasmode* (from plasma=form, and mode=measure) is a term designed to avoid confusion with model (e.g., a mathematical or mechanical model). A plasmode is meant to refer to an exemplification of a general model in a particular set of numerical values derived from a concrete situation. Its function is that it can be used to test different methods of computation and to bring out various properties of the model not easily seen or deduced on abstract grounds. In the case of the factor model the plasmodes are sets of variables made to correlate as required by a definitely known set of underlying factors, with the standard assumptions of linearity, additiveness, etc., in the scientific factor model. Thus, possessing sure knowledge of the "back stage" structure, one can experimentally evaluate the precision of various proposed extraction methods, statistical tests, etc., as, in this case, the correctness of any test of the number of non-trivial factors. The best kind of plasmode is not one resting on numbers generated by dice or random number tables,

In the use of the scree test the issue will arise whether the last non-trivial factor is that *immediately beyond* or *at the end* of the straight scree line. Since there are four factors for certain in the ball problem and 10 in "Plasmode 30" the first point beyond the scree might seem indicated by these immediate examples. But there is now evidence that the ball problem contains a small fifth factor —the extent to which the weight is distributed to the circumference (determining angular momentum variables). Further, as we shall see in a moment, the sample correlation of specifics, even in a completely errorless example like Plasmode 30 or a practically errorless example like the Ball Plasmode, produces at least one com-

FIGURE 1
Scree Decisions on Plasmodes with Definitely Known Number of Real Factors

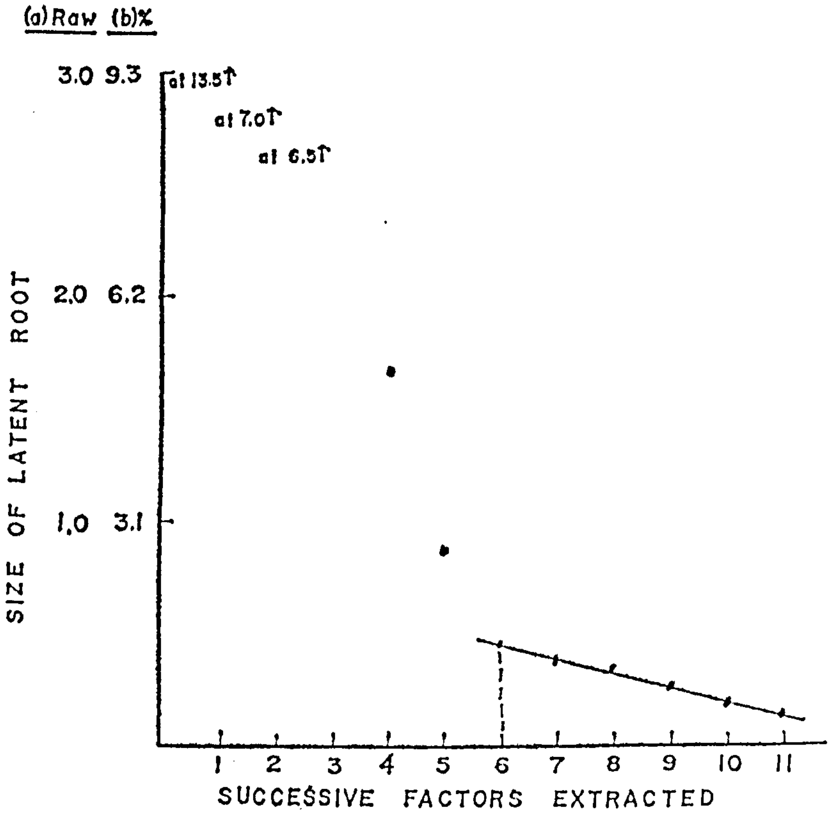

Fig. 1 (a) The Cattell-Dickman Ball Plasmode: 4 factors, plus error of measurement. Sample 80.

but taken from a physical model, as in the ball problem (Cattell and Dickman, 1962), the box problem (with actual boxes) (Thurstone, 1947), or the coffee cups problem (Cattell and Sullivan, 1962). For in these the measurement error, sampling, correlations, etc., occur in a natural manner related to the procedures of actual psychological experiments and the naturally occurring interactions.

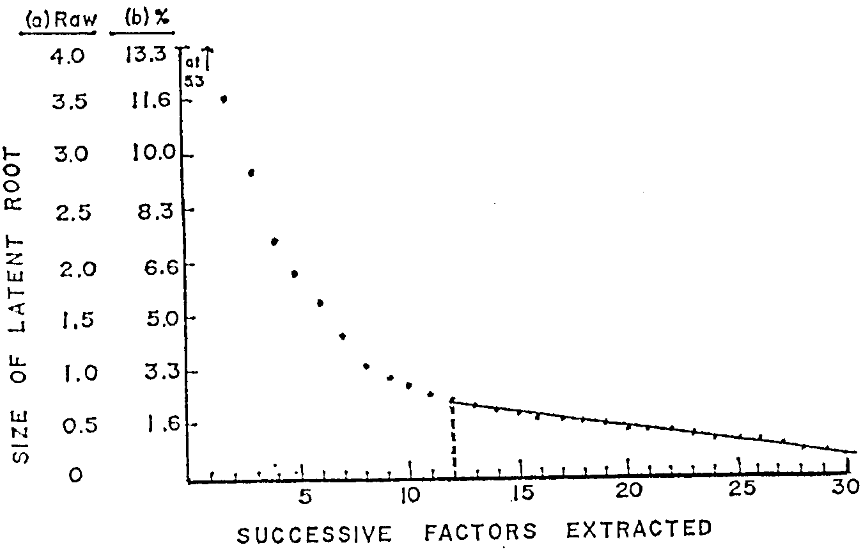

Fig. 1 (b). The 10 Factor Plasmode: without error of measurement. Sample 300.

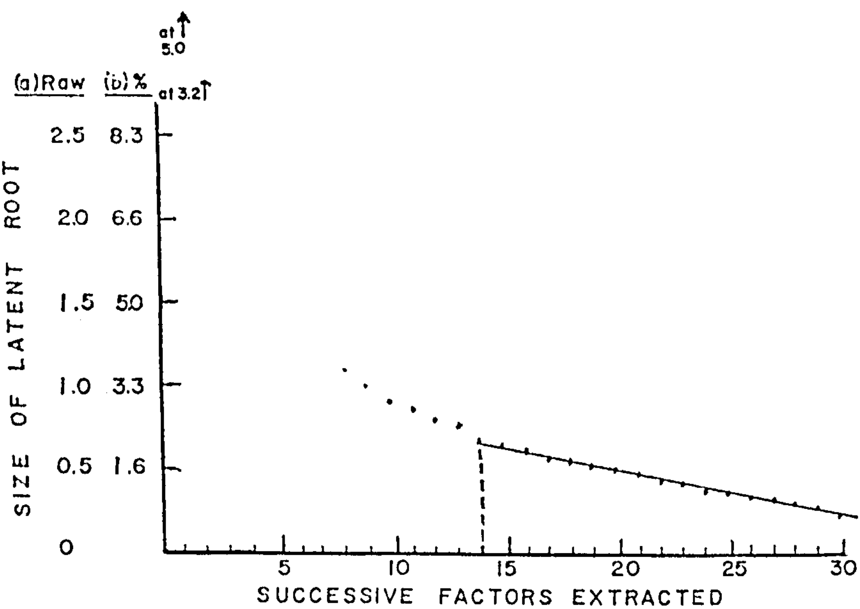

Fig. 1 (c). The 10 Factor Plasmode: with error of measurement. Sample 300.

mon (error) factor beyond the truly substantive ones, so that cut-offs at 11 in Plasmode 30 and 6 in the Ball Plasmode are correct. These are the values given by the highest point *on* each scree.

As we proceed in Figures 1(c) and 1(d) to examples with error, and in which the largest common error factor would be ex-

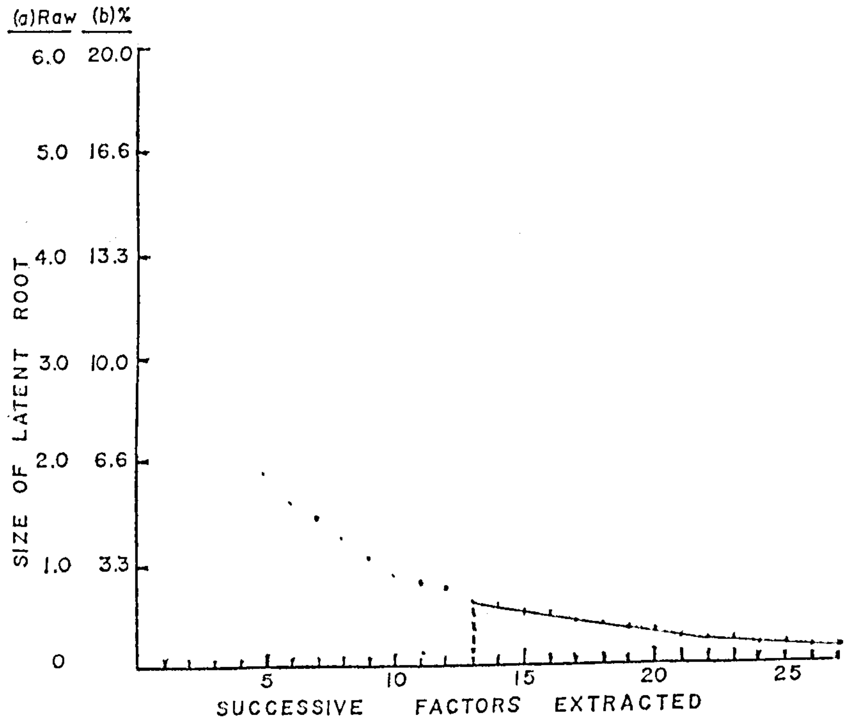

Fig. 1(d). The 10 Factor Plasmode: with error of measurement. Sample 50.

pected to exceed the smallest substantive factor, we shall see that Plasmode 30 now requires 1 or 2 more factors. Thus provided we agree to aim at inclusion of the two or even three largest specifics-plus-error common factors, in order to be sure of containing the substantives down to, say, a conception of non-trivial fixed at 95%, the best rule is evidently to *cut at the uppermost point actually on the scree.* Indeed, we argue below that it is always desirable to include at least one common error factor as a "garbage can." Moreover, as Thurstone argued (though it has more force for the centroid than the principal axis), it is always safer to take out one too many rather than the converse, since rotation will reduce it to triviality if it is in excess.

In Figure 1 we have contrasted errorless with error-encumbered data, presenting Plasmode 30 handled with errorless data and on a large sample in 1 (b), with error on a large sample in 1 (c), and on a smaller sample with error in 1 (d). Experimental factorings were thus carried out at 1 (c) and 1 (d) in fact to show the effect of (1) introducing error of measurement, and (2) severely reducing the size of sample. Under these conditions the scree test, as it theoretically should, shows increase of factors, from 12 to 13 and 14 respectively. That the larger sample, 1 (c), should show an increase of perhaps two error factors is to be expected, through the appearance of error common factors. In concrete data it would also tend to show more factors than the small sample, 1(d), because the inclusion of more varied people would bring in some dimensions accidentally absent in a small sample. Normally—but by chance this appears not to have happened here—the smaller sample with error would be expected to generate some larger common error factors, for then random r's get larger so that the *largest* non-trivial error common factor would be expected to be larger.

Evaluation of the Scree Test on Substantive Experimental Examples

No psychologist is likely to be entirely satisfied unless the scree test is evaluated on concrete plasmodes—and, indeed, on concrete plasmodes from our own psychological field of data—over and above the abstract or mechanical plasmodes just illustrated. Unfortunately, in seeking to explore this aptness directly to psychological data one is in danger of slipping a cog in logic. For a soothing tradition of circular argument has grown up in which some original research has argued for a particular number of factors in some hoary example, and psychologists now use the decision on this example as the test of any new number-of-factors test! It is easy to drift to the dangerous conclusion that we really know the numbers of factors in, say, Air Force data re-examined by Fruchter (1966), or in the 8 physical or 24 psychological variable examples (pages 82 and 137) in Harman (1960). (Parenthetically one is reminded of another concession in our Zeitgeist to "truth by apostolic succession" in the common assumption that a new intelligence test can be evaluated against the Terman Binet or Wechsler, instead of a well rotated g_c or g_f general factor (Cattell, 1963)!)

While one must explicitly reject dependence on traditional "hearsay," embodied in any conception of apostolic succession of truth one need not rule out genuine, non-direct, circumstantial evidence as to number of factors in a real data experiment. However, one must recognize that mostly this ends only in probabilities and not, as with plasmodes, in certainties. This pragmatic and circumstantial evidence in the examples now to be viewed in Figure 2 rests on (a) research with the same variables with different groups having led to the same number of factors by a consensus of factor number tests, (b) variables having been chosen from years of research on structure, and supporting physiological, learning, etc., evidence, to yield just certain factors, and (c) other circumstantial and well replicated information too various to summarize.

For example, Humphreys' data (Figure 3(f) is used because it deals with a well explored ability area, and has been around for years for examination by psychometrists. His own analysis is on record (Humphreys, 1964) and it has the advantage of so large a sample that we can suppose the non-trivial factors to be solely those designed to appear in the test battery. Examples 2(b) and 2(d) represent foreign language translations (Cattell, and

Nesselroade, 1965; Tsujioka and Cattell, 1965) of the 16 P.F. in which the item equivalence has been very carefully preserved. Since two clear markers only were used for each factor, and since the original 16 P.F. has been re-factored now at least seven times, with regular reappearance of the 16 factors in good simple structure formation, one should be able to count with reasonable certainty on the translations also containing 16 factors, as the screes here indicate.

FIGURE 2
Scree Decisions on Psychological Data with Reasonably Certain Number of Factors

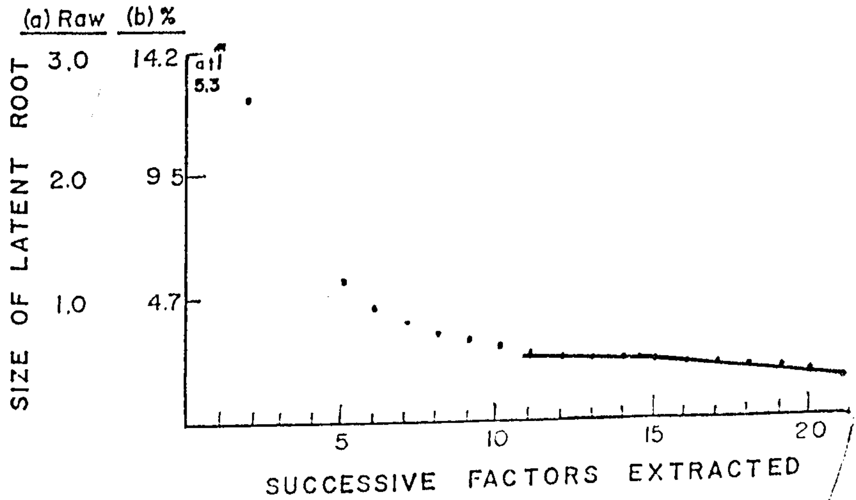

Fig. 2 (a). Humphreys' Air Force Data: 11 factors. Sample 8,158.

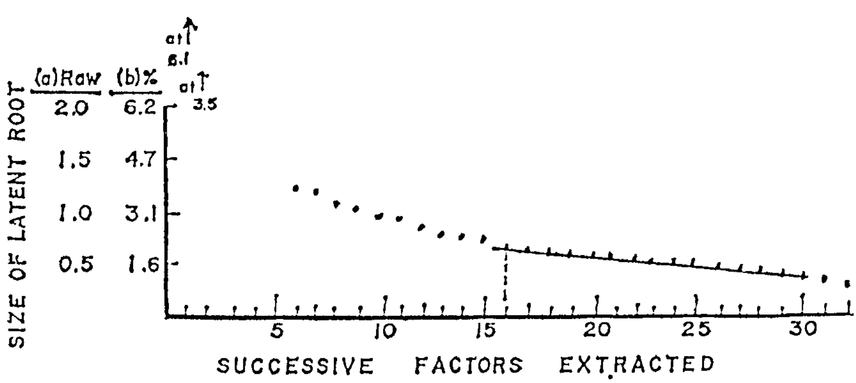

Fig. 2 (b). Japanese translation of the 16 P.F. Sample 300.

The third example, 2c, is Meredith's (1966) second-order factoring of the HSPQ, taken because (1) it also has so large a sam-

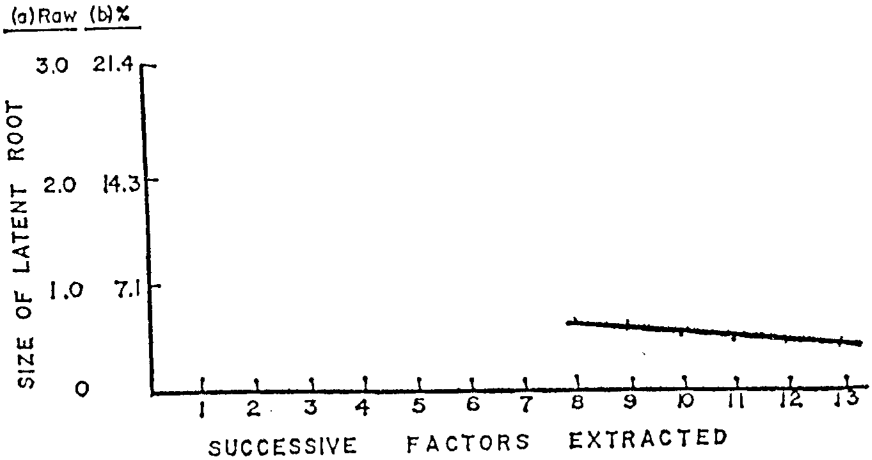

Fig. 2 (c). Data on Second Order HSPQ Test, 8 factors (Only 7 ultimately extracted to keep within $\frac{n}{2}$). Sample 1127.

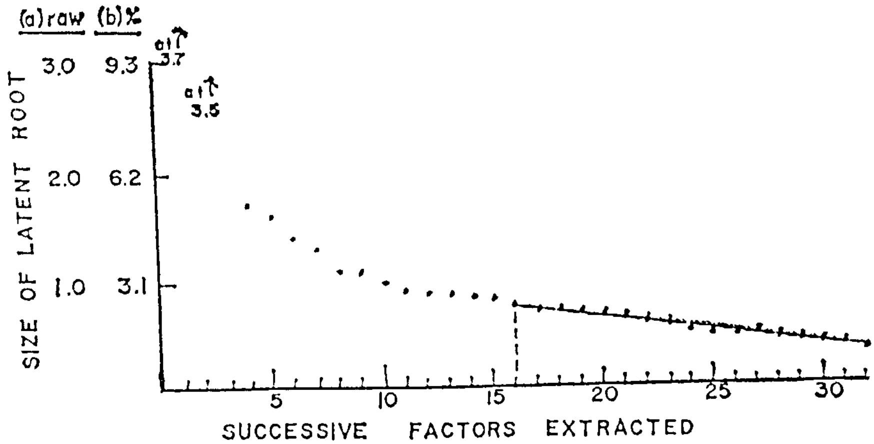

Fig. 2 (d) German translation of the 16 P.F. Sample 446.

ple that we can anticipate finding little or nothing beyond the substantive factors—anxiety, exvia, cortertia, etc.,—which have been repeatedly found in the questionnaire domain, (2) the existence of this particular structure is vouched for by analyses on quite other instruments, too, e.g., the 16PF (Gorsuch and Cattell, 1966), in which the same meaningful second-order factors appear (though the 16 P.F. has two more). This example has the interest of posing an additional problem, successfully met, namely, that two of the second order factors are virtually represented by single first order factors, having only very small loadings on other primaries.

In every one of these examples, as Figure 2 shows, the scree test decided on the correct number of factors (though indicating no error factor large enough for inclusion) on the basis of independent conclusions. In numerous other instances in our experience, excluded by space from illustration here, it has given what other evidence indicated as correct, and in none has it given a demonstrably false answer. Doubtless results of a still wider experimental try-out, hopefully always on cases with a firm ulterior basis for decision, will now appear from the research of others.

Some Issues in Practical Use

For the reader prepared to accept its working at the level of an empirical law, and who may not want to go on to the speculative theoretical basis derived in the next section, it would be appropriate here immediately to conclude with advice and observations on its practical use.

Clear plots, as in most occurring in Figures 1 and 2, present no problem, but cases sometimes occur where the person inexperienced in the technique hesitates, and it must be admitted that even a test as simple as this requires the acquisition of *some* art in administering it. First, one has to be able to distinguish between the slight irregularities and even slight wave forms which occasionally appear in an essentially straight line, on the one hand, and the relatively definite upward inflection where the straight scree line joins the major curve, on the other. We have found it useful to sight the line with the edge of a transparent ruler and to draw with a more expanded vertical scale than is possible if one includes the whole range, from the largest root, in a single drawing. This is the reason why plots in Figures 2 and 3 are truncated, omitting the upper, initial, root values.

Secondly, one has to recognize a phenomenon not hitherto mentioned which occurs with appreciable frequency, namely, the split of the scree line into two distinct straight slopes as shown in the 3(a) and 3(c) ideally, and 1(d), 2(a) and 3(e) plots, concretely. In all such "double screes" which we have seen, the empirical rule has been simply to take the higher scree and ignore the lower (a rational for this follows in the next section). In any case, the numbers of factors indicated by the tops of the upper and of the lower scree (occasionally even three steps occur) are of so radically different an order that one would generally know on ulterior evidence which lies in the range of possibility.

It has been suggested above that choice of the number of non-trivial factors might be guided by some definition of non-trivial in an exact percentage, held consistent from research to research. However, as pointed out, this percentage is intended to be relative to either the total common factor substantive variance or the total common substantive plus common error variance. The percentage calculated not under the scree in the principal components solution, with unities in the diagonal, will have only a remote relation to this true desired value, and will be numerically much lower.

FIGURE 3
The Relation of Double Scree to Sample Size, Reliability, etc.

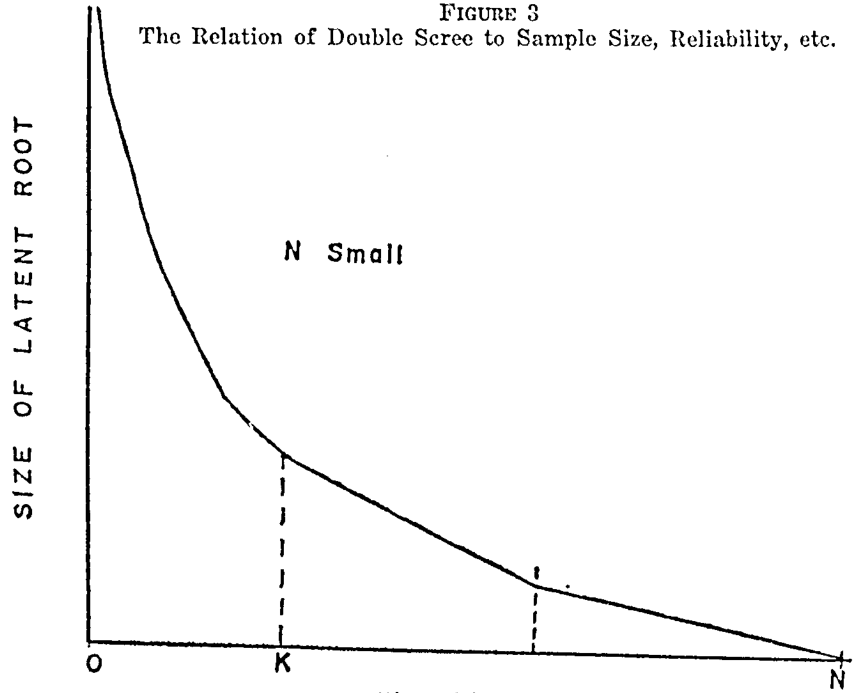

Fig. 3(a).

SUCCESSIVE PRINCIPAL AXIS FACTORS

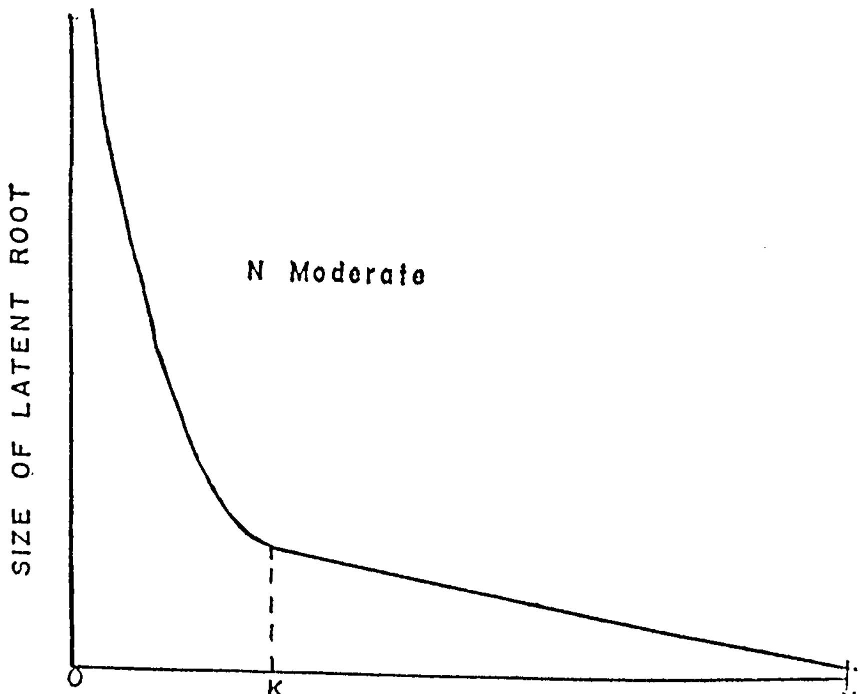

Fig. 3(b)

SUCCESSIVE PRINCIPLE AXIS FACTORS

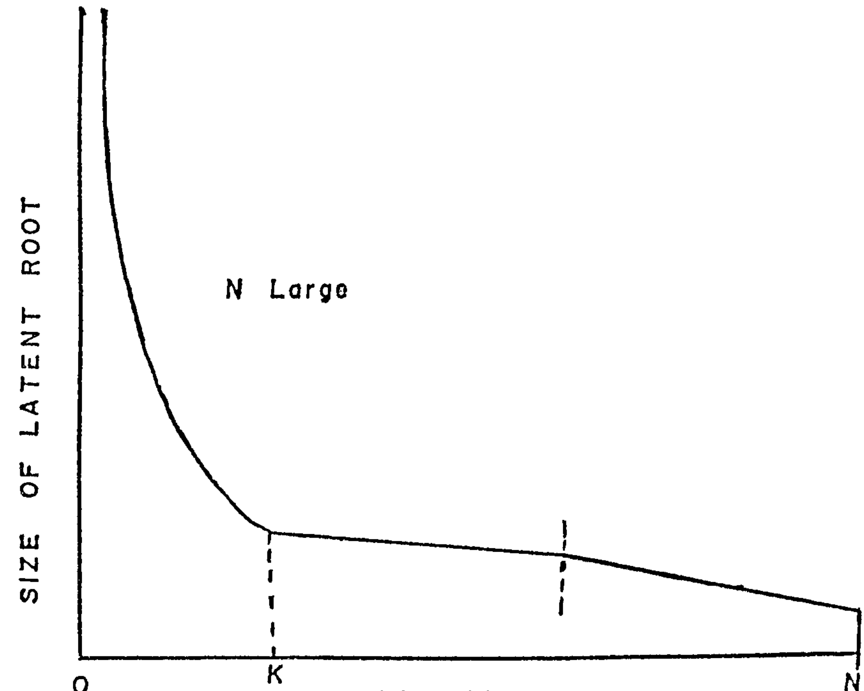

Fig. 3 (c)

SUCCESSIVE PRINCIPAL AXIS FACTORS

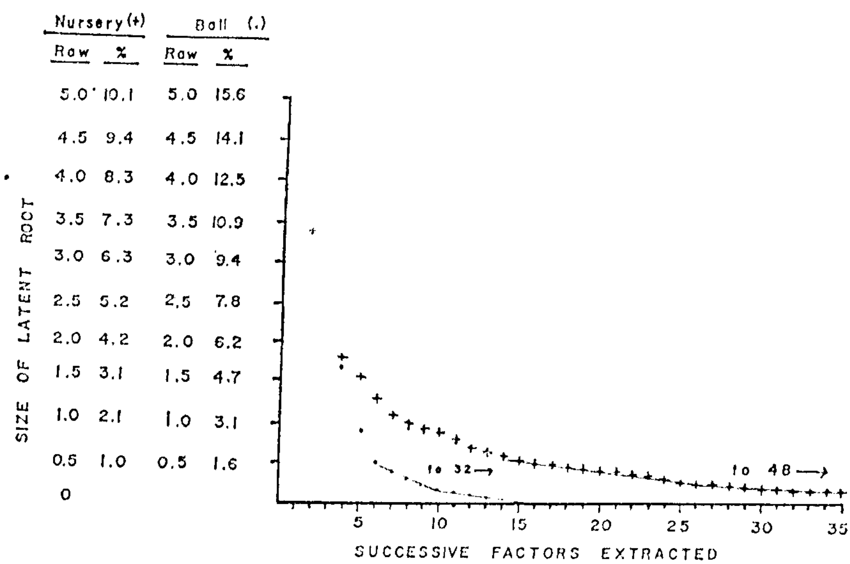

Fig. 3 (d). .... The Ball Plasmode Extended to 32 Roots. ++++ Ratings of 93 nursery 4-5-year-old children by Teachers on 48 variables chosen to represent 14 factors.

However, for the sake of information we have worked out this value for examples 1 (a) (Ball), 2 (a) (Humphreys), and 2 (b) (Japanese 16 P.F.) above and it lies at 94.5, 82.2, and 77.5 respectively.

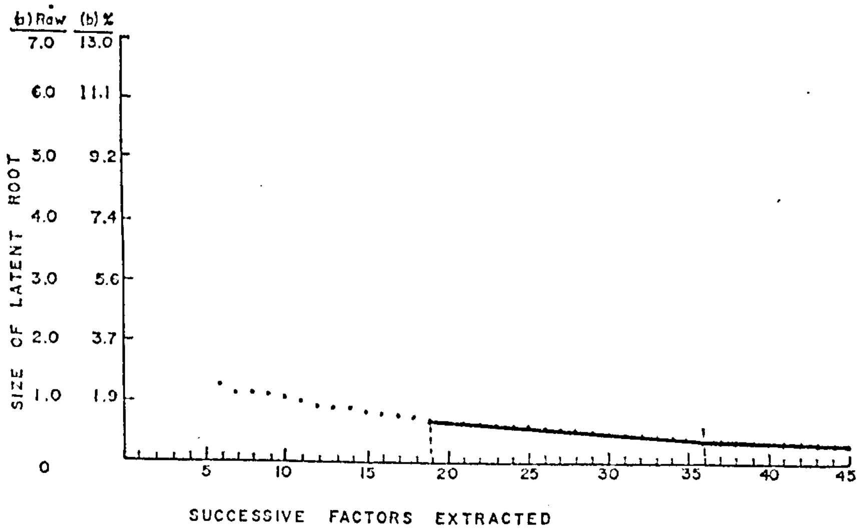

Fig. 3(e). Questionnaires (ESPQ & CPQ of previously resolved structure) on 289 8-year-olds, with 54 variables. Sample 289.

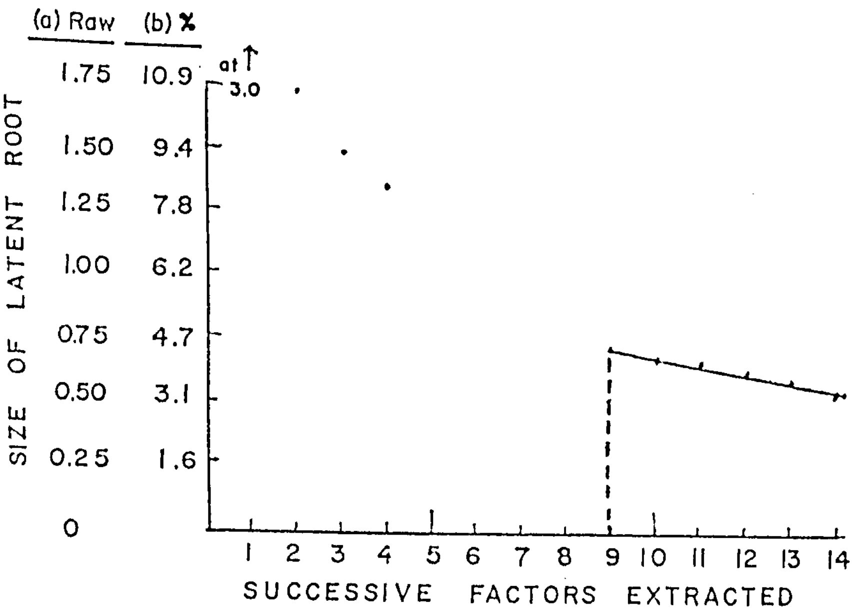

Fig. 3 (f). Data on Second Order 16 P.F., 7 factors, Gorsuch. Sample 1300.

As might be expected, the percentage before the scree cut-off point is lower when the sample size, the communality, and the test reliability of measurement are lower. One should beware, however, as pointed out in the next section, of assuming that the size of this unused, cut-off variance (up to 22.5% here) is the size of the specific-plus-error sample correlation common variance. For it is inflated here by having used unities in the diagonal which brings all specific variance (as defined in our factor model) into the common space (of the components model). Consequently, when the final estimate is made by entering with even a relatively high estimate for communalities this variance left beyond the scree cutting point is found to be very small indeed (2 or 3%).

The real question, therefore, concerns the point at which one decides to lop off, as trivial, (from further calculation and interpretation) the small substantive factors appearing *after* iteration of communalities to the chosen number of factors *and after rotation.* It is *this* percentage which is the desired objective and consistent definer of triviality. If we assume that, after rotation, the common error factors can be distinguished from substantive factors, (a) by inherent form, notably a tendency toward platykurtic distribution of loadings (because they affect many variables but never very much) and (b) by complete failure of cross matching of loading pattern with any other experiment (Linn, 1964), then this separation permits us also to calculate another couple of informative parameters concerning the properties of a given research. These values are the percentages of the total extracted (best possible communality estimate, to *n* factors) common variance which derive respectively from substantive and error-plus-specific (connotation described below) correlations. The aim of the scree is to yield this estimated total common variance lacking only the quite trivial variance in that lower part of the ogive curve of common error factor distribution which will be discussed in the next section. This latter rejected variance, we have seen, will be smaller with larger samples, etc., and will not be *entirely* free of substantive variance (because of the absence of a strictly one-to-one relation of rotated and unrotated factors even at the tail of the distribution). But this largely error variance thrown away (and which will tend to be *relatively* great in the last principal component factors) may well be only 1 or 2% of the total (communality directed) variance for *n* factors, and the substantive variance lost with it may well be only 10 to 20% of that, i.e., say, .1 to .4% of the variance used. More reliable estimates of these values need to be made by further research with plasmodes.

Prior to the scree test it had been suggested that as a practical precaution one should always employ at least two and preferably three independent tests to decide factor number, and specifically (Cattell, 1952; 1966 b), a choice from the Burt (1950), Lawley (1956), Rao (1955), Sokal (1959), and Tucker (1938) tests was indicated as best. Tucker's test, the earliest in this field, is as frankly empirical as the scree, and resembles it (and a whole class of tests), in depending on comparisons of the size of successive extracted factors. The present writer used Tucker's test through ten years of empirical work with increased certainty and consistency of results when compared with studies not using such a test, but advances in the last decade indicate a need to change. Thus Sokal (1959), comparing what he considered the best four methods, found on empirical data (essentially plasmodes) that it did not rank as high as the new criterion he proposed. For this reason, and in view of the fact that for the computer facilities of most people the Lawley-Rao Maximum likelihood is too costly (and works on a philosophy which causes extraction to stop too soon), one might reasonably advocate the consensus of the scree and the Sokal tests for decision-making in the majority of factor analyses. The verdict of the Kaiser-Guttman test which has been so widely adopted because of its ease rather than its rationale,[8] can, in our general experience, and in the light of its behavior in the above plasmodes (Figures 2 and 3), be considered misleading as a single test. (It will always be available as a third "piece of evidence," as a by-product, if one uses the scree and the Sokal test.)

Granted some skill with the scree test, and the regular inclusion of an independent supplementary check, our experience suggests that investigators will find the era of wide and wild disagreement on number of factors is at an end, and that independent studies will rarely disagree by more than one or possibly two factors in researches with ten to twenty factor dimensionality.

THEORETICAL BASIS

Although our prime concern is to demonstrate a law at an empirical, inductive level, we purpose in this section to discuss a possible theoretical explanation.

Beyond the substantive common factor variance, i.e., in the remaining unique variance of each variable, there will exist a large number of small influences divisible respectively into specific substantive factors and error of measurement factors. (Often only *one* e and s term is assigned to each variable, but strictly the formula describing the uniqueness, by this classical model, should be:

$$[1]\quad u_{ji} = s_{j1i} + s_{j2i} + \ldots + s_{jpi} + e_{j1i} + e_{j2i} \ldots + e_{jci},$$

where the s's are each specific to the variable j and the e's likewise.

According to the usual model these are uncorrelated in the population but correlated in the sample. For this the present writer will substitute, consistent with his treatment of factor analysis as a deterministic scientific model rather than merely a mathemati-

8. The important property which Kaiser (1960) has demonstrated, namely, that the α coefficient of homogeneity reaches zero at the K-G (latent root of unity) extraction limit, is sometimes taken as evidence that at any rate one cannot reliably *estimate an individual's scores* on any factors beyond the latent root = 1 limit. This misses three points:

(1) The estimation is assumed by Kaiser to take place from *all* variables: in practice one would take only the salients, which would, on an average, retain positive internal consistency for that *unrotated* factor estimation);

(2) When the factors are rotated (whereby the variance of any one factor more closely approaches the average of all extracted) *any* factor is likely to have a variance above this limit of unity even when the last *extracted* factor has dropped appreciably below the limit; and

(3) As in, say, the objective of rotating to determine correlations among primary factors, most basic research is not much interested in any substantial accuracy of factor *measurement*, but only in locating the factor as a concept, with its pattern of loadings, and its correlations with other factors.

cal model (Cattell, 1952, 1966b), a different set of postulates as far as the breakdown of the unique contribution in (1) is concerned. (Note that (1) is already more specific as to assumptions than in the most common mathematical model.)

What may be called the *complex stratified factor model* (and which the reader should be alerted is as different from the ordinary *factor model* as the latter is from the *component model*) postulates that in addition to the variance in each variable accounted for by the substantive broad primaries there are contributions from:

(a) A set of "temporarily-specific broad[9] primaries." These, written below as f_s, would ordinarily be broad primary substantive factors if appearing in a matrix sufficiently comprehensive of variables. As such (despite chancing here to affect only one variable) they must be mutually *oblique*. However, since *several* of them account for whatever variance is thus contributed to any one variable, and since each of the several primaries' contributions to variable x will be differently correlated with the bits contributed by other primaries to other variables, the covariance among variables from this source will be small, because of the averaging of diverse signs. In fact in a natural system with an indefinitely large number of such factors, the correlation due to this part of each variable will approach zero leaving us close to the simple factor model.

(b) A set of "broad error factors," designated f_{e1} to f_{ed}. The usual model, which makes error factors specific and uncorrelated, is, of course, only one of many possible. The present writer judges it to be quite improbable compared to the present theory, namely, that error of measurement (observation) is both broad and specific. For many influences can readily be conceived which would be instrumental in producing error across a whole set of variables (see Cattell and Digman, 1964). The broad error factors would again, like all primary influences, be likely to be somewhat correlated (oblique).

(c) A set of truly specific primary substantive factors, s_1 to s_p, and truly specific error factors, e_1 to e_g. It may be asked, if these are truly specific and uncorrelated (in both population and sample), what is gained by postulating several of them (Equation 1) for any one variable? Also it may be asked why these specifics are taken to be uncorrelated when all other factors, as influences, are allowed in the model to be correlated? The issues here are as fine as the influences are small. The correlational independence may

9. Consistent with a clarifying nomenclature adopted elsewhere (Cattell, 1966), a *common* factor is common to many people (as in Allport's sense, but defined operationally by R-technique) and contrasted with a unique factor trait peculiar to one person (by P-technique). A *broad* trait, loading many variables, is the antithesis of a *specific*, peculiar to one. The expression "unique variance" is perhaps irretrievably bound by mathematical usage to $1\text{-}h^2$; but in our psychological model we can at least analyze it into essentially *specific substantive* and *specific error* variance, plus the more novel components here indicated.

be considered as expectation of such high specificity, or, alternatively, we may say that we are dealing in this case with only a mathematical factor—the orthogonal residual when common error is taken out (in the Schmid-Leiman sense). That is to say, we choose to make it orthogonal because we are not interested to pursue the complication of the model further in respect to the extremely small variances which our model explicitly postulates to remain in true specifics. Although this multiplicity and independence has no role in our present paper, there are, nevertheless, other circumstances in which it could affect the theoretical inferences, and we must consistently adhere also here to what on general grounds seems the most realistic model.

The model in (a), (b) and (c) can be summarized in equation (2) as follows (omitting the subscript i for the particular individual):

$$(2) \quad a_j = b_{jf1}f_1 \ldots + b_{jfk}f_k + b_{jfs1}f_{s1} \ldots + b_{jfsp}f_{sp} + b_{jfe1}f_{e1} \ldots + b_{jfel}f_{el} + b_{js1}s_1 \ldots b_{jsx}s_x + b_{je1}e_1 \ldots + b_{jey}e_y$$

where b's are loadings; the three sets of f's—f, f_s, and f_e—are respectively broad substantives, singly represented broad substantives, and broad error factors; the s's are absolutely specific substantives, and the e's are absolutely specific error of measurement influences.

It is well, perhaps, to repeat, that the above is a postulated description of what happens in the population, not just in the sample, i.e., it supposes correlated broad error and correlated non-primary variance, and does not consider broad error factors in the sample to arise from "chance" correlation of specific error factors uncorrelated in the population, as does the traditional position. For chance has no role in a deterministic model.

If we now glance at sampling effects it is well to recognize that the differences from the traditional position arise also from the extension of the concept of sampling from people to include sampling effects on all ten Cartesian sets bounding the data box (Cattell, 1966a). Of these, perhaps people, stimulus situations, response variables, observers, and conditions (occasions) will suffice, by illustration, to remind us of how the factor structure might be affected by sampling. There is value in introducing at this point the division of factors into transients, appearing in the matrix only of one experiment, and factors which are stable and invariant across experiments, provided we recognize this to be a matter of degree, and that a factor quite constant in loading shape may nevertheless sometimes "dip below the horizon," i.e., be likely to escape a particular research, because of a decline in its general variance size.

From experiment to experiment the size, form and intercorrelation of broad error factors will change because of changing conditions and observers (to name only two sets whose sampling is involved). Similarly broad substantive ("real") factors will change not only through sampling of people (some samples having more egregious individuals) but also through changing sampling of stimulus situations and background conditions. Parenthetically, taking correlations from random numbers simulates only the first of these sampling effects.

The present degree of explicitness of development of the basic deterministic model is necessary if we are to make inferences about the distribution of factors, and its change with changing sampling, test reliabilities, etc., as a guide to understanding the empirical use of the scree test. To summarize, we have arrived at a model which recognizes five distinct classes of variance in variables entering a correlation matrix. Two—specific substantives (s's) and specific errors of measurement (e's)—contribute only to unique variances. The common, broad variance among variables (communality, covariance) actually arises from the two remaining sources, in broad factors, and, since they are oblique, also from the three contributions in the covariances possible among them and the temporarily specific broad substantives. Since the full equation would be cumbersome, let us express this in condensed form as follows:

$$(3)\quad h^2 = (f^2 + r_1ff) + (f^2_e + r_2f_ef_e) + r_3f_sf_s + r_4ff_e + r_5ff_s + r_6f_ef_s + r_7f_e + r_8f_es + r_9f_se$$

where r_1ff is the covariance among the substantive factors, r_4ff_e that between the substantive and the broad error factors, f^2 is the variance of the broad substantives and correspondingly for the others. (In the rotation process the covariance among the larger factors is "cleared out" from the primary factor analysis description and transferred to the second order, but in the initial orthogonal extraction all this variance is involved.)

If space permitted a systematic analysis could advantageously be made concerning the variance magnitudes and numbers typical for the eleven classes of factors implied in (3), and the response of these to changing sampling and experimental controls. But, as highlights, one may perhaps roughly note that (a) all common variance in (3) except f^2 and f^2_e is at the second order and therefore (since r's among factors tend, empirically, to be less than among variables) of a distinctly lower order of magnitude; (b) in a worthwhile experiment, notably as to test reliabilities, f_e variance will be of distinctly lower order than f, though the largest broad error factors will exceed the smallest broad substantive factors (Cattell, 1958); (c) r_4, r_6, r_7, r_8 and r_9 being across domains, are likely to be smaller than within domains, and perhaps only the term r_5ff_s in this class is anything but negligible; (d) presumably the f_e's will be less numerous than the f_s's for they will tend to be peculiar to one experiment; (e) with small samples the relative magnitudes of factors of all kinds—f's, f_e's, and the second orders connoted by r_1, r_2, r_3, etc.—will oscillate in size more from experiment (sample) to experiment; (f) since errors of observations are less likely to be systematically repeated than other features of an experiment the four terms in (3) involving f_e will be more variable than the others, and the main f_e's as distinct from the f's and the second orders in $r_3f_sf_s$, r_5ff_s, r_7fe, etc., may be called "transients," since they will not repeat themselves as recognizable loading patterns; (g) with reduced reliabilities of measures, both f_e's and e's will tend to be larger; (h) with smaller samples of variables the terms involving f_s will tend to be larger.

If frequency distributions, by size of mean variance contributed by broad factors (primary and secondary, e.g., f and $r_3f_sf_s$) were built up, we should expect from the above deterministic factor model a set of substantives (primary and secondary), ranging from very large to comparatively trivial variance contributions, then another, lower, mode among the broad error factors, and two or three modes at a still somewhat lower variance contribution size, arising from the second orders from these more trivial sources (the last five terms of (3) above). What we have to go upon to guide a particular WSF decision in factor extraction, however, is not this distribution of variance visible at the rotated factor resolution position, but that manifest in the distribution of the orthogonal factors (latent roots of principal axes) at extraction. Now, although the present writer has found it necessary (Cattell, 1958) to stress with tyros that broad error variance does not come *only* in the last factors from the extraction process —but, rather, suffuses all—yet it is an empirical and understandable fact that most rotational transformation matrices have larger terms down the diagonal than elsewhere. In other words, from the way the principal axis ellipsoid is commonly obtained, the larger factors in the *rotated* system tend to have more association with the larger principal components, and the smaller with the smaller. The meaningful, rotated factors are spread out over those near to their rank in the corresponding extracted series. The result is that the directly extracted, unrotated factors, when arranged in series, present a smoothed out picture of whatever slopes, frequencies, etc., exist in the true factors, through the merging of variances of all adjacent and fairly adjacent factors in the meaningful (rotated) series.

If, as our theory suggests, we have in the meaningful (stratified model) factors a distribution composed of two and possibly four modally distinct distributions, having all but the first (the broad substantives) at an altogether lower order of size and higher order of frequency, what then should we expect in the "blended" extraction series?

For easy reference let us adopt the term "rubble factors" for the totality of factors in this lower size order, neglecting, in a description kept to essentials, that there may really be three or more overlapping sizes of rubble. For present inferences we may neglect also that some (the f_e's, the $r_6f_ef_s$'s, etc.) are transients whereas others (the secondaries in r_5ff_s) are not. Then, if factors from this latter were arranged in rank order of size, after rotation, they would form an ogive curve of descending sizes which, if the original distribution were platykurtic, would be *virtually a straight line for a substantial part of the middle range.* How then, in the scree, do we get what seems to be almost *exactly* a straight line for this part of the variance which we believe to come largely from the chance distribution of small rubble factors? The answer would seem to lie in several contributory causes, partly connected with the fact that we are dealing with the distribution *before* rotation, *and* with unities instead of communalities in the diagonal.

First, this line would be straighter if the distribution is platykurtic. Secondly, there is overlap and summation of the two neighboring distributions (the smaller substantive factors being smaller than the larger error factors and similarly among the rubble groups themselves) which would smooth out the end inflections of each. Thirdly, and most importantly, the introduction of unities into the diagonal, where the error correlations are comparatively small, *will exercise a stabilizing effect in the direction of a steady fall* in the extracted random factor sizes. Note the scree is freer of slight irregularities in the examples in Figures 1-3 on 300 cases or more. That this is the most important influence is shown also by the fact that the small factor extraction curve, when communalities are used, tends to an ogive even before rotation, as

FIGURE 4

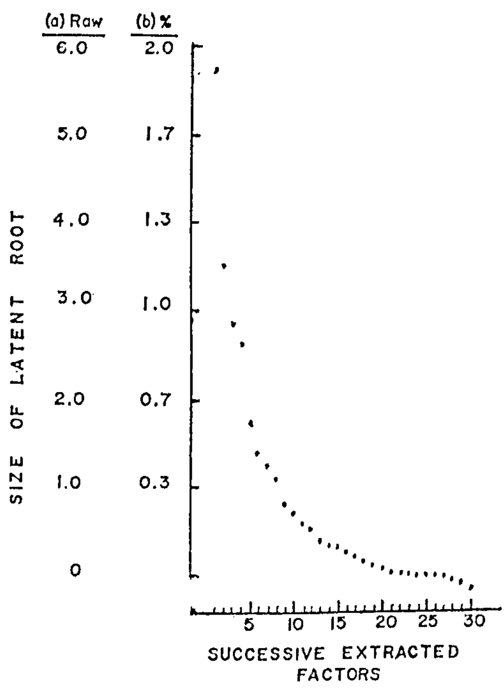

Fig. 4. Latent Root Distribution with Communalities on Data with Error (N=50).

shown in the repetition in Figure 4 of the Plasmode 30, N=50, with error, formerly represented in Figure 2 (c) with unities. Thus although there is nowhere near a one-to-one relation by rank order of unrotated with rotated factors, as pointed out above, yet the later extracted factors *tend* to be linear composities of the rubble factors so composed that they tend to form a smooth platykurtic distribution curve, and it is this which, with the above influences, produces the straight line scree.

A further empirical generalization which it seems can be reasonably made from our number of instances is that the scree tends to split into two successive straight lines, as illustrated in Figures 1(d), 2(a), 3(a), 3(c), and 3(e), and occasionally even into three. Indeed, the single scree line may well turn out to be the exception. Further work on the scree test should aim to develop the rationale relating N, n, the reliability of measurement, and the domain of variables to the length, slope, etc., of these divisions in the scree. Empirically, on the few cases yet available, we have sought associations explaining, for example, the transition from 3(a), with the upward inflection of the second part, to 3(c), with a downward inflection in the second part. The only generalization we have been able to form about this direction of inflection, when compared with magnitudes of N, n, s and e (from reliability) is that it appears to depend on N, inflecting downward as in Figure 3(c) when N is large. But on few instances this is tentative, and, theoretically, as discussed below, one would expect it to be a function of s, e and N.

However, there is no *practical* problem in applying the scree test, since the number of factors seems to be decided in these "split-scree" cases reliably by the end of the *upper* scree. One possibility of explaining a double scree by the theoretical model supposing that under certain circumstances the distribution of the f_e-generated transients in (3), on the one hand, and the f_s and f_e-generated second order variances, on the other, pull apart in size, i.e., giving three modes in all, including that of the substantives. Such a separation could be due to either (a) a marked discrepancy in size of the f_s and f_e variances e.g., high test reliabilities would reduce f_e, a lower n would tend to reduce h^2 (and therefore to increase u and thus f_s); or, (b) a marked discrepancy in the mean correlations of the f_s and f_e influences. The latter makes no sense on the classical model for both behave as random deviates and the mean r's would move together, with N. But our newer model *does* differentiate these two sources of rubble factors into two species, and later work may show that the double or treble scree does in fact correspond to experimental conditions that would be expected to separate them.

Both the empirical and theoretical arguments, however, point to the beginning of the scree as corresponding to the emergence, in the extraction process, of a class of factors different in nature and origin from the larger (real influence) broad error factors and from the non-trivial substantives. This reflection has evoked from some psychologists a dictum on design. It is asserted that in a good factor analytic design the substantive factors and their markers, should be so chosen that there are no small substantives, whereby the break between substantives and rubble will be brought out sharply, by a large and sudden descent to the scree. But if a good design requires one to put into a factor analysis just exactly so many large factors, and nothing else, then surely it is only a bad design which brings new information to the researcher! The conception of a good design must actually embrace both representative (exploratory) experiment as well as abstractive (hypotheses-testing) research (see definition in Cattell, 1966b). And even in the latter it is desirable to be able to find (by the *absence* of such a sharp break between substantive and rubble factors) that one has *not* confirmed one's particular hypothesis, i.e., one must still be able to abstract the appropriate number of factors even when this sharp difference does not exist. It is safe to say that in most good experiments as in most of the plots in Figures 1, 2 and 3 above, the scree will begin at a definite point, but smoothly, as does a tangent leaving a circle. It will begin where the non-trivial factors—substantive and broad error—end and where a new and large class of uniformly distributed trivial rubble factors starts.

The question still remains whether the start of trivial factors as assumed to occur through an approximately Gaussian error distribution of rubble factors, about a mode at a different order of magnitude from that of the substantives, corresponds with trivial as would be defined by our earlier quantitative definition of, say 2% of the total *substantive* factor variance. It obviously will not correspond to any fixed percentage of the total *extracted* factor variance, because the amount in rubble factors, which create the scree, will vary, as described above, with sampling of variables and people, and the reliability of the observers (measurement). If we are right in assuming that only the most trivial of the substantive factors will get included in the scree, however, the percentage they constitute of the total *substantive* variance will be both much smaller and more constant than the percentage of the total extraction area constituted by the area (in Figures 1, 2 and 3) under the scree.

Some investigation of the size of this total rejected variance, under the scree, covering both trivial substantives and the sources we have called rubble, could readily be undertaken, at least empirically. If we care to switch our model, and accept the classical mathematical model (simply of broad factors and [population] uncorrelated unique factors) as an approximation, some headway can also be made theoretically. Some idea of the size and distribution of factors that are closely equivalent to our rubble factors can be obtained by factoring correlation matrices obtained from score matrices of random normal deviates. (Though, as Horn [1965] points out, the distribution of random sample r's in a matrix when true r's are zero, is distinctly more complicated than is obtained from the standard error of n(n-1)/2 zero r's on N cases.)

Random deviate correlations can be factored either with "communalities" or with unities in the diagonal, and the results are very different. Dickman, Nesselroade and the present writer have at various times factored such "score" matrices, estimating communality either as largest r in the column or by the squared multiple lower bound procedure (Guttman, 1954) (iteration does not here work too well, producing alarming instabilities in communalities). With communalities the result is an initially steeply falling exponential type of curve of the same general character as for substantive factors. Such curves usually fall to zero around n/2 and pass into negative latent roots, just as a substantively based curve would do.

FIGURE 5

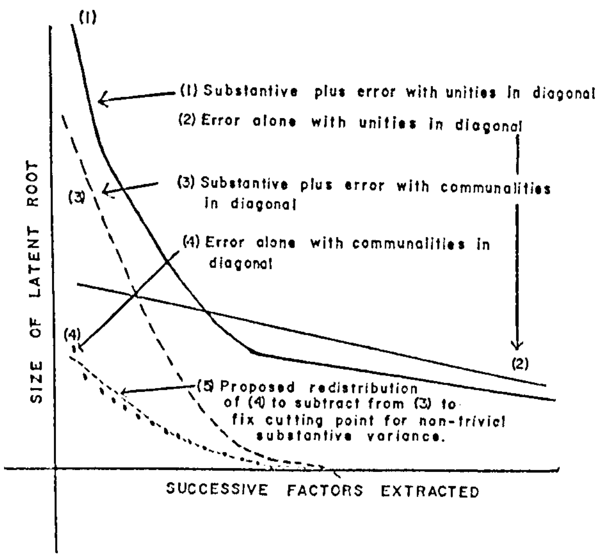

Fig. 5. Factor Extraction Latent Root Curves with Communalities and Unities in Diagonal, for Random Normal Deviate Correlations Alone and These Plus Substantive Score Correlations.

With ones in the diagonal, on the other hand, the successive latent roots drop in something very close indeed to a straight line, like the scree at (2) in Figure 5. As the sample gets larger the tangent gets smaller. This is why Figures 2(a), 2(b) and 3(f) (N = 8158, 300 and 1300) for example, finish, at the nth root, well above a zero latent root value, whereas Figures 1(d) and 3(d) (ii) (N = 50 and 93) end at n virtually at a zero root. Finally, as Horn correctly points out (1965), at the infinite population size we should expect a straight horizontal line at a latent root level of +1.0. (On the other hand we find no evidence for an inflection point—curtseying, so to speak, at the Kaiser-Guttman unity-root test value—as Horn (1965, p. 180, Figure 1) finds in his particular intersection of lines at the +1.0 line. Our general conclusions here on factoring random deviate correlations, with communalities and unities, seem to be empirically confirmed by Linn's data (1964), as far as they go, and a few other, sporadic, privately circulated independent results which have come to the writer's notice.

What happens when significant correlations corresponding to well defined substantive factors and random normal deviate correlations are factored together is a more complicated story, and one which we believe has been falsely simplified in some recent published conclusions or methodological practices. In considering this let us set aside at the beginning the case which causes no problem, namely, that wherein there is a definite number of relatively large substantive factors, unencumbered by any family of smaller associates and on so large an N that error common factors are trivial. Then, even with unities in the diagonal and before rotation, the larger factors will show some separation and end in a little precipice dropping to the scree of rubble. The nearest approach to this in our examples is Humphreys' data in Figure 2(a).

In the far more typical case, smaller substantives and larger broad error factors actually show, even after rotation, a single smooth declining curve as drawn schematically elsewhere (Cattell, 1958, Diagram 3, p. 806) and shown empirically here. If now we return to ask in this more typical case, first, what fraction of the area under the total curve is typically constituted by the rubble (here the random correlation factors), and, secondly, how this variance is distributed in the extracted factors before rotation, our answer is that it is smaller than has generally been assumed, and that it distributes itself differently from what has been supposed.

For in one of the first explicit, and theoretically clearly argued statements of what lies behind common practice, Horn (1965) (with acknowledgments to Humphreys, and apparently consistently with Linn's (1964) empirical conclusions) makes two assumptions. First he proposes that the magnitude of the rubble factors (now defined, however, in "classical" way, as the broad error factors occasioned by that correlation of uniquenesses which occurs only in the sample) is to be estimated by generating random normal deviates, for the N and n of the given experiment, and factoring their correlations with unities in the diagonal. Secondly, he proposes to subtract this curve from the substantive plus error (with ones in the diagonal) of one's actual experiment and to conclude that one's factor extraction should really proceed only to the point where the resulting difference curve falls to a certain value. This point is not zero but the K-G stipulated +1.0. One finds that the results of this procedure on a good (Horn, 1963) empirical study ($N = 297$; $n = 64$) disagree radically with those from the scree. In this careful and substantial study the random number procedure points to 9 factors instead of the 18 directed by the scree.

In rejecting the procedure just described the present writer is unfortunately compelled to disagree with Horn and other co-workers in this field on three rather widely accepted assumptions and practices: (1) The acceptance of the latent root of unity as a stopping point; (2) The "random deviates with unit communality" derivation of the magnitude of the broad error factor variance; and (3) The assumption about the way the variance in these factors will distribute itself.

As to the first, enough has been said above,[10] and regardless of the acceptance of the present writer's arguments, the objections just given to (2) and (3) would still hold firm and are sufficiently disabling to the position attacked. As to (2), entering one's in the diagonal of course considerably inflates the common error variance, and it is inconsistent at one and the same time to enter unities for the error[11] *and* for the error plus substantive common variance. Obviously the communality assigned to the error in the latter alone should be far below unity and to attempt to abstract the error by subtracting what one gets when unities are assigned to the error factoring is therefore likely to produce a gross over-correction. It is even evident to the eye that this is so on looking at Horn's Figure 2 (1965). As to (3) our empirical random deviate factorings show that, with communalities in the R diagonal, the extracted rubble factors drop off in variance size at first sharply, just like substantive factors (curve 4 in Figure 5) though in their later course, and with unities in the diagonal, they tend, as the scree shows, to form a straight line.

Reflections on this evaluation of the size and distribution of rubble factors (conceived, however, not in our model above but classically, as correlations of sample specifics and error) have naturally excited the idea of obtaining a plot of the rubble variance at extraction and subtracting it from the actually obtained curve of extraction of latent roots. This *might* leave a residual series of true substantive factor variance for the rotation process. How-

10. Footnote 7 introduces the criticisms, but perhaps they can advantageously be listed in summary together here! (1) Guttman's argument for + 1.0 really applies to the population not the sample; (2) The "psychometric" argument for this upper bound of extractable factors as the limit of a positive α coefficient (Kaiser, 1960), (a) applies to the last of k *unrotated* factors and is therefore very likely to be a marked underestimate for the last of the k *rotated* factors from these, since rotation commonly "evens up"; (b) overlooks that a significantly homogeneous estimate can be made by confining the estimate to the higher loaded variables; (See footnote 8.) (3) As Dickman (1960) himself admits, in arguing for the psychometric bound, there is, as we have stated earlier, no particular logic in stopping at a factor with the variance contribution of the average variable. Better, surely, if "triviality" is the criterion, to accept some fixed percentage (99%?) uniformly across studies as suggested above.

11. To avoid tedious repetition of "broad factor variance from correlated specifics and error contributions in the sample" we shall condense to "error" from here on, until the context indicates that we have changed to the finer analysis in equation (2).

ever, so far this line of thought has yielded nothing reliable. If one is willing to take that approximation to the deterministic model stated here which is presented by considering the unique variance as random normal deviates that are uncorrelated in the population, one could then obtain the correlations among these values in the given sample appropriate for the N of that experiment. Pre and post multiplying this correlation matrix by a diagonal matrix of u values (obtained by any good method of evaluating h in the original R matrix) one obtains covariances equivalent to those due to rubble factors. These could be subtracted from the original covariance matrix (with a rationale somewhat similar to Kaiser's alpha analysis and resemblances also to Bartlett's Chi square test) before factoring, or factored separately and subtracted as latent roots from those produced by the original R matrix (to determine at what point a factor residue reduces to zero). The difficulty with this idea is that when substantive and rubble factors are extracted together the distribution of the latter is neither all at one end (as in the naive assumption we have criticized above) nor absolutely uniformly distributed. (By this we mean retaining the same proportion to the true substantive variance throughout the sequence of extracted factors. In Figure 5 we have sketched some re-distribution of rubble latent roots in curve 5 to indicate this intention, but, as stated, there is really no firm basis for doing so.)

Subtraction at the factor stage, as has been proposed (in the form of curve 2 in Figure 5) as a WSF test, must therefore be abandoned. On the other hand the subtracting of a generated rubble correlation matrix from the original R matrix, as mentioned above, with iterative convergence on a more accurate u estimate, may have promise, but that is not part of our present theme.

In regard to the theoretical foundation of the scree, even though we would defend the deterministic theory that broad and oblique error factors exist, and, indeed, the whole model presented in equation 2 above, yet a weakness in the argument for certain postulates remains. For the theory supposes that f_s and f_e terms are so much more numerous and so much smaller than the f terms that they (and their covariances) create a smooth normal distribution (playkurtic) of rubble factors which would (a) be separate from the distribution of the (large and few) substantives, and, (b) extract as a uniformly dropping succession of latent roots. However, we know that these conditions are often likely not to be fully met in that (a) the substantive factors (f's) will also frequently tail off into a debris of small factors, and (b) in small samples or with tests of low reliability subject to broadly influential errors of observation the largest broad error factors can shoulder their way up among the moderate sized substantive factors. The first we have accepted from the beginning—that when substantives are no larger than the rubble factors they will be lost from the rotation following use of the scree test, but we have suggested this lost substantive variance might be only an acceptable

1% to 5% of the total substantive variance. The second has been checked empirically, where it has been shown that the scree, when measurement error gets large, will err on the side of indicating one or two more than the known number of substantives. The inclusion of such a broad error factor among those accepted for rotation (demonstrable not only by being extra to the demonstrable, known substantives in a plasmode, but by being transient from one experiment to another), however, creates no problem (as would an *under*-estimate), and actually improves the precision (narrowness) of hyperplanes obtained in the rotation of the substantive factors (Cattell, 1966a).

In conclusion the scree's theoretical basis clearly deserves further attention. But, meanwhile, the validity of the use of the scree test in practice does not rest upon the correctness of the above tentative theoretical analysis of its origins. It depends strictly on inductive generalization, which also, however, needs to be broadened, in its basis of observations, by further controlled experiment with plasmodes, etc.

SUMMARY

(1) There is no such thing as "the true number of factors to extract," since the only possible assumption is that both the number of substantive and the number of error common factors each exceed n, the number of variables. Consequently the cut-off point in extraction is best decided by a conception of *non-trivial common variance,* which might be adjusted between, for example, 95% and 99%, according to circumstances and objectives, but which, like the familiar $p < .05$ and $p < .01$ significance levels, could advantageously take some level consistent across researches. However, this defines a final cut-off with respect to the *rotated substantive* factor variance, and needs to be translated into an appropriate figure for the pre-rotated, substantive-plus-error variance as extracted initially.

(2) The practice of using statistical tests which cut off factor extraction at a point where error mounts to the point of giving, say, a $p < .05$ significance to the last factor extracted is questioned. Since error and substantive variance are contained in a proportion changing only slowly across successively extracted factors, this results in a serious Type 2 error in which the pattern of the ultimately obtained rotated substantive factors is distorted by absence of essential parts of their variance. It is better to take out most error too and trust to the process of rotation to separate substantive and error factors. With error rotated into a "garbage factor" or factors, the substantives will have exact hyperplanes (except for *specific measurement* error or non-linearity) regardless of size of sample.

(3) The *stratified model* adopted here is a deterministic one according to which what the classical model has called broad (com-

mon) factors due to correlation of specifics and errors (uncorrelated in the population but correlated in the sample) become considered as real factors, though of three species, different from that of the large real substantive factors. They are considered respectively to be: (a) both broad and specific *error of measurement observation factors,* due to real influences in observation; (b) second order factors from correlations among ordinary substantive oblique broad factors which appear (in the given matrix) only once, and which might be called "temporarily specific broad factors;" (c) second order factors from correlations of broad primaries and of specifics *across* the various domains (notably of f_s and f_e primaries). Sampling error (as contrasted to measurement error as here invoked) is not an issue, except in so far as it alters the *magnitudes* of the above defined factors, and causes f_e's to be perhaps totally transient. This model has been expressed above in three formulae, and is called "stratified" because it supposes that factors at different strata levels will appear in the primary extraction.

(4) According to the new model one would expect the several species of factors named in (3) to be decidedly more numerous and of a smaller order of size than the majority of the substantive factors sought in research. Considered together as "rubble factors" they should form a normal distribution modally separated from the substantives. In factoring substantives and rubble together one would expect (especially if the rubble distribution is platykurtic) to find the rubble factors succeeding each other in a prolonged ogive curve. If unities are used in the diagonal, this and the overlap of the ogives would suffice to produce a straight line—the scree.

Investigation of the form to be expected from the successive latent roots in rubble factors is attempted through accepting the approximation to our deterministic theory presented by the classical "specific substantive and specific error factors uncorrelated in the population but correlated in the sample." Empirical R matrices from random normal deviates with communalities yield factor distributions little different from those of substantives. But with unities a straight line appears, approaching the horizontal as N increases. A criticism is offered of the use of this line, along with the Kaiser-Guttman "latent root of unity" test to decide the number of factors. A more appropriate use of correlating random deviates in an experiment with N cases is suggested, using a uniqueness estimate, but it does not contribute to a scree type of test.

(5) In any case the scree test does not rest for its practical validity upon the correctness of the theory or inferences from it, but on an inductive law, some of the empirical evidence for which is presented here in a stratified sample of types of experiment. It is also illustrated by four plasmodes of varied n's and N's, and error admixture, i.e., numerical and mechanical models where the

number of substantive factors is known, as well as by several psychological examples in which the number of factors is known to a high probability from circumstantial evidence. In all of these it succeeds in giving the expected number of factors, provided we recognize that its task is also to take out at least the largest of the broad error factors too, i.e., that which is equal in size to the smallest substantive factor the researcher is interested in (given the usual definition of "trivial").

This systematic property in the scree test of taking out the largest observational error factor (or even factors) rather than stopping short and underestimating the substantives needing to be taken out, actually results in clearer hyperplanes, since the "garbage factor" collects variance which would otherwise blur substantive hyperplanes (and loadings). It fits the above theory of at least three sources of rubble factor variance (which could show three distinct distribution modes) that two and three successive straight lines (at different slopes) sometimes appear in the scree. In these cases it is demonstrable empirically that the number of factors put in, in a known plasmode, is indicated by the *upper* limit of the upper straight line in the scree, as theory would suggest.

This investigation was supported in part by Public Health Service Research Grant MH 10274-01 from the National Institute of Mental Health. The writer wishes to express his indebtedness to John Horn, John Nesselroade and Gene Glass for valuable discussion and help with the empirical examples.

References

Burt, C. L. Tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sect.*, 1950, 3, 77-85.

Cattell, R. B. *Factor analysis, for the psychologist and social scientist.* New York: Harper, 1952.

Cattell, R. B. Extracting the correct number of factors in factor analysis. *Educ. psychol. Measmt*, 1958, 18, 791-838.

Cattell, R. B. Theory of fluid and crystallized intelligence: a critical experiment. *J. educ. Psychol.*, 1963, 54, 1-22.

Cattell, R. B. The data box: its ordering of total resources in terms of possible relational systems. Chap. 3 in R. B. Cattell (Ed.) *Handbook of multivariate experimental psychology.* Chicago: Rand McNally, 1966 (a).

Cattell, R. B. The strategic use and meaning of factor analysis. Chap. 6 in R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology.* Chicago: Rand McNally, 1966 (b).

Cattell, R. B. The scientific model rationale for deciding the number and nature of factors in factor analytic experiments. In press, 1967.

Cattell, R. B., and Dickman, K. A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychol. Bull.*, 1962, 59, 389-400.

Cattell, R. B., and Digman, J. M. A theory of the structure of perturbations in observer ratings and questionnaire data in personality research. *Behav. Sci.* 1964, 4, 341-358.

Cattell, R. B., and Gorsuch, R. L. The uniqueness and significance of simple structure demonstrated by contrasting organic "natural structure" and "random structure" data. *Psychometrika*, 1963, 28, 55-67.

Cattell, R. B., and Meredith, G. M. Second and third order personality structure in Q-data at the teen age level by the pure factor method. In press.

Cattell, R. B., and Nesselroade, J. Untersuchungen der inter-kulturellen Konstanz der Persönlichkeits Faktoren im 16-P.F.-test. *Psychol. Beiträge*, 1965, 8, 502-515.
Cattell, R. B., and Sullivan, W. The scientific nature of factors: a demonstration by cups of coffee. *Behav. Sci.*, 1962, 7, 184-193.
Coombs, C. H. Criterion for significant common factor variance. *Psychometrika*, 1941, 6, 267-272.
Dickman, K. W. Factorial validity of a rating instrument. Ph.D. thesis, Univer. of Illinois, Urbana, Ill., 1960.
Fruchter, B. Manipulative and hypothesis-testing factor analytic experimental designs. Chap. 10 in R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966.
Gorsuch, R. L., and Cattell, R. B. Second strata personality factors defined in the questionnaire medium by the 16 P.F. In press *J. Soc. Psychol.*, 1966.
Guttman, L. Some necessary conditions for common factor analysis. *Psychometrika*, 1954, 19, 149-161.
Harman, H. H. *Modern factor analysis*. Chicago: Univer. of Chicago Press, 1960.
Horn, J. L. A rationale and test for the number of factors in factor analysis *Psychometrika*, 1965, 30, 179-185.
Humphreys, L. G. Number of cases and number of factors: an example where N is very large. *Educ. psychol. Measmt*, 1964, 24, 457-466.
Kaiser, H. T. The application of electronic computers to factor analysis. *Educ. psychol. Measmt*, 1960, 20, 141-151.
Lawley, D. N. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 1956, 43, 128-136.
Linn, R. L. Use of random normal deviates to determine the number of factors to extract in factor analysis. Unpublished M.A. thesis, Univer. of Illinois, 1964.
McNemar, Q. On the sampling error of factor loadings. *Psychometrika*, 1941, 6, 141-152.
Nesselroade, J. R. An empirical examination of factor invariance with different rotational methods and indices of matching. Unpublished M.A. thesis, Univer. of Illinois, 1965.
Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, 20, 93-111.
Sokal, R. R. A comparison of five tests for completeness of factor extraction. *Trans. Kansas Acad. Sci.*, 1959, 62, 141-152.
Thurstone, L. L. *Primary mental abilities*. Chicago: Univer. of Chicago Press, 1938.
Thurstone, L. L. *Multiple factor analysis*. Chicago: Univer. of Chicago Press, 1947.
Tsujioka, B., and Cattell, R. B. Cross cultural comparison of second stratum questionnaire personality factor structures—anxiety and extraversion—in America and Japan. *J. soc. Psychol.*, 1965, 65, 205-219.
Tucker, L. R. In L. L. Thurstone, Primary mental abilities. *Psychometr. Monogr.*, 1938, No. 1. Chicago: Univer. of Chicago Press.