

**UCLA**

**Working Papers in Phonetics**

**Title**

WPP, No. 16: PARAFAC

**Permalink**

<https://escholarship.org/uc/item/0410x385>

**Author**

Harshman, Richard

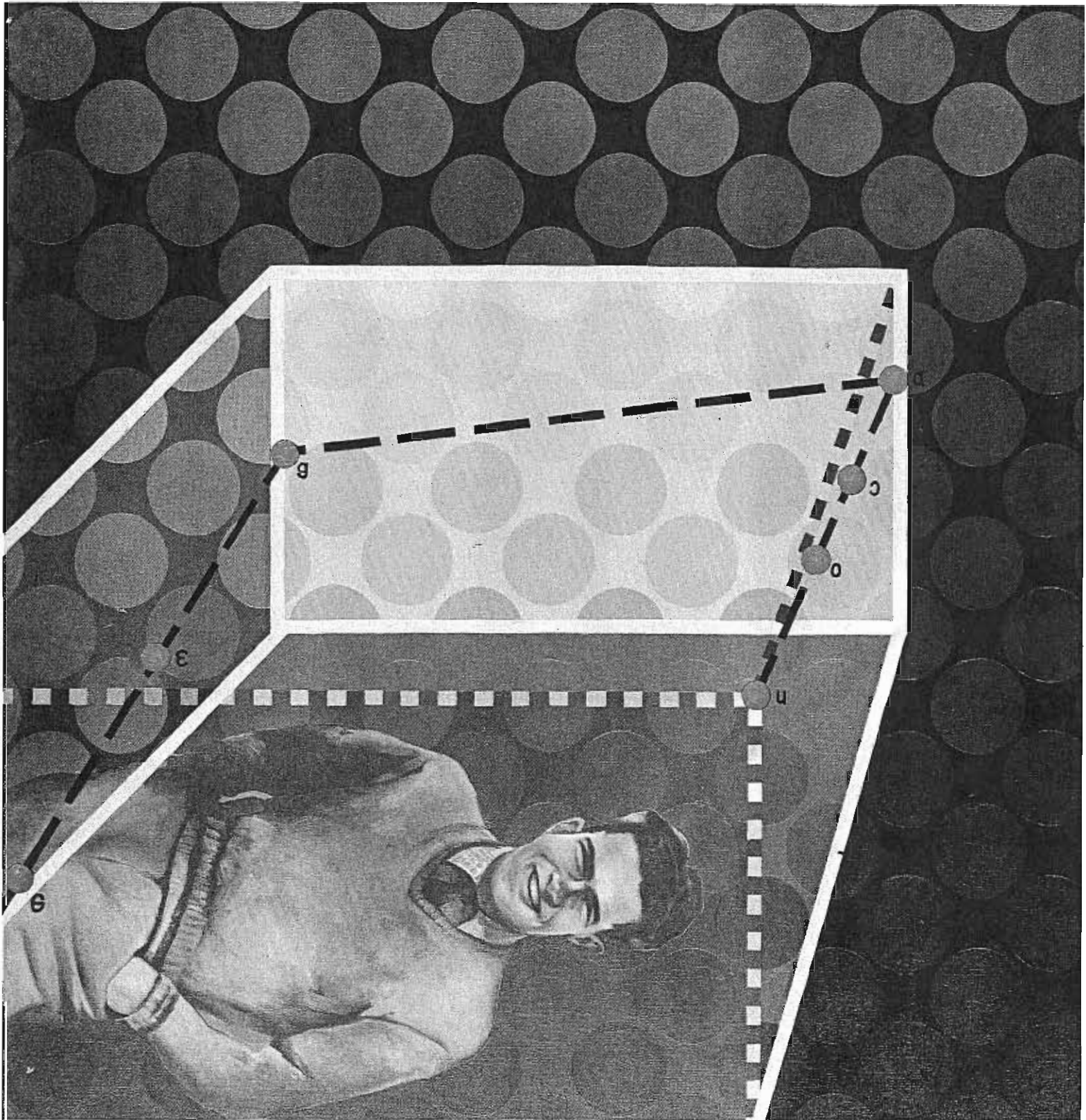
**Publication Date**

1970-12-01

U.C.L.A.  
W.P.P. NO. 16  
DEC. 1970



PARAFAC  
by  
RICHARD A. HARSHMAN



FOUNDATIONS OF THE PARAFAC PROCEDURE: MODELS AND CONDITIONS  
FOR AN "EXPLANATORY" MULTI-MODAL FACTOR ANALYSIS

by

Richard A. Harshman

U C L A

*Working Papers in Phonetics*

16

December, 1970

## ACKNOWLEDGEMENTS

The author wishes to express his great indebtedness to Peter Ladefoged, without whose support this work would probably not have been possible.

The author is also very grateful for the help of Hans Reichenbach with the optimization program, Robert Jennrich for the proof of uniqueness, and the fast algorithm, and to Dale Terbeek, Lee Cooper, Andrew Comrey, and Peter Bentler for their comments, criticisms, and the generous donation of their time.

It was the author's good fortune to have the manuscript prepared by Jeanne Yamane who contributed an unusual combination of mathematical and secretarial abilities, plus care and concern with the quality of the final result.

The computation for this project was paid for by an intramural grant from the UCLA Campus Computing Network, and performed on their IBM 360/91 computer.

Finally, the author wishes to thank Diane Vaughan (now Mrs. Diane Harshman) for her helpfulness, love and support which have contributed so much toward the successful completion of this project.

Cover: Bill Jahnke (oil on canvas)

This research has been funded in part by NSF grant GS 2741, USPHS grant NB 04595, and UCLA intramural research funds.

## ABSTRACT

Simple structure and other common principles of factor rotation do not in general provide strong grounds for attributing explanatory significance to the factors which they select. In contrast, it is shown that an extension of Cattell's principle of rotation to Proportional Profiles (PP) offers a basis for determining explanatory factors for three-way or higher order multi-mode data. Conceptual models are developed for two basic patterns of multi-mode data variation, system- and object-variation, and PP analysis is found to apply in the system-variation case.

Although PP was originally formulated as a principle of rotation to be used with classic two-way factor analysis, it is shown to embody a latent three-mode factor model, which is here made explicit and generalized from two to  $N$  "parallel occasions". As originally formulated, PP rotation was restricted to orthogonal factors. The generalized PP model is demonstrated to give unique "correct" solutions with oblique, non-simple structure, and even non-linear factor structures.

A series of tests, conducted with synthetic data of known factor composition, demonstrate the capabilities of linear and non-linear versions of the model, provide data on the minimal necessary conditions of uniqueness, and reveal the properties of the analysis procedures when these minimal conditions are not fulfilled. In addition, a mathematical proof is presented for the uniqueness of the solution given certain conditions on the data.

Three-mode PP factor analysis is applied to a three-way set of real data consisting of the fundamental and first three formant frequencies of 11 persons saying 8 vowels. A unique solution is extracted, consisting of three factors which are highly meaningful and consistent with prior knowledge and theory concerning vowel quality.

The relationships between the three-mode PP model and Tucker's multi-modal model, McDonald's non-linear model and Carroll and Chang's multi-dimensional scaling model are explored.

## CONTENTS

I.	Introduction: The Search for "Explanatory" Factor Analysis . .	5
II.	The Principle of Parallel Proportional Profiles . . . . .	15
III.	Development of the Three-Mode PP Model and Other Generalized Factor Models . . . . .	18
IV.	Testing the Three-Mode Proportional Profiles Model . . . . .	27
V.	A Uniqueness Theorem for PP Factor Analysis . . . . .	61
VI.	Generalization to Non-Linear Factor Models . . . . .	63
VII.	The PP Model in Relation to Other Non-Linear and Multi-Modal Factor Models . . . . .	69
	References . . . . .	82

## I. INTRODUCTION: THE SEARCH FOR "EXPLANATORY" FACTOR ANALYSIS

Basic to the use of factor analysis as a tool for scientific discovery is the distinction between its "descriptive" and its "explanatory" application. While descriptive factor analysis seeks merely to find a convenient, condensed representation of data relationships for a given case, explanatory factor analysis seeks to discover good estimates of the structure of "true underlying" influences that are responsible for the observed data relationships. The degree to which factor analysis can provide explanatory solutions has always been a matter of controversy.

The problem lies in the factor model itself. Descriptions in terms of the classic factor model are not sufficiently constrained by the data. For any given set of data in which there is more than one factor common to the measures, the classical factor model allows for many possible sets of factors, all of which would equally well fit the data in a manner consistent with the model. But although these solutions are mathematically equivalent within the constraints of the model, they do not lead to equivalent interpretations of the influences underlying or organizing the data relationships. The classical factor model does not give us grounds for choosing among these many different possible explanations or interpretations of the data relationships.

Is the Search for Explanatory Factors Meaningful?

How meaningful is the distinction between convenient descriptive factors and so-called "true" or "valid" explanatory factors? A basic part of the controversy over explanatory factor analysis centers on the question of whether there ever exists a unique "true" structure of factors which "really" underlies a set of data. Thurstone, for example, argues against the search for unique solutions in factor analysis:

When a factor analysis has been completed, ... the first question is naturally: "What *are* the parameters or factors?" It seems to be in the nature of science that such a question has no unique answer. *In scientific work a parameter is one of the measurable attributes of an object in terms of which it is described ... a circle on a graph is defined by the two co-ordinates (x,y) of its center and by the radius r. These parameters are easily understood and easily used in most problems, but, of course, they are not unique. For some problems they would be awkward and another set of parameters would be chosen. The same is true in scientific work ... To hunt for a unique solution in the comprehension of a set of related phenomena is an illusory hunt for absolutes.*

(Thurstone, 1947, p. 332) (his italics)

But in this comment, Thurstone seems to be conceiving of factors as playing a purely descriptive role. For such descriptive factors the selection of any particular solution from among the possible alternatives is based entirely on the simplicity, convenience and usefulness of the resulting description. In Thurstone's classic box problem (1947, p. 360), for example, the variables all consisted of different measurements on the size and shape of boxes (e.g. area of a side, length of the diagonals, length of the perimeter, etc.). Relations among these variables are "passive", in that they have no causal significance. There are no functional systems involved, and none of the factors extracted could be thought of as underlying causal influences. Since no *explanation* of the observed relations is sought, the choice of length, width, and depth as the extracted factors is somewhat arbitrary. It is simply the convenience of working with descriptions in terms of those three dimensions (as opposed to some factors which are linear combinations of length, width, and height), which justifies favoring this solution.\*

Such descriptive applications of factor analysis should be contrasted with applications in other domains, where the factors sought are the causes of patterns of covariance among the measures, or where one wishes to identify functional systems in natural processes. For example, tidal level and amount of daylight might be explanatory factors underlying the patterns of covariation in measurements of seashore activity, and functional systems might be sought in the ecology of the tidepools. The notion of an explanatory factor does not allow choice among factorial representations to be based on simplicity or convenience of the solution since a correspondence with some real external structure is sought.

Consider the classic debate over the structure of human intelligence. From the same matrix of correlations among tests many different descriptions of the structure of intelligence can be derived. In one description, intelligence consists of one general intellectual factor plus many uncorrelated group factors corresponding to specific skills; in an alternative description, intelligence consists of a set of correlated basic intellectual abilities. Can one decide among these alternatives, or is the search for a unique solution to this question simply an "illusory hunt for absolutes"?

It might be argued that the two different sets of explanatory factors are really two different ways of describing the same underlying system. And it is true that any real system, consisting of a complex network of interrelationships, can be described in a number of different ways. But *explanatory* descriptions have implications beyond the current set of measurements being described. Explanatory descriptions imply predictions about

---

\* Since these factors were determined by simple structure rotation, they presumably represent a factorially invariant description. This gives that solution another type of convenience or usefulness. The significance of factorial invariance will be discussed later in this section.



the results of other possible experiments. Therefore descriptions which are equivalent for a given data set are not generally equivalent as *explanations* of the relations in that data set. It is true that such things as changes in terminology and level of detail can alter one's perspectives while still producing valid and equivalent explanatory descriptions, but in factor analysis one's descriptions cannot be so freely reformulated. With any given factor analysis one is dealing with some particular limited set of empirical variables and these variables constitute the descriptive "vocabulary". Any factorial description of the underlying functional system must consist of specifying specific relations between each variable and each latent factor. In light of these considerations, it seems much less likely that two alternative sets of explanatory factors could be really equivalent.

It would seem necessary to try to decide between alternative sets of explanatory factors whenever the two explanations are "really different". In science, being "really different" amounts to having different empirical consequences. The fact that both models of human intelligence have the exact same mathematical consequences within classical factor analysis need not show the "meaninglessness" of the distinction. Rather, it might show the weakness of the classical factor model. In fact, the two conceptions of intelligence could not be called "really the same" unless they had the same empirical consequences (would make the same predictions) in all conceivable circumstances (e.g. in developmental studies, under the influence of drugs, etc.). But such is clearly not the case. They would, for example, predict different possible patterns of disruption of intellectual functions by drugs or as a result of brain damage. It would seem that most sets of alternative explanatory factors which are equivalent within the classical factor model are nonetheless meaningfully different, since the hypotheses they suggest would make differing predictions in certain appropriate circumstances.

Because of these differences, one approach to the explanatory use of factor analysis has been to use factor analysis to generate alternative hypotheses (such as differing models of intelligence) and then devise and perform crucial observations or tests in order to select among the alternatives. With this approach the solutions of factor analysis only have explanatory force as part of a larger experimental program.

A single factor analysis can often generate a large number of alternative hypotheses (since for a given multi-factor matrix there are an infinite number of alternative sets of factors). Therefore it would be desirable to develop a system of factor analysis which would directly determine unique explanatory factors. This could be accomplished by enlarging the conceptual-mathematical model of factor analysis in such a way that the many alternate sets of factors are no longer equivalent within the model, but indeed have differing consequences which allow one solution to be preferred. This amounts to strengthening the model or the rotation process so that it incorporates the same type of ability to distinguish among alternative classical solutions that was previously only

possible by using factor analysis in conjunction with external experimental tests. For such a strengthened type of factor analysis, the fact that a given set of factors was discovered underlying a set of data would directly give those factors explanatory plausibility. This article proposes such a system of factor analysis. But before it is developed an examination should be made of the weaknesses of the traditional techniques for selecting among alternative sets of factors.

#### Traditional Approaches to Choice of Factors

The problem of determining which set of factors will be derived from a given factor analysis has been traditionally handled by adding a second stage of analysis after the factor extraction process. This second stage of analysis is called factor rotation, and consists of successively transforming the factors into mathematically equivalent alternative sets of factors, until some added criterion or principle is fulfilled. This added criterion forms the basis for a choice among the infinite number of possible solutions derivable as rotations of any given factor solution.

An alternative, however, would be to replace the classical factor model with one which is not inherently underdetermined by the data. No additional rotation criterion would then be needed. But although Cattell (1944) came close to this alternative (see Sections II and III below), the search for explanatory factor analysis has proceeded, instead, by the development of principles of rotation for solutions derived from the classic factor model. And although great progress has been made along these lines, no fully adequate solution to the rotation problem has been found. As Comrey says "the rotation process has been the target of much criticism, and continues to be the weakest link in the entire [factor analysis] process" (Comrey, 1967, p. 143).

One approach to the rotation problem is exemplified in a number of "special principles" which have been adopted by individual researchers to choose a rotation for their particular data or type of data. These principles usually amount to finding a rotation which seems "reasonable" or "meaningful" in that it agrees with knowledge about the data based on other grounds than the factor analysis itself. Insofar as this brings additional valid information to bear on the choice of rotation, it is a desirable improvement over an unrotated arbitrary solution. But this approach can yield many different solutions when used by individuals with differing knowledge or opinions about what constitutes an approximation to a "reasonable" set of relationships. Whatever merit such special principles might have in a particular circumstance, they do not constitute an *objective general* criterion which would tend to select the rotation with explanatory validity.

#### Is Simple Structure Explanatory?

The closest that factor analysis has come to an agreed general

principle of rotation is the concept of simple structure. The object of this approach is to minimize the number of factors influencing each variable. This amounts to finding a rotation which, as much as possible, allows each variable to be described in terms of a few factors which influence it strongly, rather than many factors which influence it less strongly. This simplifies the factorial description of the variables, hence the name "simple structure". The initial definition of simple structure was given by Thurstone (1935, p. 156; 1947, p. 335) and consisted of rules giving the proportion of near zero values that should appear in the rows and columns of the factor matrix. This definition of "simplicity" has been supplemented and extended with the advent of computerized analytical rotation methods. Varimax, for example, tries to achieve a complimentary type of simplification, i.e. to explain each factor with as few variables as possible (Harman, 1960, p. 301).

Rationales have been presented for rotation to simple structure in both descriptive and explanatory factor analysis. In the case of descriptive factor analysis, the justification is simple. A few large loadings are easier to deal with than many small ones, and, in a sense, rotation to simple structure gives the most parsimonious description of a set of data.

In explanatory factor analysis, rotation to simple structure is less clearly justified. Any description, however compact and convenient, which does not correspond to the underlying causal or "functional unities" (to use Cattell's terminology) would be misleading, if interpreted in an explanatory sense.

The strongest general argument for preferring simple structure solutions when looking for explanatory factors is based on the principle of factorial invariance. This principle was formulated by Thurstone (1947, p. 361) as follows: "A fundamental requirement of a successful factorial method, [is] that the factorial description of a test must remain invariant when the test is moved from one battery to another which involves the same common factors." The description of a given variable should not fluctuate wildly from one study to another as a mere mathematical artifact of the rotation process. If reality is consistent, then any "explanatory" solution would have to reflect this consistency.

It has been argued by Thurstone (1947), Cattell (1952) and others, that simple structure rotations tend to have this property of factorial invariance. Kaiser (1958, p. 195) has made this claim for the Varimax criterion. He presents an argument to show that it would have this invariance in the special case in which all variables fall into two colinear clusters, but stated that the generalization to a greater number of factors would involve enormous difficulties.

Cattell seems to argue that some types of invariance give convincing proof that such factors represent the true form of underlying organic unities:

... the rediscovery of the same factors despite (a) partially different test batteries or (b) populations of different age, education, or dispersion, and (c) independent factorizations and rotations, is a proof that they have an existence as something more than mathematical equivalents -- that they are in fact functional unities in nature. (Cattell, 1955, p. 90)

Such arguments, based on a consistent finding, do not rule out the real possibility of a consistent distortion of the true form of the latent factors being consistently generated by the repeated application of consistent, but inappropriate, rotation criteria.

Even if simple structure could be clearly demonstrated to have perfect factorial invariance across data sets (as defined by Thurstone), this would not be a convincing argument for the explanatory validity of simple structure rotations. This is because factorial invariance is a necessary but not sufficient condition of explanatory validity. To see this, simply imagine that there is a certain real latent structure underlying a set of variables (let us call this structure  $S$ ), but also imagine that this latent structure is not of simple structure form. Now if a factor analysis is performed of data with latent structure  $S$  and the simple structure criterion is used for rotation, then a different factor structure will be extracted, let us call it structure  $S'$ . Now it is likely that the form of  $S'$  is largely determined by the form of  $S$ , i.e. some sort of systematic distortion of  $S$ . Since  $S$  is a real and invariant latent structure measured in several different factorial studies, one would expect  $S'$ , the simple structure distortion of  $S$ , to be consistently extracted from all these studies. The stability of  $S'$  across a number of different studies shows (perhaps) that it is a consistent transformation of a real latent data structure, but it does not in any way show that it is an undistorted expression of that latent data structure. Thus the finding that certain simple structure factors have "stood up" under the test of time and in a number of different factorial studies does not in itself give any evidence that they are accurate descriptions of the latent influences underlying the data. It only gives evidence for the existence of *some* factors that are related, by a consistent but unknown rotational transformation, to the simple structure solution.

An inverse and equally objective criterion of "most complex structure" could probably be defined and would probably show similar consistency across studies. If not "most complex structure", clearly there are going to be some principles, related in a systematic way to simple structure, which would show similar invariance of results across studies. Thurstone admits that there are probably principles other than simple structure which would show factorial invariance (Thurstone, 1947, p. 364). But if this is so, then factorial invariance alone is clearly an insufficient argument for attributing explanatory significance to a principle of rotation.

An important additional requirement for any "true" factor solution was proposed by Tucker (1958, p. 112). It is the converse of the principle of factorial invariance of variable description, namely "that the scores of individuals on a factor should remain invariant as the individuals are tested with different batteries which involve the factor". Clearly, if a person is found to have a small loading on the hostility factor with one battery, and a large loading with a different battery of tests, both cannot be measuring a "true" objective factor pattern (assuming that the individual hasn't had time or circumstances which might change his personality).

Tucker's principle, which might be called a principle of factor score invariance, is also a necessary but not sufficient condition of explanatory validity. The same argument just advanced in relation to factorial invariance could be used here. Nonetheless, when these two principles are fulfilled in the rotation, we can have *more* confidence in the explanatory validity of a resulting solution than we could if it were just an arbitrarily selected rotation. Only a subset of possible rotations should fulfill these criteria, and the true solution would be found among the members of that subset.

If there is not, in general, a sound basis for attributing explanatory significance to factors determined by simple structure rotation, might there nonetheless be a valid explanatory application in some particular circumstances? The most sophisticated use of simple structure involves just such restricted application to certain selected types of data sets. The reasoning behind this approach has been well developed by Cattell (1955). He argues that not all correlation matrices will yield factors which can be rotated to a clear and compelling simple structure (with a number of loadings very near zero for each factor), and therefore the discovery of such a clear structure among the possible rotations of a factor analysis is of great significance.

In this approach, it is not justifiable to take the results of a factor analysis and rotate them to the *best approximation* of a simple structure, in order to maximize the invariance and interpretability of the solution. Instead, it is acknowledged that a simple structure might not really underlie the data, and therefore, if the results of a given factor analysis cannot be rotated to a sufficiently clear and compelling simple structure, the results are not interpreted. In such a circumstance it is concluded that the latent factors did not show a sufficiently well defined simple structure, and thus the rotation criterion cannot be trusted to have correctly discovered their true underlying form.

In this approach it is admitted that simple structure is *not* a general procedure for extracting explanatory factors from data sets. Instead, the criterion is applied only to data sets in which the variables have been specially selected, on the basis of prior knowledge and

experience, so that none of the presumed latent influences would be expected to affect all of the variables included in the set. This approach recognizes that simple structure rotation will give explanatory solutions only if the true latent factors in fact have simple structure form. Therefore, an attempt is made to design data sets to fit this requirement. If, after extracting a set of factors from such a data set, it is indeed possible to find a rotation which shows a clear simple structure, this is taken as a confirmation of the original expectations. It is concluded that the attempt to build a data set whose factors would have simple structure was successful, and therefore simple structure rotation can be used as an appropriate guide to the true form of the underlying factors.

But even this restricted use of simple structure for explanatory factor analysis is subject to serious theoretical objections. The most fundamental objection is to the central claim that the discovery of a clear simple structure provides a confirmation of one's initial conception that the latent influences would have simple structure form. *Considered by itself*, such a discovery is at best only suggestive. This alleged "confirmation" could in fact occur when the latent factors were not in simple structure form but instead were merely in some other form which was capable of rotation into a clear simple structure. While it is true that there exist many sets of factors which cannot be rotated into a clear, compelling simple structure, there of course also exist many others which can be so rotated.

What, then, is the likelihood that the discovery of a clear simple structure is really a "false confirmation", revealing not the true latent structure but instead one of its alternative rotations? This question is difficult to answer. Certainly, the less clear and compelling the discovered pattern (i.e. the fewer near-zero loadings there are for each variable, and the further these loadings are from zero) the more likely it is that the resemblance to simple structure could have occurred by chance. But even if a configuration is discovered which is so clear that it could *not* have occurred by chance, it still need not have been due to simple structure patterns of influence. Other non-random influences or data characteristics can produce systematic factor relationships (hyperplanes oblique to true factor axes) which are rotatable to an apparent simple structure, but which would not display the zero loadings of a simple structure when the factors have their true orientation. A partial example occurs with the real data analyzed later in this article (see Figure 2, and Table 8, Section IV). The fact that alternative possible simple structures can sometimes be detected in a given data set (Cattell, 1955, p. 251, quoted below) further indicates that orderly patterns can arise which are not due to simple structure, but which can mimic the zero loadings of simple structure if rotated into a certain position.

In actual application, Cattell and others use many grounds besides the discovery of a clear simple structure to argue for the explanatory

validity of any set of factors (e.g. interpretability of the factors, relation to prior findings, etc.). The case for choosing the simple structure rotation from among the many possible rotations is strengthened by arguments based on prior knowledge of the data and its presumed latent influences, and in this way the factors gain in explanatory plausibility. But as with any choice of factors based partly on prior expectations and on "meaningfulness" of the solution, the stronger one can make such an argument for explanatory validity, the more one must know about the presumed latent structure *before* the analysis, so the less one stands to learn from a factor analysis.

Because of the tediousness of graphical rotation, and because some researchers object to its alleged "subjectivity", most simple structure solutions are now discovered through one of the analytical computer rotation procedures. But although the various computer rotation procedures give somewhat similar solutions, the results are sufficiently different to affect interpretation of the rotated factors. For example, "normal" Varimax rotation (Kaiser, 1958) tends to distribute the variance evenly across many factors, while the Quartimax method (see Harman, 1960, Chapter 14) tends to concentrate it on a general factor, with the remaining factors having smaller influences (Harman, 1960, p. 302).

An additional uncertainty arises from the controversy over orthogonal vs. oblique rotations. Restricting the factors to orthogonal positions is equivalent to requiring that all factors be uncorrelated with one another. But this complete independence of underlying influences seems unlikely to hold in many real situations. On the other hand, some workers have pointed out weaknesses in the principles of oblique rotation (Comrey, 1967; Harman, 1960).

Even the uniqueness of simple structure solutions is sometimes questionable. Cattell comments that more than one simple structure position is sometimes found for the same set of data (using graphical rotation) "and in absence of any generally acceptable criterion we cannot decide, from the given matrix alone and without extraneous evidence, which is preferable" (Cattell, 1955, p. 251). It has recently been demonstrated that Varimax is not always going to give a unique rotation, but it does not seem to diverge too seriously from this ideal (Gebhardt, 1968).

The basic question, of course, remains; what grounds do we have for believing that a given rotation corresponds to the true underlying structure of influence that generated the data? Comrey questions the reasonableness of simple structure for many applications:

If we sample at random from the entire universe of factors and use predominately factor-pure measures, simple structure will no doubt give results which are reasonably satisfactory. In many real-life factor analyses, however, where selection of variables is anything but random, and measures of considerable factor complexity sometimes predominate, one can

only hope that the simple structure criterion is approximately applicable. Simple structure pursued to its logical conclusion, i.e. by increasing to a maximum the number of near zero loadings per factor while allowing the axes to go oblique, can give results which may be misleading. The varimax criterion can also present problems. There is no particular reason why the variance of the squared factor loadings must be maximized, except that loadings are more easily interpreted if they are either high or low rather than medium in absolute value. It may well be, however, that they should be in medium range, rather than high or low, depending on the data being analysed.

(Comrey, 1967, p. 143)

Comrey presents an alternative pair of rotation criteria for factor analysis, namely that variables which appear on the same factor should be correlated, and uncorrelated variables should not appear on the same factor. This principle, Comrey admits, would not always be valid in real life, but "the principle will be substantially appropriate for most correlation matrices taken in their entirety". The argument for this principle relies on expected probabilities. As with the simple structure criterion, we are making implicit assumptions about the probable latent structure (i.e. that it is not of the sort that would systematically violate the two principles). But perhaps the probabilities are more in our favor with Comrey's assumption than with an assumption of simple structure. (Until one is more familiar with the significance of the conditions wherein the principle might be expected to be *systematically* violated, it would seem hard to judge.)

In view of all of the questions raised in this section, there does not appear to exist a principle of rotation from which one could make a strong and general argument for explanatory validity. One wants to find conditions where the rotation selected necessarily corresponds most closely to the latent structure that created the data relationships (except for deviations due to error variance). Such conditions would presumably be based on logical or mathematical considerations deriving from the factor model itself. Many factor theorists have given up hope of such a principle. Yet in 1944 Cattell came tantalizingly close to a solution to the problem when he formulated his principle of Parallel Proportional Profiles (Cattell, 1944). His approach is the foundation for the solution to be proposed in this article.



## II. THE PRINCIPLE OF PARALLEL PROPORTIONAL PROFILES

In 1944 Cattell published a review of principles which might be used for determining the desired rotation of a set of factors (Cattell, 1944). His objective was to discover the principle of rotation which had the strongest likelihood of explanatory validity. He rejected the notion that psychologists should be satisfied with a "convenient description" and nothing more. "Psychological research, as such, cannot ... be content with this limited goal. It strives toward psychologically meaningful functional unities ... 'How can one decide which one, among many possible sets of factors, alone corresponds to the real functional unities in the psychological situation?'" (Cattell, 1944, p. 267).

Cattell reviewed seven principles for choice of rotation. These were: (1) Rotation to agree with clinical and general psychological findings; (2) Rotation to agree with factors from past factor analyses; (3) Rotation to put factors through the center of clusters; (4) Rotation for simple structure; (5) The principle of orthogonal additions: rotation to agree with successively established factors; (6) The principle of expected profiles: rotation to produce loading profiles congruent with general psychological expectations; (7) The principle of parallel proportional profiles.

After discussing and finding objections to principles 1-6, Cattell focused his attention on (7), which he called "the most fundamental principle". He starts with the notion that "if real psychological functional unities exist they are bound to appear as possible mathematical factors in many different kinds of situations, whereas mathematical factors which are artifacts will stand only the test of fitting the particular matrix in which they happen to appear and may not be reproducible elsewhere" (Cattell, 1944, p. 274). But agreement across two different factorizations is not, itself, sufficient.

... it is clear that to require agreement in factors and factor loadings among correlation matrices derived from the same or similar test variables on the same or similar population samples, is an empty challenge. No new source of rotation determination is introduced, for such matrices will differ only by sampling errors and there will be an infinite series of possible parallel rotations in the two or more analyses. The special and novel required condition is that any two matrices should contain the same factors but *that in the second matrix each factor should be accentuated or reduced in influence by the experimental or situational design*, so that all its loadings are proportionately changed, thereby producing, from the beginning, an actual correlation matrix different from the first.

(Cattell, 1944, p. 274) (his italics)

The most concise statement of the logic of the parallel proportional profiles criterion occurs in a later article:

The basic assumption is that, *if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study.* No inorganic factor, a mere mathematical abstraction, would behave in this way ... This principle suggests that every factor analytic investigation should be carried out on at least two samples, under conditions differing in the *extent* to which the same psychological factors ... might be expected to be involved. We could then anticipate finding the 'true' factors by locating the unique rotational position (simultaneously in both studies) in which each factor in the first study is found to have loadings which are proportional to (or some simple function of) those in the second: that is to say, a position should be discoverable in which the factor in the second study will have a pattern which is the same as in the first, but is stepped up or down.

(Cattell, 1955, p. 84) (his italics)

This rotation criterion overcomes the objections leveled at the other techniques. (1) It is completely general, in the sense that it makes no assumptions about the size or distribution of factor loadings in a given study, but only asks that they be proportional to the loadings for that same factor in another study. (Although for philosophical reasons Cattell also called this principle "simultaneous simple structure", it in fact bears no relation to the mathematical assumptions and procedures of simple structure rotation criteria.) (2) It is objective, in that it does not require judgments of "reasonableness" based on other knowledge of the data under study. (3) Finally, the criterion has explanatory validity, since we would expect the parallel proportionality conditions to be fulfilled with real latent factors and *only* with real latent factors.

But there was a basic problem with the implementation of the Proportional Profiles criterion. Cattell was only able to establish a procedure for discovering proportional profiles with orthogonal factors (Cattell, 1955). If the true underlying factors were not in fact orthogonal, then Cattell's technique would not be able to discover them. Since it is not normally possible to be sure, in advance, that the factors underlying a given data set are orthogonal, the explanatory validity of factors selected by Cattell's PP procedure would be questionable, at best.

Demonstrations were made of the successful application of parallel proportional profiles with the orthogonal restriction to synthetic data

of known (orthogonal) factor composition. But the failure of a study using real data raised serious questions about the principle's effectiveness:

Still seeking for living and organic (rather than merely physical) examples that would nevertheless be controllable we were encouraged by the recent discovery of drives as factors to suggest that hunger and thirst be factored from the usual learning and deprivation variables on two populations of rats. This work, by Haverland (Haverland, 1954) had larger variation of food deprivation in one sample and larger variations of water deprivation in the other. The many interesting findings cannot be described here; but it is questionable whether proportional profiles gave a meaningful solution, whereas simple structure gave easily recognisable drive patterns (best oblique, but acceptable even when orthogonal).

(Cattell, 1955, p. 85)

Because of the use of correlations as the base for analysis, Cattell encountered problems with his simple proportionality criterion. "If a variable is loaded by a factor 0.5 and the variance of that factor (still orthogonal) is doubled relative to other factors the loading will not equal 1.0. If we apply the formulae of Thompson and Ledermann (1939) and Thurstone (1947) for predicting the correlations to be expected after the variance of the selection variable has been changed (counting the factor as the variable on which the selection is made), it will be found that not all the loadings of a factor will change in the same proportion when the factor variance is changed" (Cattell, 1955, p. 86).

At the fundamental philosophical level, parallel proportional profiles seemed clearly superior to all other rotation criteria.\* It held out promise, in 1944, of answering the fundamental rotational problem of factor analysis in a satisfactorily general, objective, and persuasive fashion. It would have presented a strong argument of explanatory validity, since it sought to select a factor solution based on a form of invariance which one could expect only in "true, underlying factors".

But parallel proportional profiles has not been developed and successfully used as a principle for factor rotation. It has languished, apparently bogged down in mathematical difficulties and restricted to orthogonal factors. Thus explanatory factor analysis has remained in the unsatisfactory condition described earlier. But (as subsequent sections of this article will show), when Cattell's principle of proportional profiles is incorporated into a revision and generalization of the factor model itself, techniques can be developed for arriving at unique explanatory factor solutions which will allow not only orthogonal and oblique factors, but non-linear ones as well.

---

\* for data showing system variation. This qualification will be developed in Section III.

### III. DEVELOPMENT OF THE THREE-MODE PP MODEL AND OTHER GENERALIZED FACTOR MODELS

Cattell developed his principle of proportional profiles (PP) simply as a criterion for selection among the possible rotations of a pair of classical factor analyses. In fact, however, it provides a fundamental clarification or enrichment of the notion of a "true underlying" factor. As such, it can be considered as a modification of the classical factor model.

Whereas the classical factor model was concerned only with the behavior of factors within a given study, Cattell's contribution was to consider also the expected behavior of factors in a pair of related but differing studies. Proportional profiles is a precise definition\* of the expected behavior of the loadings of "true underlying" factors in a pair of studies where the relative influence of the factors changes.

To use Tucker's terminology (Tucker, 1964) classical factor analysis only considered two "modes of measurement" at a time. Typical of this would be the study of the scores of a number of persons, on a number of measures. The resulting data consisted of a two-way matrix of persons against measures. But the principle of proportional profiles is concerned with persons, on measures, *on two different occasions*. It is fundamentally a three-mode concept.

#### The Three-Mode PP Model

The classical equation of factor analysis could be written, for  $m$  factors, as:

$$(1) \quad x_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \dots + a_{jm}F_{mi} + U_{ji}$$

Here  $x_{ji}$  is a score on measure  $j$  for person  $i$ . Contribution of any factor (e.g. factor 2) to that score, is given by taking the product of  $a_{j2}$  the loading of the  $j^{\text{th}}$  measure on that factor, and  $F_{2i}$  the score of the  $i^{\text{th}}$  person on that factor. These values represent the relative influence of that factor on the  $j^{\text{th}}$  measure and in the  $i^{\text{th}}$  person. The last term,  $U_{ji}$  represents the contribution of a "unique factor" for each data value, and consists of the "specific factor" contribution plus error.

---

\* Assuming, as shall be done here, that conventions for expressing loadings are revised so that proportionality is retained when factors change relative influence (e.g. change from correlations to functions which allow loadings to exceed 1.0). This will be developed later.

The three-mode interpretation of parallel proportional profiles gives a straightforward generalization of this classical factor equation. The generalized equation would predict scores in a three-way data matrix, composed of elements  $x_{kji}$  representing the score for the  $i^{\text{th}}$  person on the  $j^{\text{th}}$  measure on the  $k^{\text{th}}$  occasion of measurement. The contribution of a given factor (e.g. factor 2) to a given score is computed in part as it was before, by multiplying the degree of that factor's influence for that individual,  $F_{2i}$ , times the degree of that factor's contribution to that variable,  $a_{j2}$ . But this estimate of the factor's contribution has to be modified by a coefficient describing the relative influence of that factor on that occasion:  $o_{k2}$ . Thus all the contributions of factor 2 on occasion  $k$  are stepped up or down in a parallel proportional fashion by the coefficient  $o_{k2}$ . The classical factor model becomes generalized as follows:

$$(2) \quad x_{kji} = o_{k1}a_{j1}F_{1i} + o_{k2}a_{j2}F_{2i} + \dots + o_{km}a_{jm}F_{mi} + U_{kji}$$

In this formulation, the principle of proportional profiles applies to any number of occasions, rather than just two. It is no longer a principle for rotation of a classical solution, but instead it has become incorporated into a revision of the basic factor equation itself. Not only will this formulation play a basic role in the development of a PP analysis procedure, but its incorporation of more than two occasions will prove crucial for the successful practical application of PP analysis to data sets involving more than a very few factors (in Section IV, below).

However, this form of the factor model is not completely general. It is not always true that the contributions of the factors from one occasion to the next would vary systematically in this manner. Cattell was aware that the proportionality would not always hold, and he discussed experimental designs in which one might ensure the desired systematic variation in factor influence. His unsuccessful experiment with drives in rats (Haverland, 1954, as cited in Cattell, 1955, p. 85) was an example of an attempt to ensure that the relative influences of the presumed factors of hunger and thirst were systematically different from one group of rats to the other. But, perhaps because of his close association with personality-measurement applications, Cattell did not fully develop the theory of factor analysis to which the parallel proportional model could be applied, and distinguish it from the theory of a different type of factor for which an entirely different type of model would be necessary. This theoretical basis will be developed here in terms of two corresponding conceptual models for factor analysis.

## Two Conceptual Models for Factor Analysis

Behind the *mathematical* model of factor analysis (the fundamental equation describing scores in terms of contributions from factors)

there must lie a *conceptual* model, a logical or semantic interpretation of the terms in such an equation by virtue of which its application to real data is deemed reasonable. Only when factor analysis has its foundation in some conceptual model can one hope that it will tell us something of explanatory value about the world. It would seem, therefore, that one logical approach to a three-mode factor model would be to start at the conceptual level, to apply the classical conceptual model of factors and factor analysis to three modes of measurement, and then from this enlarged *conceptual* model to develop an appropriate generalization of the *mathematical* model. Unfortunately, this is not possible.

It soon becomes apparent that the classic two-way *mathematical* model of factor analysis is, in fact, an ambiguous representation of at least *two* fundamentally different *conceptual* models. Although these conceptual models have equivalent algebraic representation in the two-way case, their algebraic generalizations to three-way and higher order data relationships are substantially different.

For the purposes of discussion, the two basic models will be called (1) the "system-variation" model and (2) the "object-variation" model of latent data influences. In the "system-variation" model, unitary latent influences act across many objects, and thus the factor scores of the objects will vary proportionally across any third mode, (e.g. across occasions of measurement). In the "object-variation" model, latent influences are present separately in each object, and "factors" represent common *types* of these individual influences. Here the factor scores of each object will vary independently of the variations of those scores in other objects, when measured across a third mode. It will be useful, before proceeding further, to develop a clear picture of these two conceptual models.

#### (1) The System-Variation Model

Consider the example of an economic system. For a set of data, different individual businesses might be measured on a number of variables (e.g. production, investment, salaries, layoffs, etc.) on each month throughout a five year period. In such a set of data it would be reasonable to look for factors present in the system as a whole, operating across a number of individual businesses. One such system-factor might be described as "inflationary pressure". When this factor increased from one month to the next, its contribution to individual scores in the data matrix would rise in different businesses in proportion to their loadings on that factor. It would appear (in terms of the classical factor model) that their factor scores had all increased proportionately. The loadings of the individual businesses ( $F_{1i}$ ,  $F_{2i}$ , ...) on that factor (classically called factor scores) would here represent the relative susceptibility of each business to the influence of that factor. The loadings of the different months ( $\alpha_{k1}$ ,  $\alpha_{k2}$ , ...) would represent the relative degree of activity or influence of that factor in the system as a whole for that month (in this case the fluctuations in inflationary pressure in the

economy from month to month). The loadings on the variables ( $a_{j1}$ ,  $a_{j2}$ , ...) would represent the degree to which each type of measurement was influenced by (or reflected) changes in that factor (in this case, the degree to which layoffs, salaries, etc. are influenced by "inflationary pressure"). The system-variation conceptual model would be appropriate for such data, and thus it would be appropriate to apply the parallel proportional profiles mathematical model for three-mode factors which was developed earlier in this section.

Another source of data which could be expected to show system-variation of factors would be any biological system, for example the brain of a laboratory animal. Electrodes could be placed in different brain centers, and the EEG activity recorded from each electrode could be subject to spectral analysis. Sets of these spectra could be measured for each of a number of different occasions, yielding a three-way data set of measured spectral intensities. Its modes would be frequency by brain center by occasion. The system-factors sought in such an experiment would be the hypothetical "generators" of EEG activity within the brain. Each factor might correspond to one such generator. It would have large loadings on those spectral frequencies corresponding to the type of EEG "wave" it generated. Its factor loadings in different brain centers would describe the relative influence of that generator in the brain activity at those locations: the "effective proximity" of those brain centers to the generator. The loadings on different occasions would represent the relative "activity" of that generator across those occasions. Certain EEG generators could be expected to be much more active during certain stages of sleep, others during alert goal-oriented behavior, etc. Again, the factors are acting *in* the system, *across* "individuals". The loadings of "individuals" (e.g. brain centers) represent their different susceptibilities to these common latent influences. Therefore it is reasonable, with this conception of brain activity, to expect parallel proportional changes in factor influence across a third mode of measurement as the latent factors change in their relative degree of activity or influence.

Other likely examples of system-variation data might be found in studies of languages, weather patterns, ecological systems such as a forest, and social systems. The key feature of the factor model in all these systems is that the factors are interpreted as single, unitary influences, existing in the system as a whole, and through the system influencing individuals in a coordinated fashion.

The system variation concept will immediately generalize to cases of four-mode and higher order relationships. A four-mode application might arise, for example, if an attempt were made to analyze (1) a number of measurements of (2) a number of different speech sounds, as made by (3) a number of different individuals, and as occurring in (4) a number of different languages. If  $c_{pn}$  is taken to represent the loading of the  $n^{\text{th}}$  factor on the  $p^{\text{th}}$  condition of measurement, then a four-mode linear model for system-variation would be written:

$$(3) \quad x_{pkji} = c_{p1} a_{k1} a_{j1} F_{1i} + c_{p2} a_{k2} a_{j2} F_{2i} + \dots + c_{pm} a_{km} a_{jm} F_{mi} + U_{pkji}$$

## (2) The Object-Variation Model

Now consider, by contrast, the example of personality or aptitude measurement. In a typical personality study a group of several hundred persons are administered questionnaires containing items like "When I get angry, I throw things." and are asked to rate how much each item applies to them. The result is a two-way data set of answers by persons. If one re-administers the questionnaires to the same individuals on a number of different occasions, a three-way data set (answers by persons by occasions) will be generated. But the properties of such a data set are fundamentally different from the system-variation examples previously described, and therefore require a different conception of "factors" and their relation to the data.

In cases such as personality or aptitude measurement, the "factors" are not conceived of as residing in the *system*, but instead in the *objects* or individuals being measured. Such "factors" are not unitary influences across individuals, but are merely "unitary" across a set of measures within a given individual. The meaning of extracting factors from a whole population of individuals lies in identifying characteristic *types* of latent influences within individuals or objects. When a researcher says "a factor of hostility clearly emerged in this population," he means that a characteristic *type* of influence was identified. The factor of "hostility" occurs in a distinct, separate way in each individual. There is no reason to suppose that the "hostility" factor scores of two different persons would necessarily vary in parallel from one occasion of measurement to the next. This concept of the data and its variation is here called the object-variation model for factor analysis.

For two-mode factor analysis, the *mathematical* model is the same, whether the underlying conception is of a system-variation or object-variation case. In both cases the loadings of the individuals on the factors (the factor scores) represent the influence of the given factor for that particular individual. It does not matter whether that factor is conceived of as a unitary influence acting simultaneously on all the individuals of the system, or whether the "factor" is in fact conceived of as a type of influence which occurs separately within each individual. In both cases the  $(F_{1i}, F_{2i}, \dots)$  coefficients suffice for expressing in the two-way model the degree of the factor's influence in a given individual.

This mathematical equivalence of the two conceptual models breaks down for three-way and higher order data variation. As noted earlier, this is because the system-variation model predicts a coordinated variation of factor influence across a third mode (and any fourth or higher mode), whereas the object-variation model predicts that the changes in



the degree of expression of a given factor across any third mode should be uncorrelated from one individual to the next.

Equation 2 gives a mathematical model appropriate for a linear three-mode application of the system-variation concept. The corresponding linear three-mode model can be constructed to express the object-variation concept, as follows: the difference between the influence of a given factor on one occasion and any other consists of a number of uncoordinated shifts (generally small) in the amount of this factor in some of the individuals; this small shift or variation "V" in the factor score of a given factor for individual *i* on occasion *k* is symbolized  $V_{kli}$  for factor 1,  $V_{k2i}$  for factor 2, etc. The general object-variation expression for a given data value then becomes (in the linear, three-mode case):

$$(4) \quad x_{kji} = a_{j1}(F_{1i} + V_{k1i}) + a_{j2}(F_{2i} + V_{k2i}) + \dots + a_{jm}(F_{mi} + V_{kmi}) + U_{kji}$$

This model resembles a set of independent two-way factor models with respect to the factor scores of individuals, but it is a common model with respect to the factor loadings on variables. In this weak general form it will have no unique solution determined by the data. To try to secure a unique, "explanatory" solution, an additional constraint must be placed on the *V* values. Is there a criterion which can be applied to take advantage of the additional information supplied by the third mode, without demanding parallel proportional relationships which would only be found in system-variation data?

One possible constraint, suitable for many sets of object-variation data (including personality questionnaire data) can be derived from considerations of continuity. Even when an individual's variations in factor scores are not related to those of *other* individuals, they are still related to his *own* previous scores, if the occasions of measurement are "sufficiently close" to one another in the third mode (e.g. sufficiently close in time if the third mode covers successive occasions of measurement). Just how close is "sufficiently close" will depend, of course, on the rates of change expected in the factors underlying the particular data. Perhaps this could be estimated from the rates of change of the variables themselves. If the occasions are "sufficiently close" to one another, then the shifts or *V* values, measured from one occasion to the immediately prior or succeeding one, should always be small, even when the cumulative shifts over many occasions might be large for some individuals. We could expect most types of underlying factors to show this sort of continuity over samples taken at sufficiently short intervals. Therefore a reasonable additional constraint upon the *V* values of equation 4 would be the condition:

$$(5) \quad \text{minimize} \quad \sum (V_{kfi} - V_{(k-1)fi})^2$$

This elementary continuity measure has many shortcomings, and we shall immediately take advantage of an insightful and highly related paper by Shepard and Carroll (Shepard and Carroll, 1966, p. 581) and borrow their adjusted criterion (translated into our notation):

$$(6) \quad \text{minimize} \quad \sum_k \sum_{k'} \left[ \frac{V_{kfi} - V_{k'fi}}{k - k'} \right]^2 w_{kk'}, \quad k \neq k'$$

where  $w_{kk'}$  is some monotone decreasing function of  $k - k'$ . In effect, this considers not only immediately adjacent factor scores but (to a lesser degree) scores in the vicinity, as determined by the  $w_{kk'}$  weighting function.

Whereas the proportional profile system-variation model will be studied in detail in Section IV, the actual properties of the object-variation model have not yet been investigated. The model is developed here because of its great theoretical interest and its relations to the system-variation case. As an initial comment, however, it seems clear that the factor scores of a number of individuals would have to shift significantly and in independent fashion in order for a unique solution to be obtained with the proposed continuity criterion. Since no tests have been conducted, it is not known how effective such a criterion would be, and under what conditions it could be expected to give an "explanatory" unique solution. But it seems definitely to merit further investigation.

#### Mixed Types of Variation

With some types of data, one might expect significant effects involving both system- and object-variation. Consider, for example, studies of the development of human abilities. As children grow older, general systematic changes can be expected to occur in the relative influence and expression of different ability factors. There will also, however, be variations in these factors for each individual due to the unique circumstances, history, and genetic complement of that child. A similar interplay of system-variation and object-variation would be expected in measurements of individuals going through a series of learning trials, mastering some complex skill.

One simple way of analyzing data containing such mixed variation patterns might be to ignore the object-variation and merely try to "fit" the system-varying factors. If, in a given case, all factors of interest could be expected to show sufficient system-variation, then the object-variation could be considered part of the "noise" or error of the model's fit to the data. On the other hand, if techniques for analysis of object-variation were successfully developed (along the lines suggested earlier

in this section), and if such an analysis were performed on mixed type data, then both the object- and the system-variation would be described by the tables of "V" values for each factor. An analysis of these tables would reveal precisely how much variation of each type was present.

The concept of analysis of the "V" tables needs to be explored more fully. When considering the possible structures of these tables one becomes aware of possible complex patterns which could not be attributed to either system-variation or object-variation, or even to any "mixture" of these two cases. These complex new types of possible data variation pose difficult questions of interpretation, apparently requiring new conceptual models. But since the appropriate models for such situations appear to be strongly related to Tucker's concepts developed for his three-mode factor analysis (Tucker, 1963, 1964, 1966), further discussion of these patterns will therefore be deferred until Section VII, where they will be discussed in relation to Tucker's approach.

The object-variation model given in equation 4 allows for change in the amount of a factor in any individual or object, but retains constant factor loadings on the variables. Even this restriction will sometimes seem inappropriate. It might sometimes be the case (e.g. in the child development example given above) that one could expect shifts in the loadings of the factors on the measures, as well as on the individuals. As children grow, they might alter their interpretation of a given question, or there might be shifts in the degree to which certain abilities are used to solve a given problem. A more general model of object-variation might therefore be written:

$$(7) \quad x_{kji} = (a_{j1} + V'_{klj})(F_{1i} + V_{kli}) + (a_{j2} + V'_{k2j})(F_{2i} + V_{k2i}) + \dots \\ \dots + (a_{jm} + V'_{kmj})(F_{mi} + V_{kmi}) + U_{kji}$$

With this generalized object-variation model, the continuity constraint would have to apply to both the  $V$  values of factor score variations, and the  $V'$  values of factor loading variations.

The generalized object-variation model of equation 7 suggests that a classic factor analysis might be performed on each of the successive occasions of a three-way data matrix, and then an attempt should be made to rotate these separate solutions into the greatest degree of agreement with one another. Techniques have been developed for such rotation of two sets of factors (Wrigley and Neuhaus, 1955; Horst, 1965) and all that might be needed for their application here is a generalization of the rotation criterion from that of best agreement between *two* sets of factors, to the most continuous agreement of factors over the entire series of occasions ("most continuous" defined as in the continuity criterion in expression 6).

### Extending the Application of the PP Model

The distinction developed in this section between system-variation and object-variation would seem to rule out applying the PP model to many interesting types of data sets, for example personality test data. The same apparent restriction confronted Cattell in his development of the original proportional profile rotation criterion. One of his primary interests was the application to personality test data, and so he devised ways in which one could deliberately impose the required systematic variation on what would otherwise be object-variation type data (Cattell, 1944, p. 275). Though not all of these techniques apply to our generalized model, several of his suggestions are quite relevant. The basic idea is simply to alter the conditions under which measurements are made on the successive occasions in such a way that the relative influence of different factors will change for all individuals. Administering a test under speeded vs. unspeeded conditions, under different "sets", using different instructions, etc. might all be used to impose overall enhancement or suppression of particular factors. For further discussion the reader is referred to Cattell (1944, p. 275).

#### IV. TESTING THE THREE-MODE PROPORTIONAL PROFILES MODEL

The PP model has been tested in two basic ways: (1) by using it to analyze synthetic data of known factor composition; (2) by applying it to real data. The tests with synthetic data helped to determine the behavior of the model and of the analysis technique with different size data sets and different latent factor structures. The conditions in which uniqueness breaks down and the minimal requirements of an "adequate" data set were determined in a trial and error fashion. This knowledge then guided the applications of the model to real data.

##### Synthesis of the Data

In order to judge whether or not PP three-mode factor analysis can correctly extract the true latent structure of a set of data, it is of course necessary to know in advance what that latent structure is. To this end, a FORTRAN IV computer program was written to generate synthetic data. Input to the program was a description of the factors and the dimensions of the desired data set, output was the set of data generated by those factors. Exact error free data was created, since this simplified the recognition of optimal fit and allowed the initial tests to concentrate on those aspects of the solutions determined by the data structure rather than by error variations. Specification of the factors used to generate a given set of data consisted simply of specifying the loadings of those factors on each measure, on each individual or object, and on each occasion.

The term "loading" as applied to the PP model, should not be taken to mean more than simply a weighting coefficient, describing the variations of relative influence of a given factor from one occasion to the next (or one person or variable to the next). This use is a generalization of the term "loading" as it normally occurs in factor analysis. Since the analysis procedure developed below completely bypasses the computation of correlation coefficients, either among components of the data or among data and factors, the mathematical properties of the "loadings" developed by the PP analysis are not the same as those of conventional loadings. This is particularly significant since (as Cattell discovered) strict proportionality does not hold across conventional loadings when a factor increases its relative influence (Cattell, 1955, p. 86). In contrast, the "loadings" used in the PP analysis do maintain proportionality as a factor changes its relative influence. These "loadings" are simply the weights which would occur in equations, such as 8, predicting the raw data from the factors.

A set of factor loadings input to the program generating synthetic data is given in Table 1, and the resulting data in Table 2. The data was synthesized from the input loadings as follows:

TABLE 1

## Typical Loadings Used to Generate Synthetic Data

<u>Occasion Loadings</u>				
	Factor 1	Factor 2	Factor 3	Factor 4
1	1.000	1.000	1.000	1.000
2	1.500	2.150	0.0	0.100
3	1.000	0.350	0.300	0.400
4	0.500	0.900	1.200	0.300
5	0.900	0.500	1.300	1.100
6	1.000	1.100	2.800	2.300
7	1.100	1.000	0.200	1.800
8	0.0	1.000	1.100	0.500
9	0.900	0.0	0.600	0.700
10	2.100	2.000	1.500	1.800

<u>Person Loadings</u>				
	Factor 1	Factor 2	Factor 3	Factor 4
1	1.000	0.600	1.000	2.000
2	1.100	0.700	1.400	0.600
3	0.900	0.800	1.300	0.700
4	1.200	0.900	1.200	0.800
5	0.800	1.000	1.100	0.900
6	1.300	1.100	1.000	2.000
7	0.700	1.200	0.900	0.900
8	1.400	1.300	0.800	0.800
9	0.600	1.400	0.700	0.700
10	1.000	1.000	0.600	0.600

<u>Test Loadings</u>				
	Factor 1	Factor 2	Factor 3	Factor 4
1	-1.100	0.010	0.120	0.110
2	1.000	0.030	0.210	0.070
3	4.000	0.030	0.120	0.030
4	0.190	3.000	0.320	0.600
5	0.230	7.190	0.500	0.110
6	0.010	0.740	5.250	0.130
7	0.500	1.500	9.500	1.500
8	-0.900	0.200	2.100	0.300
9	0.010	0.100	0.030	9.990
10	0.110	-0.200	-0.800	7.640

A typical set of factor loadings used to generate synthetic data. Note the absence of orthogonality and simple structure.

TABLE 2

## Data Computed for Occasion 1

1	-0.754	1.368	4.198	3.510	5.264	5.964	13.900	1.920	20.080	14.470
2	-0.969	1.457	4.607	3.117	6.052	7.957	15.800	2.270	6.117	3.445
3	-0.749	1.246	3.801	3.407	6.686	7.517	15.050	2.290	7.121	4.247
4	-1.079	1.535	4.995	3.792	7.435	7.082	14.550	1.860	8.130	5.104
5	-0.639	1.124	3.389	4.044	8.023	6.640	13.700	2.060	9.132	5.884
6	-1.079	1.683	5.413	5.067	8.928	6.337	14.800	1.750	20.133	14.403
7	-0.551	0.988	2.971	4.561	9.338	5.737	12.050	1.770	9.145	5.993
8	-1.343	1.663	5.759	4.902	10.157	5.280	11.450	0.920	8.160	5.366
9	-0.485	0.838	2.547	4.958	10.631	4.808	10.100	1.420	7.160	4.574
10	-0.952	1.198	4.120	3.742	7.786	3.978	8.600	0.740	6.122	4.014

## Data Computed for Occasion 2

1	-1.622	1.532	6.024	2.205	4.681	0.485	1.950	-1.170	2.073	1.573
2	-1.801	1.675	6.623	2.449	5.419	0.542	1.965	-1.327	0.686	0.500
3	-1.469	1.379	5.426	2.698	6.070	0.615	1.980	-1.034	0.793	0.523
4	-1.962	1.833	7.229	3.090	6.894	0.694	2.370	-1.416	0.907	0.629
5	-1.300	1.236	4.833	3.282	7.476	0.764	2.235	-0.853	1.011	0.620
6	-2.112	1.997	7.839	3.790	8.379	0.859	2.925	-1.475	2.127	1.522
7	-1.133	1.092	4.239	3.853	8.879	0.910	2.460	-0.678	1.030	0.563
8	-2.288	2.145	8.441	4.347	9.839	0.993	3.120	-1.606	0.950	0.582
9	-0.968	0.947	3.644	4.413	10.281	1.054	2.655	-0.509	0.848	0.354
10	-1.633	1.534	6.032	3.321	7.542	0.763	2.340	-1.132	0.714	0.423

.

.

.

## Data Computed for Occasion 9

1	-0.764	1.124	3.714	1.203	0.661	3.341	8.250	0.870	14.013	10.315
2	-0.942	1.196	4.073	0.709	0.694	4.474	9.105	0.999	4.231	2.646
3	-0.743	1.008	3.348	0.697	0.630	4.167	8.550	1.056	4.927	3.209
4	-1.040	1.270	4.423	0.772	0.670	3.864	8.220	0.708	5.627	3.821
5	-0.643	0.903	2.978	0.726	0.565	3.554	7.575	0.927	6.321	4.364
6	-1.061	1.394	4.794	1.254	0.723	3.344	8.385	0.627	14.016	10.345
7	-0.559	0.787	2.604	0.670	0.484	2.923	6.390	0.756	6.316	4.450
8	-1.267	1.400	5.114	0.729	0.591	2.605	6.030	0.042	5.621	4.033
9	-0.490	0.662	2.225	0.531	0.388	2.274	4.995	0.543	4.913	3.467
10	-0.901	1.005	3.656	0.538	0.433	1.954	4.500	0.072	4.216	3.020

## Data Computed for Occasion 10

1	-1.722	2.703	8.724	6.639	10.257	9.252	22.500	2.580	36.150	26.295
2	-2.156	2.869	9.566	5.959	11.766	12.224	24.825	2.935	11.015	6.545
3	-1.690	2.436	7.880	6.539	13.052	11.604	23.760	3.092	12.825	7.954
4	-2.380	3.053	10.393	7.319	14.580	10.994	23.220	2.304	14.645	9.479
5	-1.452	2.200	7.027	7.819	15.770	10.370	21.945	2.839	16.450	10.842
6	-2.405	3.363	11.274	9.759	17.592	9.998	24.315	2.213	36.256	26.164
7	-1.253	1.939	6.163	8.883	18.447	9.089	19.590	2.478	16.479	10.978
8	-2.906	3.371	12.025	9.607	20.129	8.441	18.930	0.826	14.711	9.845
9	-1.093	1.653	5.288	9.731	21.085	7.761	16.695	2.009	12.911	8.365
10	-2.063	2.425	8.600	7.335	15.432	6.366	14.220	0.724	11.037	7.362

Data synthesized with the latent structure defined by Table 1.

$$(8) \quad x_{kji} = O_{k1} I_{j1} M_{i1} + O_{k2} I_{j2} M_{i2} + \dots + O_{km} I_{jm} M_{im}$$

where the "O" values represent weights for the different occasions, the "I" values represent weights for the different individuals, and the "M" values represent weights for the different measures.

A relativity of scale is apparent in the PP model. Consider equation 8, which generates data according to that model. If all of the "O" values for any factor are doubled, and all the "I" values are halved, the resulting data is unchanged. In general, any factor could have all of its loadings for one mode multiplied by some constant, and so long as all its loadings on one of the other modes were divided by the same constant, the results would be equivalent. Such an operation changes the factor's absolute scale in any given mode, but it does not alter the relationships among the loadings within any mode. Therefore, all these changes of scale yield descriptions of the factors which are completely equivalent with respect to interpretation.

It was decided to adopt a scaling convention so that solutions with the PP model would not be subject to this trivial type of indeterminacy. For this purpose, the average value of the influence of any factor, across all "occasions", was taken to be 1.0. Similarly, the average amount of influence of any factor, taken across all "individuals", was also set to 1.0. Thus a loading of 2.0 on either of these two modes meant that a factor was twice as active as average on that occasion or in that individual, 0.5 meant half as active as average, etc. Loadings on the third mode, measures, would therefore give directly the average contribution of that factor to the predicted score of that measure. The input specifications of factors for the data synthesis program (Table 1) were selected to follow these conventions.

#### PARAFAC: The Computational Technique for Data Analysis

The PP three-mode factor model (2) is sufficiently different from the classical factor model (1) to make conventional factor analysis algorithms inapplicable. An analysis procedure called PARAFAC (for *parallel factors*) was developed to accomplish the required analysis. One objective of this work has been to develop a technique which is general enough to fit both the PP model and a family of related models to real data. For this reason, an approach was adopted which involved direct optimization of the parameters of the model. As Green points (Green, 1966, p. 440), "Nearly all psychometric procedures can be looked upon as attempts to optimize some criterion ... Factor analysis seeks the minimum number of parameters to describe the maximum amount of inter-correlation among the variables. Whenever one is fitting a model to data, one seeks parameters of the model that fit the data as closely as possible -- parameters that optimize some measure of fit." In this case, the PARAFAC analysis procedure tries to fit the mathematical model of (2) to the data as closely as possible, using the least-squares criterion of fit. This is similar to the approach recently



developed by Harman and Jones (1966) for conventional factor analysis models. The  $U$  term of equation 2 is here taken as equivalent to the error of estimate, i.e.  $(x - \hat{x})$ . In a sense, then, it "drops out" of the model. The equation used for computation of the estimated data values was

$$(9) \quad \hat{x}_{kji} = \sum_m O_{km} I_{jm} M_{im}$$

and it was required that the parameters (the  $I$ ,  $O$ , and  $M$  values) minimize the sum of the  $(x - \hat{x})^2$  residuals. Unlike Harman and Jones (1966), this technique estimates three sets of parameters (or factor "loadings"), rather than just one. As a result, a large number of parameters have to be "fit" simultaneously to the data. For a matrix of 10 measures, on 10 individuals, on 10 occasions, there are 1000 data points, and for each factor fit to such a data set there are 30 loadings to be estimated (10 weights for occasions, 10 weights on individuals, and 10 weights on measures). For four factors, a total of 120 parameters have to be estimated simultaneously. Such large problems can bog down conventional optimization techniques based on gradients.

Fortunately, a powerful optimization routine was made available for this work by Hans Reichenbach. The program is called Routine TBU, and is thoroughly described in a 400+ page manual and theoretical article (Reichenbach, 1969) of which only an earlier and more primitive version has been published (Reichenbach, 1962). A common cause of convergence difficulties with gradient techniques occurs when the routine becomes trapped on a "thin ridge" in its upward path on the multi-dimensional hypersurface. In order to avoid this difficulty, Routine TBU uses trends in the changes of the variables rather than gradients. A complete iteration of this routine contains a section where each variable is changed individually as a function of its current short-range and long-range trends, and another section where all variables are changed simultaneously according to some optimal proportion of their trends. The routine alternates between these two methods of changing variables. Both Harman and Jones (1966) and Joreskog (1966) adopted modifications of the steepest descent technique to improve convergence, and these modifications are related in different ways to features of TBU. The reader interested in further details on TBU is referred to Reichenbach (1962, 1969).

TBU proved successful in optimizing 30-120 variable models involving one to four factors, but convergence became markedly slower as the number of factors increased. On the IBM 360/91 computer, one factor could be fitted to a 10 by 10 by 10 data set in a few seconds. Two factors would often require twenty seconds, three factors one hundred seconds, and four factors several hundred seconds or more, depending on the characteristics of the data set. Beyond four factors, convergence took too long to be practical. This working range of 1-4 factors provided sufficient leeway to explore many of the theoretical properties of the model, such as the

conditions for a unique solution and the effects of extracting more or fewer factors than were used to create the data. But for practical application and for more general theoretical studies, an improvement of the algorithm was clearly necessary.

Dr. Robert Jennrich of the UCLA Department of Biomathematics proposed a special purpose computation algorithm for the solution of the *linear* PP three-mode model which proved to be much simpler and faster than the generalized approach which was being used. His development of this algorithm was as follows:

Given the representation

$$(10) \quad x_{ijk} = \sum_l O_{il} I_{jl} M_{kl} + E_{ijk}$$

we seek values of  $O_{il}$ ,  $I_{jl}$ , and  $M_{kl}$  which minimize the sum of squared errors  $\sum_{ijk} E_{ijk}^2$ . For fixed values of  $O_{il}$  and  $I_{jl}$  the selection of values  $M_{kl}$  which minimize this sum is a simple multiple linear regression problem whose solution is:

$$(11) \quad M_{kl} = \sum_m x^{lm} y_{mk}$$

where

$$(12) \quad x_{lm} = \left( \sum_i O_{il} O_{im} \right) \left( \sum_j I_{jl} I_{jm} \right) ,$$

$$(13) \quad y_{lk} = \sum_{ij} O_{il} I_{jl} x_{ijk} ,$$

and the matrix  $(x^{lm})$  is the inverse of the matrix  $(x_{lm})$ . Similar formulas give  $I_{jl}$  in terms of  $O_{il}$  and  $M_{kl}$  and  $O_{il}$  in terms of  $I_{jl}$  and  $M_{kl}$ . A complete algorithm consists in solving for the matrices  $(O_{il})$ ,  $(I_{jl})$  and  $(M_{kl})$  in order and repeating until convergence is obtained.

This algorithm has been found to converge in less than ten seconds for many data sets involving four to six factors. But in other cases (particularly certain types of data sets to be described below) convergence would proceed only to a certain point, beyond which a large number of iterations would produce little improvement of fit to the data.

To cope with this difficulty, a relaxation factor was added to the computation procedure. Examination of the behavior of the factor loadings after convergence slows down reveals that they continue to move in small, steady steps, usually in the direction of the desired unique solution.

In order to increase the size of these steps, a modification is made in the computation given in equation 11. An adjusted estimate  $M^*$  is computed for each factor loading according to the relation

$$(14) \quad M^*_{kZ} = M'_{kZ} + R(M_{kZ} - M'_{kZ})$$

where  $M_{kZ}$  is the value for the factor loading which would be computed for this step according to the unmodified procedure of equations 10 through 13, and  $M'_{kZ}$  is the old value for the loading computed during the last iteration.  $R$ , the relaxation factor, is assigned a value between 1.0 and 2.0.

The effect of this modification is to push each loading further in the direction that it has been moving, beyond the value which is optimum simply for this cycle, and hopefully towards the final overall optimum value. Although a relaxation factor seems to provide a small gain, since it increases any step size by less than 2.0, under the proper conditions it has a "snowballing" effect, since the resulting adjusted loadings are used as the basis for the unadjusted values computed in the next cycle, and the values of that cycle are then themselves adjusted, thus moving even further in the desired direction. Thus the effect of a small relaxation factor can continue to multiply itself on successive iterations.

Modifications to this algorithm are still being explored, and the results are as yet tentative and incomplete. The effectiveness of the procedure depends critically on both the size of the relaxation factor and the characteristics of the data set. To date, the best overall value for the relaxation factor seems to be in the range 1.2-1.3. Using this modified algorithm, it often becomes possible to push the residuals to very low limits until a point is reached where the factor loadings themselves cease to change. The ability to push analysis much closer to true convergence and a perfect fit to the data has led to the discovery of unique solutions where they were previously not believed to exist. This has forced a reevaluation of the minimal necessary conditions for uniqueness (to be discussed below). Therefore, even though these new algorithms are still in the process of development and only incomplete results are available, these results are significant enough to warrant including them in the discussion which follows.

#### The Analysis of Synthetic Data

Approximately 30 tests were performed by analyzing synthetic data using the optimization technique of Routine TBU. In these tests the number of factors in the data was varied from two to four, and the number of factors extracted ("fit" to the data) was varied from 1 to 5, ranging both below and above the "true" number of factors used to construct the data. The size of the data set was varied systematically, the number

of intervals in any mode ranging from 1 to 10. The pattern of factor loadings was initially all positive, but negative and zero loadings were subsequently introduced to ensure generality. As can be seen from inspection of a typical example given in Table 1, the factors were not generally orthogonal and did not display simple structure. They were selected to conform to the scale convention discussed earlier, and were varied so that a row of factor loadings would not repeat the relations of any other row in that mode.

After the development of the quick algorithm for the linear case, approximately 50 additional tests were performed with synthetic data. In these experiments the number of factors in the data ranged as high as 10, and the dimensions of the data set were expanded to a maximum of 20 by 20 by 20. In most of these experiments the factor loadings were generated from a sequence of uniform pseudo-random numbers\* between 0.0 and 2.0 (before normalization). Since this left the factors in general uncorrelated, specific tests were performed with correlated factors. These factors were constructed by adding a percentage of each loading of a given factor to the loading of the factor with which it was to be correlated.

Because error free data was used, classification of the resulting solutions was usually clear cut. Complete fit was recognizable by near-zero residuals (from 1% to .01%, depending on the amount of computer time used). A solution matched the "true" latent structure if its factors had the same loadings within 0.1% (or, when convergence was carried further, to within a few hundredths of a percent). The columnar order of factors might differ from that used on input to create the data, but factors were easily "identified" regardless of their column. A "match" was not expected to exhibit the same columnar order as the factors used to create the data.

Initially, the test of uniqueness was the consistent finding of a matching solution regardless of the starting values for the loadings. This required repeating each test from several different random starting configurations. This condition was eventually dropped and the discovery of a perfect matching solution with one starting configuration was taken to demonstrate that this match would be found with any starting configuration (since the likelihood of stumbling on the one perfect match by chance when the solution was not unique seemed quite small, and earlier results showed that when any solution matched, all solutions with that data and that number of factors would match, regardless of the starting configuration). Non-unique solutions were always found to clearly diverge from the "true" loadings.

---

\* This was accomplished by using the RANDM function included as an addition to the FORTRAN IV level G compiler by the UCLA Campus Computing Network.

Since decisions of uniqueness and good fit were non-controversial, individual sets of "true" and "reproduced" loadings will not, in general, be reproduced here. Instead, results from the 80 experiments with synthetic data will be summarized according to those conditions which were found to affect the explanatory validity of the solutions. Three things should be considered in this regard: (1) the effect on the solution of the number of factors chosen to be extracted from a given data set; (2) the data characteristics necessary to adequately determine the correct solution for a given latent factor structure; (3) the data characteristics which affect the ability of the analysis procedure to reach convergence, and the properties of different types of solutions prior to convergence.

#### A. Effects of extracting different numbers of factors

Before performing any analysis, one must decide how many factors to extract from the data, i.e. how many factors to specify in the model which is to be fit to the data. An understanding of the consequences of extracting the correct number of factors, versus too few or too many factors, is essential for interpretation of the solutions obtained. On the basis of such information one can then revise the estimate of the number of factors present, and reanalyze the data.

(1) If the number of factors extracted from the data is the same as the number of factors actually present (i.e. used on input to create the data) then the solution is unique and "true" in that it exactly discovers the latent factor structure built into the data. (This uniqueness holds, however, only if the data set is "adequate". The conditions of data adequacy will be discussed below.) The factor loadings need not fall into any simple structure or other natural pattern. The factors may be orthogonal or oblique. An example of a set of factors which are neither orthogonal nor displaying simple structure, yet which were correctly recovered from both a 1000 and a 64 point data set is given in Table 1. (The 64 point solution was for the first 4 intervals in each mode.)

(2) If fewer factors are extracted than are present in the data, and the data is adequate, then the solution is unique, but (of course) incorrect. The resemblance between the recovered solution and the true factor pattern will depend on how "wrong" the estimate of the number of factors was. In a limiting case, if the true number of factors was only 1 larger than the number extracted, and one of the factors used to create the data made only small contributions to the data values, then the resulting solution will closely approximate the true solution except that the small factor will be omitted. In less ideal cases, the extracted factor structure will be proportionately more distorted, as it tries to account for the contributions of the additional (missing) factors within the fewer factors allowed on analysis. Examples for three and four factors are reproduced in Tables 3A and 3B.

(3) If more factors are extracted than are present in the data (i.e. used to create the data), the solution which is extracted is not

TABLE 3A

## Distortions Due to Extracting Too Few Factors

## Occasion Loadings (Input)

	Factor 4	Factor 3	Factor 1	Factor 2
1	1.724	1.315	1.020	1.020
2	0.172	0.0	1.531	2.194
3	0.690	0.395	1.020	0.357
4	0.517	1.579	0.510	0.918
5	1.897	1.710	0.918	0.510

## Person Loadings (Input)

	Factor 4	Factor 3	Factor 1	Factor 2
1	1.620	2.258	0.980	1.078
2	1.268	0.161	1.078	0.980
3	0.352	0.887	0.0	0.980
4	0.493	0.484	0.882	0.0
5	1.268	1.210	2.059	1.961

## Test Loadings (Input)

	Factor 4	Factor 3	Factor 1	Factor 2
1	1.647	0.942	1.000	0.600
2	0.494	1.319	1.100	0.700
3	0.577	1.225	0.900	0.800
4	0.659	1.131	1.200	0.900
5	0.741	1.037	0.800	1.000

\* \* \*

## Occasion Loadings (Extracted)

	Factor 1	Factor 2	Factor 3
1	2.068	1.424	0.961
2	0.054	-0.340	2.048
3	0.234	0.124	0.720
4	0.335	1.869	0.668
5	2.309	1.922	0.603

## Person Loadings (Extracted)

	Factor 1	Factor 2	Factor 3
1	1.621	3.424	1.398
2	1.268	-0.618	0.906
3	0.176	0.731	-0.079
4	0.575	-0.147	-0.285
5	1.359	1.610	3.061

## Test Loadings (Extracted)

	Factor 1	Factor 2	Factor 3
1	1.285	0.409	0.739
2	0.518	0.549	0.800
3	0.504	0.521	0.763
4	0.632	0.462	0.993
5	0.561	0.447	0.849

Three factors extracted from data generated by four input factors. Many of the loadings are similar despite the too few factors extracted. Note, for example, the correspondence between the input loadings for factor 4, and the extracted loadings for factor 1 (factors are labeled according to their position in the extracted matrix, but their columnar position has been rearranged to facilitate comparison with the input factors).

TABLE 3B

## Distortions Due to Extracting Too Few Factors

Occasion Loadings (Input)

	Factor 3	Factor 2	Factor 1
1	1.316	1.020	1.020
2	0.0	2.194	1.531
3	0.395	0.357	1.020
4	1.579	0.918	0.510
5	1.710	0.510	0.918

Person Loadings (Input)

	Factor 3	Factor 2	Factor 1
1	2.258	1.078	0.980
2	0.161	0.980	1.078
3	0.887	0.980	0.0
4	0.484	0.0	0.882
5	1.210	1.961	2.059

Test Loadings (Input)

	Factor 3	Factor 2	Factor 1
1	0.942	0.600	1.000
2	1.319	0.700	1.100
3	1.225	0.800	0.900
4	1.131	0.900	1.200
5	1.037	1.000	0.800

\* \* \*

Occasion Loadings (Extracted)

	Factor 1	Factor 2
1	1.509	1.050
2	-0.306	2.155
3	-0.103	0.580
4	1.876	0.551
5	2.023	0.664

Person Loadings (Extracted)

	Factor 1	Factor 2
1	3.924	1.511
2	-0.927	0.808
3	0.425	-0.592
4	-0.314	-0.863
5	1.891	4.136

Test Loadings (Extracted)

	Factor 1	Factor 2
1	0.287	0.417
2	0.466	0.525
3	0.419	0.486
4	0.391	0.646
5	0.337	0.527

Two factors extracted from data created by three factors. The extracted factors show less correspondence to the input factors than in Table 3a.

TABLE 3C

## Non-uniqueness Due to Extracting Too Many Factors

## Occasion Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	0.306	0.066	1.700	1.772
2	1.835	2.103	0.142	2.048
3	0.143	0.994	0.622	1.100
4	0.572	0.281	2.453	0.066
5	2.144	1.557	0.084	0.015

## Person Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	1.817	0.354	1.187	0.988
2	0.629	0.741	1.110	0.854
3	0.791	1.255	0.223	1.235
4	0.996	0.867	1.247	0.969
5	0.766	1.783	1.233	0.954

## Test Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	0.368	0.677	0.556	1.623
2	0.345	0.312	1.751	2.033
3	0.309	0.382	1.746	0.277
4	0.294	0.408	0.346	2.186
5	0.013	0.853	1.378	2.559

\* \* \*

## Occasion Loadings (Extracted)

	Factor 5	Factor 3	Factor 4	Factor 1	Factor 2
1	0.306	0.065	1.663	1.781	1.738
2	1.835	2.103	-0.824	2.291	1.149
3	0.143	0.995	0.380	1.161	0.874
4	0.572	0.280	3.664	-0.239	1.192
5	2.144	1.557	0.118	0.007	0.048

## Person Loadings (Extracted)

	Factor 5	Factor 3	Factor 4	Factor 1	Factor 2
1	1.817	0.354	1.338	0.969	1.070
2	0.629	0.741	1.304	0.830	0.959
3	0.791	1.255	-0.544	1.333	0.819
4	0.996	0.867	1.458	0.942	1.083
5	0.766	1.783	1.445	0.927	1.069

## Test Loadings (Extracted)

	Factor 5	Factor 3	Factor 4	Factor 1	Factor 2
1	0.368	0.676	0.129	1.010	1.040
2	0.345	0.311	0.480	0.885	2.419
3	0.309	0.382	0.520	-0.423	1.925
4	0.294	0.408	0.052	1.505	0.974
5	0.013	0.852	0.355	1.412	2.170

Five factors extracted from data created by four factors. Since the ratio of factors extracted to the true number input is not much greater than one, some factors (in this solution factors 1 and 2) were exactly recovered. Other extracted factors show various degrees of resemblance to input factors.



unique. Its resemblance to the true solution depends on the ratio of the number of factors extracted to the number of factors actually present. If this ratio is only slightly greater than one, then some extracted factors will match true factors, although which of the true factors will be recovered will depend on the starting configuration for a given analysis. Given the following ratios of the number extracted to the true number of factors, the number of matching factors might typically be 2 for  $5/4$ , 3 for  $6/5$ , 0 for  $3/2$ , and 0 for  $6/3$ . The non-matching factors will usually include both some which bear a rough resemblance to true factors and some which are completely different from any of the true factors. An example displaying matching and both kinds of non-matching factors is given in Table 3C.

Because of the characteristics of PP analysis just described, the investigator has two interlocking criteria for determining the correct number of factors to extract from adequate data: (1) the largest number of factors that still gives a unique solution; and (2) the traditional criterion of the largest number of factors which still significantly improves "fit".

#### B. Conditions of an adequate data set

According to the fundamental theory of PP factor analysis (Sections II and III, above) a unique solution requires data for a number of measures on a number of persons collected on at least two occasions, with system-variation of the factors present across these two occasions. This basic requirement was confirmed in experiments with synthetic data. Data sets containing only one occasion of measurement (in effect, two-way data sets) failed to yield unique solutions when factors were fitted to them by TBU or the quick algorithm. Similarly, three-way data sets in which all the factors had constant loadings across the occasions (i.e. with no system-variation) did not yield unique solutions regardless of the number of occasions.

On the other hand, it was possible for all factors to be uniquely determined and correctly recovered if their columns of loadings on the different occasions were distinct. (For two columns to be distinct, it is apparently sufficient that they be linearly independent.) Thus, even when one factor did not change its influence across occasions, it could be correctly recovered in the analysis so long as no other factor was similarly constant.

If, in a given data set, only some factors had occasion loading patterns distinct from all other factors (e.g. see Table 4A), then only the distinct factors would be uniquely determined in the solution. For example, a data set was created in which three of the four latent factors had the same increasing pattern of loadings across occasions. Analysis of this data gave non-unique solutions for three of the four factors extracted. Further, the one factor which was uniquely determined in the otherwise different solutions (as evidenced by its occurrence in

the same form in each solution) was found to be an exact match of the one factor which had a loading pattern across occasions which was distinct from the other three factors. This type of result is encouraging, since it indicates that the conditions underlying the data need not be completely adequate in order to recover some valid factors. When there is inadequacy for some factors, the solution does not collapse as a whole, but rather one factor at a time. Any factors which do show the necessary distinctness across occasions of measurement can be uniquely recovered in the solution.

Two features of the non-unique solutions obtained with partially or completely inadequate data sets serve to distinguish them from the non-unique solutions obtained by too many factors. First, even when the ratio of estimated to true number of factors is small enough that some factors will be correctly recovered, it is not always the *same* factors which are recovered when different starting configurations are used. With partially inadequate data, however, the same factors will always be uniquely determined and recovered in the solution, regardless of the starting configuration used. Secondly, in these experiments the inadequate *occasion* loading patterns still seemed to be correctly recovered, even though the loadings for the other modes could not be correctly recovered. Thus the occasion loading patterns always displayed that lack of distinctness between the non-unique factors which was the cause of their non-uniqueness. By contrast, when too many factors are extracted, the loading patterns are not uniquely determined for any mode except for those factors correctly recovered as a whole. Considerations such as these provide important guides to interpretation of solutions obtained with real data. An example of loadings used to generate a partially inadequate data set and one solution obtained from data so generated is given in Table 4A.

Although distinct loading patterns across two occasions might be theoretically sufficient to uniquely determine any number of factors, this sufficiency could not be established in these experiments because of practical convergence difficulties. These difficulties made it impossible to solve for more than 4 or 5 factors using a data set with only two occasions. Two occasions were never found *insufficient*, since convergence on a non-unique solution was never obtained with otherwise adequate data when there were only 2 occasions. However, since convergence could not be obtained with two-occasion data sets which had a large number of latent factors, the adequacy of a two-occasion data set for uniquely determining these large numbers of factors could not be empirically confirmed.

Apparently, when the number of occasions is very small, specifically two or three, the properties of the hypersurface where the routine is seeking an optimum become such that the upward paths are narrow and twisted and convergence is difficult or impossible within a few thousand iterations. Before the development of the quick algorithm, these problems led to the conclusion that there must be at

TABLE 4A

## Factors Extracted from "Partially Inadequate" Data

Occasion Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	0.333	0.333	0.333	1.772
2	0.667	0.667	0.667	2.048
3	1.000	1.000	1.000	1.100
4	1.333	1.333	1.333	0.066
5	1.667	1.667	1.667	0.015

Person Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	1.817	0.354	1.187	0.988
2	0.629	0.741	1.110	0.854
3	0.791	1.255	0.223	1.235
4	0.996	0.867	1.247	0.969
5	0.766	1.783	1.233	0.954

Test Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	4.478	4.478	2.515	1.623
2	4.204	2.062	7.915	2.033
3	3.761	2.529	7.892	0.277
4	3.574	2.703	1.565	2.186
5	0.157	5.642	6.228	2.559

\* \* \*

Occasion Loadings (Extracted)

	Factor 1	Factor 3	Factor 4	Factor 2
1	0.333	0.333	0.333	1.772
2	0.667	0.667	0.667	2.048
3	1.000	1.000	1.000	1.100
4	1.333	1.333	1.333	0.066
5	1.667	1.667	1.667	0.015

Person Loadings (Extracted)

	Factor 1	Factor 3	Factor 4	Factor 2
1	1.043	5.207	3.931	0.988
2	0.861	0.877	1.704	0.854
3	0.731	-1.183	-2.072	1.235
4	1.050	1.654	2.058	0.969
5	1.316	-1.556	-0.621	0.954

Test Loadings (Extracted)

	Factor 1	Factor 3	Factor 4	Factor 2
1	11.641	0.971	-1.141	1.623
2	13.192	0.095	0.893	2.033
3	13.337	-0.101	0.946	0.277
4	7.851	0.910	-0.920	2.186
5	12.401	-1.399	1.025	2.559

Loadings used to generate "inadequate" data, and one resulting solution extracted from such data. The latent structure used on input was "inadequate" to distinguish factors 1, 2, and 3, since they had the same pattern of loadings across occasions. A typical solution extracted from this data set shows that only factor 4 was correctly recovered.

least as many intervals in each mode of an adequate data set as there are factors being extracted from the data set. For small numbers of factors this still seems to be a good rule. For large numbers of factors, 4 to 8 occasions might provide a good base for reasonably quick convergence, depending on the particular characteristics of the data set. (To date, experiments have been limited to a maximum of 10 factors because of trivial array size limitations, but the program is being modified for larger arrays in future experiments.) Table 4B gives some approximate convergence times for data sets of different sizes and containing differing numbers of factors. It should be noted that a number of other characteristics of the data set will affect its convergence properties, as will be discussed below. Therefore, the iteration counts listed in Table 4B are only indicative of reasonable but not necessarily average or minimal values. They all reflect values obtained with the quick algorithm using a relaxation factor of 1.275, starting from a random configuration of loadings.

The discussion of data adequacy has focused, up to this point, on the conditions of only one mode. For the purposes of discussion this mode has been called "occasions", although the results would apply by symmetry to any single mode of a three mode matrix. The assumption of the adequacy of the other two modes has been implicit in the discussion of the effect of changes in the "occasions" mode. The experiments listed in Table 4B (with the exception of the last one listed) were consistent with this assumption. They each had at least as many intervals in the second and third mode as there were factors underlying the data. Further, in these two modes the pattern of loadings was distinct for each factor (i.e. was always linearly independent of the loading patterns for all other factors).

The case of a single inadequate mode was studied most intensively both because it was a simplified approach to the general question of data adequacy, and because it was thought to represent the most common way in which data would be found to be inadequate. When there are difficulties with obtaining adequate data, these difficulties will frequently involve only one of the three modes. For example, it is sometimes easy to secure many measurements, and many individuals to measure, but more difficult to establish more than perhaps two or three occasions or conditions of measurement where there will be systematic differences in relative factor influence.

A few experiments have been performed, however, where more than one mode was restricted (e.g. had fewer intervals than there were factors to be extracted). The results dramatically demonstrate the differences between the theoretical capabilities and properties of classical two-mode and the new three-mode PP factor analysis. In one experiment, 10 factors were correctly recovered from an 8 by 8 by 8 data matrix (this experiment corresponds to the last entry of Table 4B). Clearly, the mathematical rank of any two-way slice of this data matrix cannot be greater

TABLE 4B

Computation Times for Some Data Sets Having  
Fewer Occasions Than Factors

	Number of Factors	Dimensions of Data Set	Number of Iterations to Convergence
a.	3	2 x 4 x 5	1000
b.	4	2 x 6 x 7	1500
c.	6	2 x 7 x 8	not close to conver- gence after 2000 iterations
d.	4	3 x 8 x 10	400
e.	4	3 x 8 x 10	300
f.	5	3 x 10 x 10	600
g.	6	3 x 11 x 10	500
h.	6	4 x 8 x 10	300
i.	10	8 x 8 x 8	800

Convergence times for 9 selected experiments. The number of iterations required to reach convergence increases with fewer intervals in the smallest mode of the data matrix, particularly as the number of factors becomes larger.

than 8, and classical two-mode factor analysis of any such slice could only reveal at most 8 factors. Yet when all the "slices" are assembled into a three-way matrix and subjected to three-mode PP factor analysis, 10 factors are correctly and uniquely recovered.

The complete set of rules defining the *minimal* conditions of an adequate data set have not yet been determined. Non-minimal *sufficient* conditions can be described with some assurance, however. It appears that in order to uniquely recover N factors, it is sufficient that the data set have N different intervals in each mode (except where this would provide fewer data points than factor loadings, as in the case of 2 factors). This empirically derived rule of sufficiency will be supported by a mathematical proof in Section V. In addition, some weaker sufficiency rules have been described above for certain selected cases, and a type of sufficiency for the correct recovery of a given factor from an otherwise inadequate data set has also been defined. These results will have direct application to the analysis of real data, but further mathematical and experimental work needs to be done in order to rigorously define the minimal necessary conditions of an adequate data set, as a function of the number of latent factors present.

### C. Conditions affecting convergence

The ratio of the number of factors to the number of occasions (or to the number of intervals in whichever is the smallest of the three modes) is one characteristic of the data set which strongly affects convergence rate (as was discussed in B above). This points out the practical necessity of the generalization of the PP criterion from two to any number of "parallel" occasions. Additional conditions which affect convergence include: the presence of positively correlated factors; the presence of factors with highly unequal contributions to the data; the number of near zero loadings in the latent structure; non-uniqueness of the solution; and the starting configuration from which an attempt at optimization was begun.

If two of the latent factors have a strong positive correlation (e.g. +.75 or higher) the similarity of their loading patterns apparently begins to pose problems for the optimization routine. Faced with two similar but not identical loading patterns across a given mode, the PP analysis procedure usually reaches a rough fit and then struggles across a plateau where it tries to disentangle the influence of the two correlated factors. Convergence can be slowed in such circumstances as much as 4 to 10 times.

Similar difficulties arise when one or more factors make only very small contributions to the data values relative to the other factors. Here, too, the program first finds a rough fit to the data and then has to spend a great deal of time transforming this "first guess" into the often very different solution required for a final good fit. The rough "first guess" gives results similar to extracting too many factors, since the small factors haven't exerted their influence effectively.

Often values are very different from the correct final values, and if one stopped at this point the solution would appear non-unique from different starting configurations. After this "first guess" the program goes through a long series of successive corrections to achieve small improvements in fit, as it tries to respond to the feeble influence of the small factors. One possible way to handle such cases may be to first solve for fewer factors. The program should converge fairly quickly on a unique solution. The factors in this solution would resemble the larger factors, only slightly distorted because of the influence of the unrepresented small factors. The loading values for these large factors can be used as a starting point for the solution involving more factors. Starting values for the small factors can be columns of zeros, or any other good first estimate. From such a starting configuration, convergence should proceed much faster than it otherwise would in the case of very unequal factors.

If a number of zero loadings are present in the latent structure, convergence can be facilitated. Apparently this simplifies the path up the hypersurface by reducing the interdependencies among the variables. This effect has been most clearly noticed for Routine TBU.

Convergence is greatly speeded when the solution is not unique, either because of data inadequacy or because of the extraction of too many factors. This might have been expected, since it should be easier to climb to the optimum by finding and following one of the many paths that would end up on that hyperplane defined by an infinity of alternative optimum solutions, than it is to find and follow one of the fewer paths that would converge on a unique point. The "more non-unique" the solution (e.g. the more the number of factors extracted exceeds the true number of factors, and thus the more divergent the possible loadings in the alternative solutions), the more the convergence is accelerated. Thus convergence speed might provide an additional confirmatory guide to the choice of the number of factors.

With analyses having unique solutions, the particular starting configuration used was often found to have a very significant effect on convergence time, sometimes causing variations in convergence time as great as 500%. But with the error free synthetic data, there appeared to be no problems with local optima. Although both TBU and the quick algorithm sometimes experienced slowdowns where they would display steady but very gradual progress up the hypersurface, the values of the parameters would never stop changing completely unless the program had achieved a near perfect fit to the data.

In general, it seems that although difficulties are encountered with some types of data sets, the quick algorithm with the relaxation factor is adequate to allow analysis of most real data to which the PP model might apply. To cope with the remaining difficult cases, further development of the algorithm is being explored, in part by combining features of the quick algorithm with additional trend estimation techniques such as are found in Routine TBU.

## The Analysis of Real Data from Vowel Sounds

Although the experiments with synthetic data provide the only certain way to demonstrate the discovery of a true latent structure, they may not adequately simulate some other properties of real data which could affect performance of the analysis procedure. Therefore a set of real data was selected for analysis.

The data consists of a set of measurements of formant frequencies (converted into pitch values in mels) for eight vowel sounds, as uttered by 11 individuals. The resulting data was a 4 by 8 by 11 matrix of pitch values whose dimensions are formant, by vowel, by individual (Table 5). This data is a subset of a four-way matrix of formant by vowel by individual by occasion, originally published by Ladefoged (1967, pp. 88-89).

The theoretical interest in the data lies in the question of the nature of vowel quality. Phoneticians have been seeking an objective description of the features in the acoustic signal which are characteristic of the different vowels. It is important in this respect to distinguish between those differences in vowel quality that do not serve to identify the vowel (e.g. the different quality of different voices) and those characteristic qualities which do serve to identify the vowel. The second type of difference is of primary interest here.

A set of dimensions for vowels has been developed by phoneticians through consideration of the positions of the tongue and lips, and through evaluation of subjective sound quality. But the systematic classification of vowels on this basis has not been clearly related to the particular objective characteristics of the acoustic signal which (presumably) the brain uses in recognizing vowels in speech. It is, however, generally agreed that the most important acoustic features are certain characteristic peaks in the spectrum of vowel sounds (Fant, 1960). These peaks are called "formants", and they are closely related to resonances of the vocal tract, as determined by the shape of the tongue, lips, etc. In the production of different vowels, the shape of these vocal organs is changed and thus the location of the formants in the sound spectrum alters.

Careful study has failed to find a satisfactory relationship between the formant frequencies (or corresponding pitch values in mels) and the presumed dimensions of vowel quality. In fact, phoneticians have been unable to discover any physical dimensions which would serve to define the perceptual qualities which distinguish vowels (Ladefoged, 1967, pp. 100-103; Ladefoged, 1970). It was in hopes of discovering such relationships and giving objective acoustic definition to the dimensions of vowel quality that factor analysis of the formant data was carried out, using the proportional profile three-mode factor model.

Since, as has already been noted, the PP model is not completely general, its application to this three-way data set must be justified.



TABLE 5

## Data Input for Formant 0

1	215.000	215.000	215.000	215.000	215.000	215.000	215.000	215.000
2	255.000	255.000	255.000	255.000	255.000	255.000	255.000	255.000
3	135.000	135.000	135.000	135.000	135.000	135.000	135.000	135.000
4	205.000	205.000	205.000	205.000	205.000	205.000	205.000	205.000
5	350.000	350.000	350.000	350.000	350.000	350.000	350.000	350.000
6	260.000	260.000	260.000	260.000	260.000	260.000	260.000	260.000
7	300.000	300.000	300.000	300.000	300.000	300.000	300.000	300.000
8	245.000	245.000	245.000	245.000	245.000	245.000	245.000	245.000
9	325.000	325.000	325.000	325.000	325.000	325.000	325.000	325.000
10	300.000	300.000	300.000	300.000	300.000	300.000	300.000	300.000
11	365.000	365.000	365.000	365.000	365.000	365.000	365.000	365.000

## Data Input for Formant 1

1	405.000	475.000	655.000	930.000	0.0	590.000	465.000	420.000
2	275.000	455.000	695.000	855.000	760.000	555.000	460.000	370.000
3	480.000	535.000	710.000	915.000	765.000	615.000	505.000	400.000
4	395.000	505.000	745.000	995.000	810.000	590.000	535.000	340.000
5	345.000	540.000	635.000	920.000	715.000	715.000	495.000	350.000
6	290.000	415.000	730.000	945.000	0.0	725.000	430.000	290.000
7	340.000	505.000	690.000	980.000	0.0	665.000	515.000	360.000
8	335.000	455.000	715.000	910.000	715.000	720.000	415.000	360.000
9	355.000	550.000	715.000	1045.000	0.0	435.000	505.000	335.000
10	315.000	475.000	680.000	980.000	0.0	650.000	485.000	395.000
11	370.000	505.000	765.000	1115.000	760.000	595.000	595.000	365.000

## Data Input for Formant 2

1	1720.000	1620.000	1460.000	1370.000	0.0	0.0	0.0	0.0
2	1620.000	1605.000	1485.000	1260.000	980.000	825.000	825.000	0.0
3	1650.000	1620.000	1530.000	1305.000	960.000	795.000	695.000	695.000
4	1790.000	1660.000	1620.000	1400.000	1015.000	895.000	855.000	0.0
5	1765.000	1805.000	1620.000	1350.000	975.000	960.000	790.000	740.000
6	1760.000	1645.000	1510.000	1455.000	0.0	0.0	745.000	735.000
7	1670.000	1615.000	1510.000	1400.000	0.0	980.000	830.000	0.0
8	1770.000	1620.000	1470.000	1380.000	980.000	940.000	0.0	0.0
9	1880.000	1820.000	1700.000	1405.000	0.0	845.000	0.0	0.0
10	1725.000	1670.000	1585.000	1430.000	0.0	0.0	0.0	0.0
11	1775.000	1765.000	1645.000	1405.000	985.000	870.000	0.0	855.000

## Data Input for Formant 3

1	2030.000	1760.000	1760.000	1835.000	1835.000	1875.000	1965.000	2070.000
2	1895.000	1870.000	1905.000	1720.000	1785.000	1720.000	1705.000	0.0
3	1960.000	1885.000	1750.000	1855.000	1865.000	1825.000	1735.000	1735.000
4	2135.000	1960.000	1895.000	1855.000	1940.000	1900.000	1795.000	2100.000
5	2075.000	1955.000	2030.000	2050.000	2045.000	2040.000	2015.000	0.0
6	2145.000	1865.000	2135.000	1670.000	1775.000	1750.000	2115.000	2160.000
7	2040.000	1820.000	1820.000	1940.000	1800.000	1750.000	1680.000	1710.000
8	2045.000	1860.000	1715.000	1840.000	2035.000	1945.000	1855.000	0.0
9	2260.000	2060.000	2015.000	1995.000	2080.000	1985.000	2175.000	2240.000
10	2200.000	1900.000	1850.000	1895.000	1965.000	1905.000	2045.000	2145.000
11	2140.000	1950.000	1930.000	1985.000	2090.000	2095.000	2085.000	2175.000

Formant pitch values (in mels) for the fundamental frequency ("formant 0") and the first three formants of eight cardinal vowels as spoken by 11 persons. Zero values represent missing data points, which during analysis were simply skipped and not "fitted" by the optimization program.

The application is of particular interest since none of the three modes represents a time dimension. Although throughout the earlier discussion of the PP model, the third mode was called "occasions", it could in fact be any third set of circumstances in which systematic variations of relative contribution of the factors might be expected. In the case of this vowel data, the measurements (formant pitch values) for the individuals (persons producing the vowels) are taken for eight different vowels (i, e, ε, a, ɑ, ɔ, o, u).

The PP or system-variation model is thought to be justified here because systematic general differences of factor influence are expected not only when one vowel is compared to another, but also from one person to the next, and from one formant to the next.

By definition, the individual vowels must differ in their relative positions on the underlying dimensions of vowel quality, since it is these dimensional differences which are presumed to distinguish them. But it can also be expected that different persons will not use each of the underlying dimensions to exactly the same degree in their vowel production. The speech of some will tend, in general, to be slightly more prominent in the use of a certain aspect of vowel quality, and on the other hand have perhaps slightly less of a different aspect, consistently across all vowels. These differences could arise because some speakers tend to use a greater amount of lip rounding on all vowels, other speakers tend to speak with the tongue slightly higher in the mouth than is usual, etc. In addition, physiological differences in the shape of each person's vocal tract, differences in tongue size, etc., could be expected to cause shifts in the quality of that person's vowels compared to those produced by other speakers. These shifts would presumably be of different degrees for each of the underlying dimensions, since alterations of vocal tract shape would probably not have precisely equal consequences for all dimensions of vowel quality.

The argument for system-variation seems less convincing with respect to the formant mode. There seems to be no obvious reason for expecting any particular pattern of factor loadings across the three formants. But of the various possibilities, the vast majority would show unequal and distinct loading patterns for each factor across the three formants. Therefore, it does not seem unlikely that the required system-variation would be present in the formant mode as well.

Before actually performing an analysis, one can only argue that this system variation model of vowel quality is a plausible one. One cannot really know whether the model is, in fact, correct. But after the analysis the appropriateness of the model can be confirmed or not by examining the results. If one or more of the factors do not show the postulated system-variation, this should become apparent in the results of the analysis. From the experiments with synthetic data it was learned that if system-variation of a given factor's influence is not present across all modes of a data set, then that data set would not be "adequate"

for that factor, and the solution for that factor would not be unique.

The analysis of this real data was performed by fitting one, two, and three factor PP solutions to the data using Routine TBU. The quick algorithm could not be applied in order to speed computation because of the many missing entries in the data matrix (the missing entries appear as zeros in Table 5). Although attempts are being made to overcome the difficulty, the quick algorithm has not yet been modified to adequately deal with missing data. The problem arises because the computation of new values for a given mode requires intermediate products which involve the complete data set (equation 13). Routine TBU, on the other hand, can deal easily with missing values, since it does not require such intermediate products but relies instead on changes in the sum of the squared errors resulting with each attempt to "fit", by estimation from factor loadings, the non-missing values of the data set. Any missing value can simply be ignored by the program.

After applying TBU to the vowel data, it became apparent that the data set possessed characteristics which substantially slowed the process of convergence. Since TBU is already slow compared to the special purpose quick algorithm, true convergence could not be obtained with many of the analyses. Nonetheless, by pushing convergence as far as computer time would allow, results have been obtained which are believed to represent reasonable approximations to what the values would have been at convergence. Because of the additional slowdown of TBU as the number of factors increases, solutions containing four factors were not attempted. This omission is less serious however, because the results with two and three factors show that any fourth factor, even if present, would only make a very small contribution to the values of the data points. And it has been demonstrated with synthetic data that omitting a small fourth factor will not substantially distort the loadings of the other three factors.

In light of these limitations, the results reported as the two and three factor solution for this vowel set must be viewed as likely approximations, to be refined further when the quick algorithm is adjusted to cope with missing values, and also to be confirmed by obtaining and analyzing different sets of vowel formant data which do not have missing values. Both lines of work are currently being pursued.

Nonetheless, PP factor analysis of this vowel data is not simply an exercise of the program, but has inherent interest from the viewpoint of linguistic phonetics. Further, it will be seen that the solution obtained possesses a highly meaningful form. The results therefore seem to demonstrate the usefulness of PP analysis, even when operating under handicaps.

#### Results of the Analysis of Vowel Data

The analysis was carried out first for the four vowels (*i*, *e*, *ɛ*, *a*).

In order to determine the number of real factors, comparisons were made of both the amount of error remaining after convergence slowed or stopped, and of the uniqueness of the solutions obtained as one, two, and three factors were extracted. These criteria had to be used flexibly, however, because of the approximations caused by imperfect convergence, and also because of the distortions which would be caused by the presence of error variation in the data. On the basis of 10 analyses of the restricted four vowel data set, an apparent cutoff was obtained at two factors. The mean squared error of estimation was consistent at around 6,600 mels both for four different two-factor and for four different three-factor analyses. The analyses were performed from different random starting configurations. There were variations of only around 10% in the final mean squared error values for all eight of these analyses, and the three-factor solutions did not tend to have lower scores. With regard to uniqueness, the results were less clear. Both for the two-factor and three-factor solutions, there were similarities among solutions, but no clear uniqueness which could be said to hold for the loadings in all three modes. The vowel mode seemed to give the most consistent loading patterns. In all two-factor solutions the vowels showed a systematically increasing set of loadings on one factor and a systematically decreasing set of loadings on the other factor. Loadings for the person mode showed some similarities in successive analyses but could not be said to be recognizably unique even for two factors. The same was true for the formant mode. Even the three-factor solutions for the four vowel data set showed some interesting similarities in the vowel mode, as can be seen in Table 7. It was concluded that the error variation in the data, the convergence difficulties, and the possible distorting presence of a third factor were making the analysis difficult.

In an attempt to minimize the inaccuracy due to incomplete convergence, a two-factor solution was carried to 350 iterations of Routine TBU (requiring 300 seconds of CPU time on the 360/91). This represented at least five times as many computations as on any of the other analyses of the four vowel data set. Yet convergence still had not been achieved at the end of this time. These results were therefore taken as the best approximation, short of true convergence, obtainable for the loadings of two factors on the four vowel data. The obtained loadings are given in Table 6, and the position of the four vowels in the plane determined by the two factors is given in Figure 1.

The data set was then expanded to include all eight vowels, and the analysis procedure repeated. An immediate improvement was noticeable in the stability of the solutions obtained for both two and three factors. Further, with this expanded data set the mean squared error of fit could be seen to drop from 7000 mels for two factors to 2800 mels for three factors. Although the precise values of the vowel loadings varied somewhat across analyses, they were roughly similar, and the patterns of the loadings was more clearly unique, for the three-factor as well as the two-factor solutions. In all cases there was one factor which would increase across the eight vowels, another factor which would decrease, and

TABLE 6A

## Two-Factor Analysis of Four Vowels

Formant Loadings		
	Factor 1	Factor 2
1	0.248	0.240
2	-0.036	1.227
3	1.836	0.996
4	1.952	1.536

## Vowel Loadings

	Factor 1	Factor 2
1	1.375	0.615
2	1.151	0.783
3	0.915	1.082
4	0.559	1.520

## Person Loadings

	Factor 1	Factor 2
1	552.258	508.999
2	553.917	489.476
3	540.112	524.531
4	587.279	533.850
5	600.630	545.492
6	597.261	486.751
7	546.270	546.178
8	567.719	502.570
9	631.319	554.582
10	589.327	526.453
11	578.573	576.265

TABLE 6B

## Two-Factor Analysis of Four Vowels

Formant Loadings		
	Factor 1	Factor 2
1	143.0430	126.4400
2	-20.7644	646.4220
3	1058.9800	524.7240
4	1125.8900	809.2130

## Vowel Loadings

	Factor 1	Factor 2
1	1.375	0.615
2	1.151	0.783
3	0.915	1.082
4	0.559	1.520

## Person Loadings

	Factor 1	Factor 2
1	0.9575	0.9662
2	0.9604	0.9291
3	0.9364	0.9956
4	1.0182	1.0133
5	1.0413	1.0354
6	1.0355	0.9239
7	0.9471	1.0367
8	0.9843	0.9540
9	1.0946	1.0527
10	1.0218	0.9993
11	1.0031	1.0938

Two factors extracted by extended analysis of the data for the first four vowels. Table 6b is the same as 6a, except that the normalization has been performed across persons so that formant loadings reflect true pitch values in mels.

TABLE 7

Three Attempts at Fitting Three Factors to Four Vowel Data

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	-0.578	-9.213	-1.154
2	0.442	10.930	4.140
3	2.081	0.972	0.000
4	2.055	1.312	1.014

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	0.054	0.222	2.015
2	0.260	2.031	-1.192
3	1.767	0.477	1.412
4	1.918	1.270	1.765

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	0.896	-3.276	4.764
2	-2.052	3.040	-2.174
3	2.605	2.952	-0.803
4	2.551	1.285	2.213

Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	1.111	0.983	0.811
2	1.025	0.942	0.862
3	0.973	0.998	1.038
4	0.892	1.076	1.289

Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	1.173	0.487	1.199
2	1.065	0.702	0.987
3	0.957	1.116	0.971
4	0.805	1.695	0.843

Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	2.237	0.921	0.894
2	1.610	0.930	0.901
3	0.733	1.015	1.017
4	-0.580	1.134	1.188

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	792.646	-120.157	380.405
2	782.005	-125.907	387.647
3	790.147	-109.530	363.755
4	836.736	-124.931	400.044
5	857.004	-147.075	443.417
6	820.388	-127.126	387.872
7	805.151	-136.716	427.242
8	802.046	-125.324	390.155
9	885.816	-147.557	455.140
10	834.089	-138.044	423.661
11	845.975	-151.325	474.186

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	752.396	241.872	59.777
2	720.434	240.489	83.052
3	788.191	236.796	20.412
4	805.078	253.468	53.670
5	757.301	275.162	121.408
6	760.334	236.580	88.424
7	716.428	279.801	100.674
8	750.628	243.421	73.744
9	805.671	281.166	104.401
10	755.095	264.342	100.735
11	735.898	310.031	120.937

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	82.893	559.428	415.109
2	86.003	544.676	414.191
3	77.347	567.430	409.648
4	91.828	588.885	432.536
5	86.034	600.636	471.115
6	97.677	561.921	426.369
7	77.982	574.820	441.968
8	88.603	560.100	419.356
9	99.474	617.177	471.446
10	91.351	581.856	444.990
11	87.519	602.929	465.261

Sets of loadings resulting from separate attempts to fit three factors to the first four vowels. Although the vowel loading patterns were fairly consistent, the formant loadings and person loadings showed fairly large variations, indicating relative lack of uniqueness or very high distortion by error variations in the data.

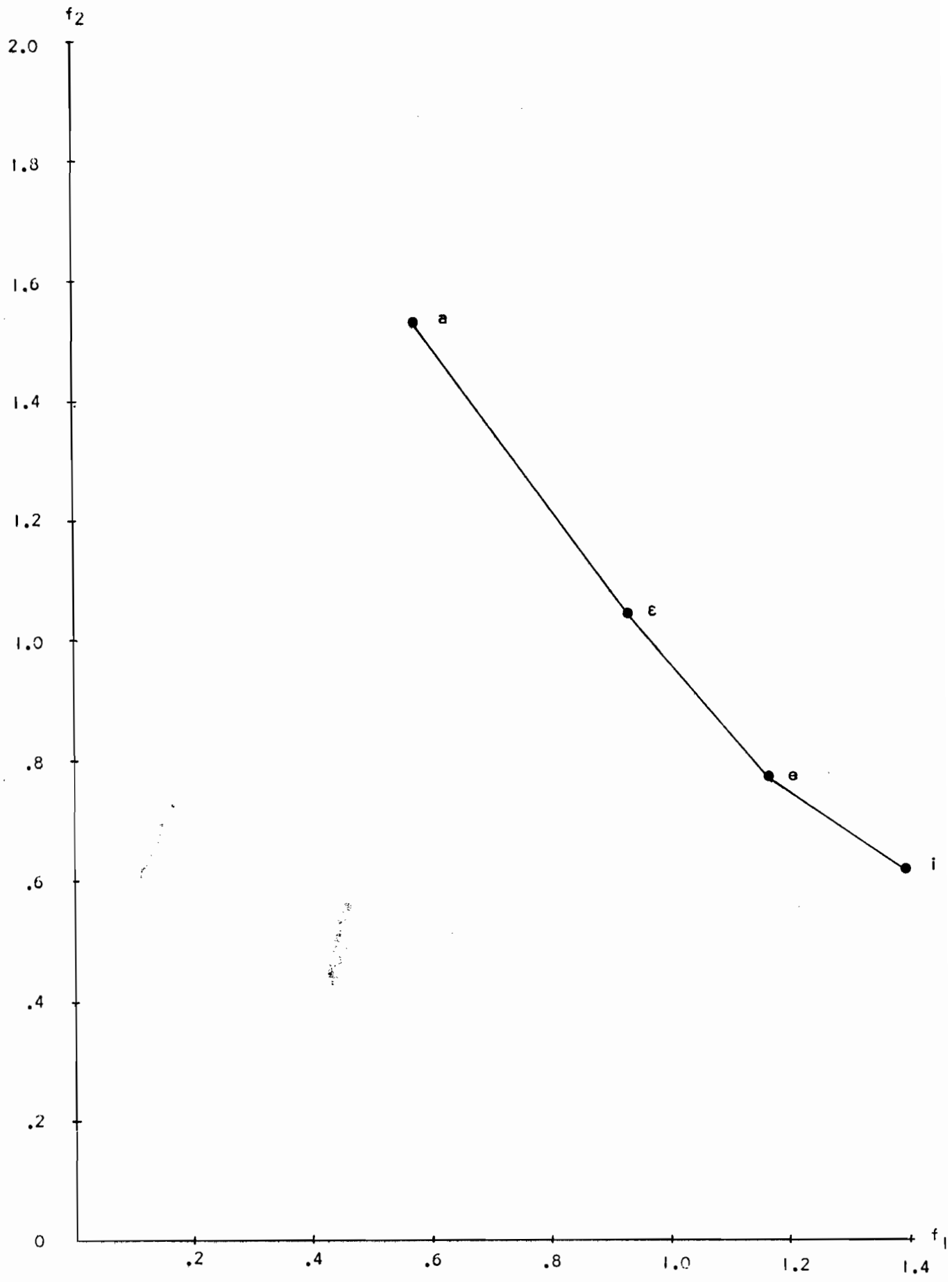
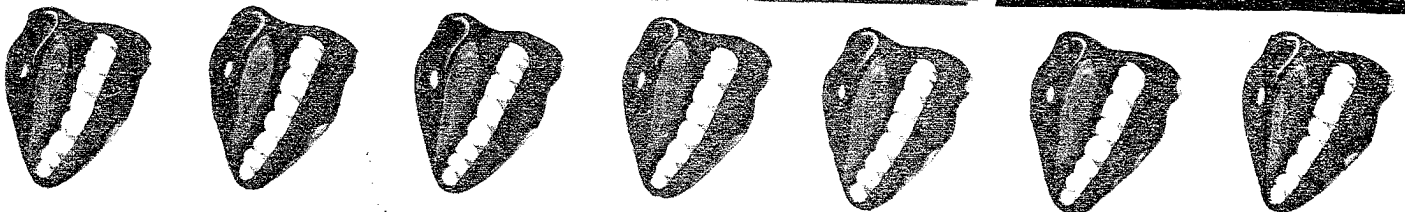
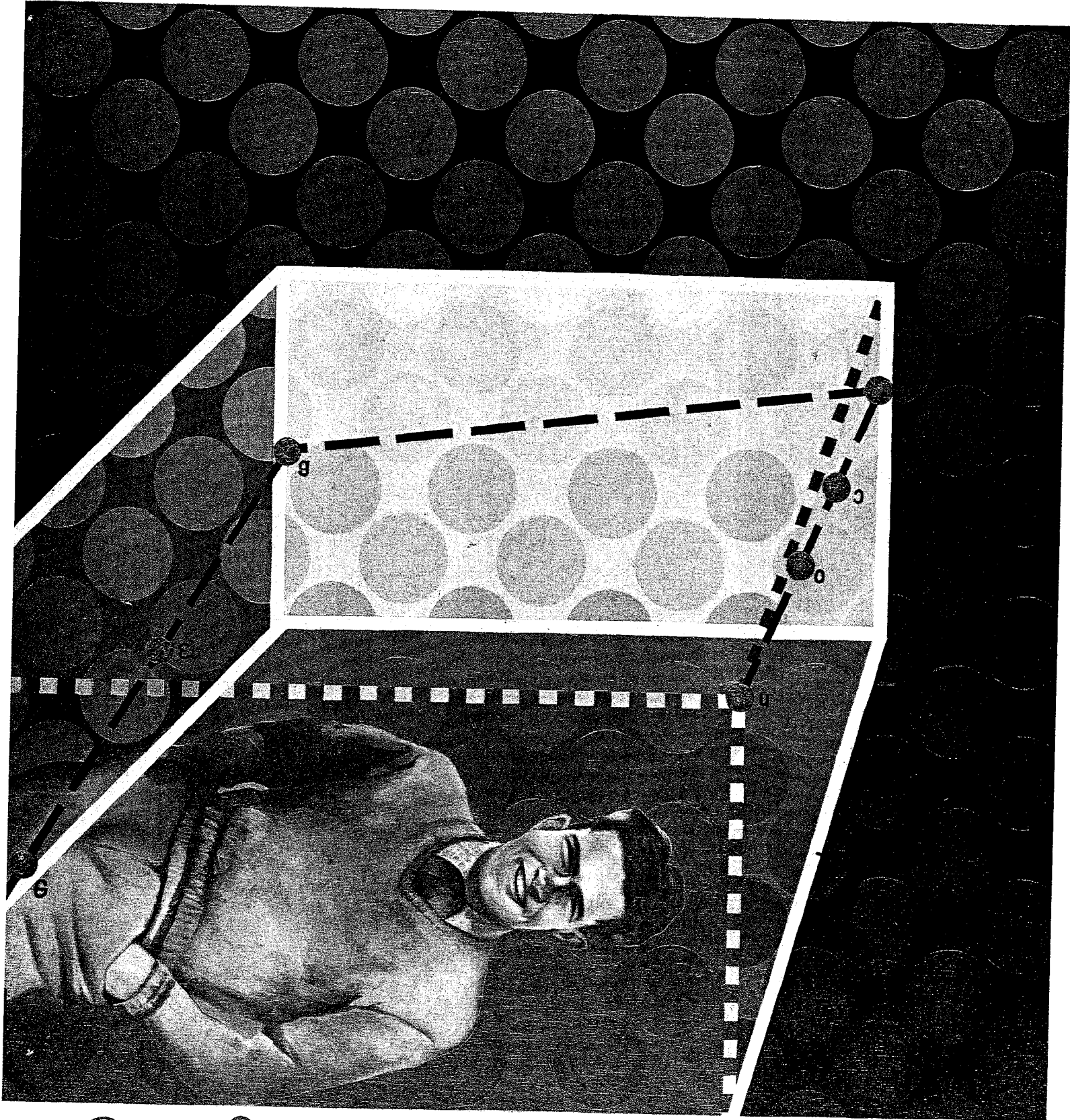


Figure 1. Plot of the position of the first four vowels in the two-factor plane extracted from the four vowel partial data set.

U.C.L.A.  
W.P.P. NO. 16  
DEC. 1970



PARAFAC  
by  
RICHARD A. HARSHMAN





FOUNDATIONS OF THE PARAFAC PROCEDURE: MODELS AND CONDITIONS  
FOR AN "EXPLANATORY" MULTI-MODAL FACTOR ANALYSIS

by

Richard A. Harshman

U C L A

*Working Papers in Phonetics*

16

December, 1970

## ACKNOWLEDGEMENTS

The author wishes to express his great indebtedness to Peter Ladefoged, without whose support this work would probably not have been possible.

The author is also very grateful for the help of Hans Reichenbach with the optimization program, Robert Jennrich for the proof of uniqueness, and the fast algorithm, and to Dale Terbeek, Lee Cooper, Andrew Comrey, and Peter Bentler for their comments, criticisms, and the generous donation of their time.

It was the author's good fortune to have the manuscript prepared by Jeanne Yamane who contributed an unusual combination of mathematical and secretarial abilities, plus care and concern with the quality of the final result.

The computation for this project was paid for by an intramural grant from the UCLA Campus Computing Network, and performed on their IBM 360/91 computer.

Finally, the author wishes to thank Diane Vaughan (now Mrs. Diane Harshman) for her helpfulness, love and support which have contributed so much toward the successful completion of this project.

Cover: Bill Jahnke (oil on canvas)

This research has been funded in part by NSF grant GS 2741, USPHS grant NB 04595, and UCLA intramural research funds.

## ABSTRACT

Simple structure and other common principles of factor rotation do not in general provide strong grounds for attributing explanatory significance to the factors which they select. In contrast, it is shown that an extension of Cattell's principle of rotation to Proportional Profiles (PP) offers a basis for determining explanatory factors for three-way or higher order multi-mode data. Conceptual models are developed for two basic patterns of multi-mode data variation, system- and object-variation, and PP analysis is found to apply in the system-variation case.

Although PP was originally formulated as a principle of rotation to be used with classic two-way factor analysis, it is shown to embody a latent three-mode factor model, which is here made explicit and generalized from two to  $N$  "parallel occasions". As originally formulated, PP rotation was restricted to orthogonal factors. The generalized PP model is demonstrated to give unique "correct" solutions with oblique, non-simple structure, and even non-linear factor structures.

A series of tests, conducted with synthetic data of known factor composition, demonstrate the capabilities of linear and non-linear versions of the model, provide data on the minimal necessary conditions of uniqueness, and reveal the properties of the analysis procedures when these minimal conditions are not fulfilled. In addition, a mathematical proof is presented for the uniqueness of the solution given certain conditions on the data.

Three-mode PP factor analysis is applied to a three-way set of real data consisting of the fundamental and first three formant frequencies of 11 persons saying 8 vowels. A unique solution is extracted, consisting of three factors which are highly meaningful and consistent with prior knowledge and theory concerning vowel quality.

The relationships between the three-mode PP model and Tucker's multi-modal model, McDonald's non-linear model and Carroll and Chang's multi-dimensional scaling model are explored.

## CONTENTS

I.	Introduction: The Search for "Explanatory" Factor Analysis . .	5
II.	The Principle of Parallel Proportional Profiles . . . . .	15
III.	Development of the Three-Mode PP Model and Other Generalized Factor Models . . . . .	18
IV.	Testing the Three-Mode Proportional Profiles Model . . . . .	27
V.	A Uniqueness Theorem for PP Factor Analysis . . . . .	61
VI.	Generalization to Non-Linear Factor Models . . . . .	63
VII.	The PP Model in Relation to Other Non-Linear and Multi-Modal Factor Models . . . . .	69
	References . . . . .	82

## I. INTRODUCTION: THE SEARCH FOR "EXPLANATORY" FACTOR ANALYSIS

Basic to the use of factor analysis as a tool for scientific discovery is the distinction between its "descriptive" and its "explanatory" application. While descriptive factor analysis seeks merely to find a convenient, condensed representation of data relationships for a given case, explanatory factor analysis seeks to discover good estimates of the structure of "true underlying" influences that are responsible for the observed data relationships. The degree to which factor analysis can provide explanatory solutions has always been a matter of controversy.

The problem lies in the factor model itself. Descriptions in terms of the classic factor model are not sufficiently constrained by the data. For any given set of data in which there is more than one factor common to the measures, the classical factor model allows for many possible sets of factors, all of which would equally well fit the data in a manner consistent with the model. But although these solutions are mathematically equivalent within the constraints of the model, they do not lead to equivalent interpretations of the influences underlying or organizing the data relationships. The classical factor model does not give us grounds for choosing among these many different possible explanations or interpretations of the data relationships.

Is the Search for Explanatory Factors Meaningful?

How meaningful is the distinction between convenient descriptive factors and so-called "true" or "valid" explanatory factors? A basic part of the controversy over explanatory factor analysis centers on the question of whether there ever exists a unique "true" structure of factors which "really" underlies a set of data. Thurstone, for example, argues against the search for unique solutions in factor analysis:

When a factor analysis has been completed, ... the first question is naturally: "What *are* the parameters or factors?" It seems to be in the nature of science that such a question has no unique answer. *In scientific work a parameter is one of the measurable attributes of an object in terms of which it is described ... a circle on a graph is defined by the two co-ordinates (x,y) of its center and by the radius r. These parameters are easily understood and easily used in most problems, but, of course, they are not unique. For some problems they would be awkward and another set of parameters would be chosen. The same is true in scientific work ... To hunt for a unique solution in the comprehension of a set of related phenomena is an illusory hunt for absolutes.*

(Thurstone, 1947, p. 332) (his italics)

But in this comment, Thurstone seems to be conceiving of factors as playing a purely descriptive role. For such descriptive factors the selection of any particular solution from among the possible alternatives is based entirely on the simplicity, convenience and usefulness of the resulting description. In Thurstone's classic box problem (1947, p. 360), for example, the variables all consisted of different measurements on the size and shape of boxes (e.g. area of a side, length of the diagonals, length of the perimeter, etc.). Relations among these variables are "passive", in that they have no causal significance. There are no functional systems involved, and none of the factors extracted could be thought of as underlying causal influences. Since no *explanation* of the observed relations is sought, the choice of length, width, and depth as the extracted factors is somewhat arbitrary. It is simply the convenience of working with descriptions in terms of those three dimensions (as opposed to some factors which are linear combinations of length, width, and height), which justifies favoring this solution.\*

Such descriptive applications of factor analysis should be contrasted with applications in other domains, where the factors sought are the causes of patterns of covariance among the measures, or where one wishes to identify functional systems in natural processes. For example, tidal level and amount of daylight might be explanatory factors underlying the patterns of covariation in measurements of seashore activity, and functional systems might be sought in the ecology of the tidepools. The notion of an explanatory factor does not allow choice among factorial representations to be based on simplicity or convenience of the solution since a correspondence with some real external structure is sought.

Consider the classic debate over the structure of human intelligence. From the same matrix of correlations among tests many different descriptions of the structure of intelligence can be derived. In one description, intelligence consists of one general intellectual factor plus many uncorrelated group factors corresponding to specific skills; in an alternative description, intelligence consists of a set of correlated basic intellectual abilities. Can one decide among these alternatives, or is the search for a unique solution to this question simply an "illusory hunt for absolutes"?

It might be argued that the two different sets of explanatory factors are really two different ways of describing the same underlying system. And it is true that any real system, consisting of a complex network of interrelationships, can be described in a number of different ways. But *explanatory* descriptions have implications beyond the current set of measurements being described. Explanatory descriptions imply predictions about

---

\* Since these factors were determined by simple structure rotation, they presumably represent a factorially invariant description. This gives that solution another type of convenience or usefulness. The significance of factorial invariance will be discussed later in this section.

the results of other possible experiments. Therefore descriptions which are equivalent for a given data set are not generally equivalent as *explanations* of the relations in that data set. It is true that such things as changes in terminology and level of detail can alter one's perspectives while still producing valid and equivalent explanatory descriptions, but in factor analysis one's descriptions cannot be so freely reformulated. With any given factor analysis one is dealing with some particular limited set of empirical variables and these variables constitute the descriptive "vocabulary". Any factorial description of the underlying functional system must consist of specifying specific relations between each variable and each latent factor. In light of these considerations, it seems much less likely that two alternative sets of explanatory factors could be really equivalent.

It would seem necessary to try to decide between alternative sets of explanatory factors whenever the two explanations are "really different". In science, being "really different" amounts to having different empirical consequences. The fact that both models of human intelligence have the exact same mathematical consequences within classical factor analysis need not show the "meaninglessness" of the distinction. Rather, it might show the weakness of the classical factor model. In fact, the two conceptions of intelligence could not be called "really the same" unless they had the same empirical consequences (would make the same predictions) in all conceivable circumstances (e.g. in developmental studies, under the influence of drugs, etc.). But such is clearly not the case. They would, for example, predict different possible patterns of disruption of intellectual functions by drugs or as a result of brain damage. It would seem that most sets of alternative explanatory factors which are equivalent within the classical factor model are nonetheless meaningfully different, since the hypotheses they suggest would make differing predictions in certain appropriate circumstances.

Because of these differences, one approach to the explanatory use of factor analysis has been to use factor analysis to generate alternative hypotheses (such as differing models of intelligence) and then devise and perform crucial observations or tests in order to select among the alternatives. With this approach the solutions of factor analysis only have explanatory force as part of a larger experimental program.

A single factor analysis can often generate a large number of alternative hypotheses (since for a given multi-factor matrix there are an infinite number of alternative sets of factors). Therefore it would be desirable to develop a system of factor analysis which would directly determine unique explanatory factors. This could be accomplished by enlarging the conceptual-mathematical model of factor analysis in such a way that the many alternate sets of factors are no longer equivalent within the model, but indeed have differing consequences which allow one solution to be preferred. This amounts to strengthening the model or the rotation process so that it incorporates the same type of ability to distinguish among alternative classical solutions that was previously only

possible by using factor analysis in conjunction with external experimental tests. For such a strengthened type of factor analysis, the fact that a given set of factors was discovered underlying a set of data would directly give those factors explanatory plausibility. This article proposes such a system of factor analysis. But before it is developed an examination should be made of the weaknesses of the traditional techniques for selecting among alternative sets of factors.

#### Traditional Approaches to Choice of Factors

The problem of determining which set of factors will be derived from a given factor analysis has been traditionally handled by adding a second stage of analysis after the factor extraction process. This second stage of analysis is called factor rotation, and consists of successively transforming the factors into mathematically equivalent alternative sets of factors, until some added criterion or principle is fulfilled. This added criterion forms the basis for a choice among the infinite number of possible solutions derivable as rotations of any given factor solution.

An alternative, however, would be to replace the classical factor model with one which is not inherently underdetermined by the data. No additional rotation criterion would then be needed. But although Cattell (1944) came close to this alternative (see Sections II and III below), the search for explanatory factor analysis has proceeded, instead, by the development of principles of rotation for solutions derived from the classic factor model. And although great progress has been made along these lines, no fully adequate solution to the rotation problem has been found. As Comrey says "the rotation process has been the target of much criticism, and continues to be the weakest link in the entire [factor analysis] process" (Comrey, 1967, p. 143).

One approach to the rotation problem is exemplified in a number of "special principles" which have been adopted by individual researchers to choose a rotation for their particular data or type of data. These principles usually amount to finding a rotation which seems "reasonable" or "meaningful" in that it agrees with knowledge about the data based on other grounds than the factor analysis itself. Insofar as this brings additional valid information to bear on the choice of rotation, it is a desirable improvement over an unrotated arbitrary solution. But this approach can yield many different solutions when used by individuals with differing knowledge or opinions about what constitutes an approximation to a "reasonable" set of relationships. Whatever merit such special principles might have in a particular circumstance, they do not constitute an *objective general* criterion which would tend to select the rotation with explanatory validity.

#### Is Simple Structure Explanatory?

The closest that factor analysis has come to an agreed general



principle of rotation is the concept of simple structure. The object of this approach is to minimize the number of factors influencing each variable. This amounts to finding a rotation which, as much as possible, allows each variable to be described in terms of a few factors which influence it strongly, rather than many factors which influence it less strongly. This simplifies the factorial description of the variables, hence the name "simple structure". The initial definition of simple structure was given by Thurstone (1935, p. 156; 1947, p. 335) and consisted of rules giving the proportion of near zero values that should appear in the rows and columns of the factor matrix. This definition of "simplicity" has been supplemented and extended with the advent of computerized analytical rotation methods. Varimax, for example, tries to achieve a complimentary type of simplification, i.e. to explain each factor with as few variables as possible (Harman, 1960, p. 301).

Rationales have been presented for rotation to simple structure in both descriptive and explanatory factor analysis. In the case of descriptive factor analysis, the justification is simple. A few large loadings are easier to deal with than many small ones, and, in a sense, rotation to simple structure gives the most parsimonious description of a set of data.

In explanatory factor analysis, rotation to simple structure is less clearly justified. Any description, however compact and convenient, which does not correspond to the underlying causal or "functional unities" (to use Cattell's terminology) would be misleading, if interpreted in an explanatory sense.

The strongest general argument for preferring simple structure solutions when looking for explanatory factors is based on the principle of factorial invariance. This principle was formulated by Thurstone (1947, p. 361) as follows: "A fundamental requirement of a successful factorial method, [is] that the factorial description of a test must remain invariant when the test is moved from one battery to another which involves the same common factors." The description of a given variable should not fluctuate wildly from one study to another as a mere mathematical artifact of the rotation process. If reality is consistent, then any "explanatory" solution would have to reflect this consistency.

It has been argued by Thurstone (1947), Cattell (1952) and others, that simple structure rotations tend to have this property of factorial invariance. Kaiser (1958, p. 195) has made this claim for the Varimax criterion. He presents an argument to show that it would have this invariance in the special case in which all variables fall into two colinear clusters, but stated that the generalization to a greater number of factors would involve enormous difficulties.

Cattell seems to argue that some types of invariance give convincing proof that such factors represent the true form of underlying organic unities:

... the rediscovery of the same factors despite (a) partially different test batteries or (b) populations of different age, education, or dispersion, and (c) independent factorizations and rotations, is a proof that they have an existence as something more than mathematical equivalents -- that they are in fact functional unities in nature. (Cattell, 1955, p. 90)

Such arguments, based on a consistent finding, do not rule out the real possibility of a consistent distortion of the true form of the latent factors being consistently generated by the repeated application of consistent, but inappropriate, rotation criteria.

Even if simple structure could be clearly demonstrated to have perfect factorial invariance across data sets (as defined by Thurstone), this would not be a convincing argument for the explanatory validity of simple structure rotations. This is because factorial invariance is a necessary but not sufficient condition of explanatory validity. To see this, simply imagine that there is a certain real latent structure underlying a set of variables (let us call this structure S), but also imagine that this latent structure is not of simple structure form. Now if a factor analysis is performed of data with latent structure S and the simple structure criterion is used for rotation, then a different factor structure will be extracted, let us call it structure S'. Now it is likely that the form of S' is largely determined by the form of S, i.e. some sort of systematic distortion of S. Since S is a real and invariant latent structure measured in several different factorial studies, one would expect S', the simple structure distortion of S, to be consistently extracted from all these studies. The stability of S' across a number of different studies shows (perhaps) that it is a consistent transformation of a real latent data structure, but it does not in any way show that it is an undistorted expression of that latent data structure. Thus the finding that certain simple structure factors have "stood up" under the test of time and in a number of different factorial studies does not in itself give any evidence that they are accurate descriptions of the latent influences underlying the data. It only gives evidence for the existence of *some* factors that are related, by a consistent but unknown rotational transformation, to the simple structure solution.

An inverse and equally objective criterion of "most complex structure" could probably be defined and would probably show similar consistency across studies. If not "most complex structure", clearly there are going to be some principles, related in a systematic way to simple structure, which would show similar invariance of results across studies. Thurstone admits that there are probably principles other than simple structure which would show factorial invariance (Thurstone, 1947, p. 364). But if this is so, then factorial invariance alone is clearly an insufficient argument for attributing explanatory significance to a principle of rotation.

An important additional requirement for any "true" factor solution was proposed by Tucker (1958, p. 112). It is the converse of the principle of factorial invariance of variable description, namely "that the scores of individuals on a factor should remain invariant as the individuals are tested with different batteries which involve the factor". Clearly, if a person is found to have a small loading on the hostility factor with one battery, and a large loading with a different battery of tests, both cannot be measuring a "true" objective factor pattern (assuming that the individual hasn't had time or circumstances which might change his personality).

Tucker's principle, which might be called a principle of factor score invariance, is also a necessary but not sufficient condition of explanatory validity. The same argument just advanced in relation to factorial invariance could be used here. Nonetheless, when these two principles are fulfilled in the rotation, we can have *more* confidence in the explanatory validity of a resulting solution than we could if it were just an arbitrarily selected rotation. Only a subset of possible rotations should fulfill these criteria, and the true solution would be found among the members of that subset.

If there is not, in general, a sound basis for attributing explanatory significance to factors determined by simple structure rotation, might there nonetheless be a valid explanatory application in some particular circumstances? The most sophisticated use of simple structure involves just such restricted application to certain selected types of data sets. The reasoning behind this approach has been well developed by Cattell (1955). He argues that not all correlation matrices will yield factors which can be rotated to a clear and compelling simple structure (with a number of loadings very near zero for each factor), and therefore the discovery of such a clear structure among the possible rotations of a factor analysis is of great significance.

In this approach, it is not justifiable to take the results of a factor analysis and rotate them to the *best approximation* of a simple structure, in order to maximize the invariance and interpretability of the solution. Instead, it is acknowledged that a simple structure might not really underlie the data, and therefore, if the results of a given factor analysis cannot be rotated to a sufficiently clear and compelling simple structure, the results are not interpreted. In such a circumstance it is concluded that the latent factors did not show a sufficiently well defined simple structure, and thus the rotation criterion cannot be trusted to have correctly discovered their true underlying form.

In this approach it is admitted that simple structure is *not* a general procedure for extracting explanatory factors from data sets. Instead, the criterion is applied only to data sets in which the variables have been specially selected, on the basis of prior knowledge and

experience, so that none of the presumed latent influences would be expected to affect all of the variables included in the set. This approach recognizes that simple structure rotation will give explanatory solutions only if the true latent factors in fact have simple structure form. Therefore, an attempt is made to design data sets to fit this requirement. If, after extracting a set of factors from such a data set, it is indeed possible to find a rotation which shows a clear simple structure, this is taken as a confirmation of the original expectations. It is concluded that the attempt to build a data set whose factors would have simple structure was successful, and therefore simple structure rotation can be used as an appropriate guide to the true form of the underlying factors.

But even this restricted use of simple structure for explanatory factor analysis is subject to serious theoretical objections. The most fundamental objection is to the central claim that the discovery of a clear simple structure provides a confirmation of one's initial conception that the latent influences would have simple structure form. *Considered by itself*, such a discovery is at best only suggestive. This alleged "confirmation" could in fact occur when the latent factors were not in simple structure form but instead were merely in some other form which was capable of rotation into a clear simple structure. While it is true that there exist many sets of factors which cannot be rotated into a clear, compelling simple structure, there of course also exist many others which can be so rotated.

What, then, is the likelihood that the discovery of a clear simple structure is really a "false confirmation", revealing not the true latent structure but instead one of its alternative rotations? This question is difficult to answer. Certainly, the less clear and compelling the discovered pattern (i.e. the fewer near-zero loadings there are for each variable, and the further these loadings are from zero) the more likely it is that the resemblance to simple structure could have occurred by chance. But even if a configuration is discovered which is so clear that it could *not* have occurred by chance, it still need not have been due to simple structure patterns of influence. Other non-random influences or data characteristics can produce systematic factor relationships (hyperplanes oblique to true factor axes) which are rotatable to an apparent simple structure, but which would not display the zero loadings of a simple structure when the factors have their true orientation. A partial example occurs with the real data analyzed later in this article (see Figure 2, and Table 8, Section IV). The fact that alternative possible simple structures can sometimes be detected in a given data set (Cattell, 1955, p. 251, quoted below) further indicates that orderly patterns can arise which are not due to simple structure, but which can mimic the zero loadings of simple structure if rotated into a certain position.

In actual application, Cattell and others use many grounds besides the discovery of a clear simple structure to argue for the explanatory

validity of any set of factors (e.g. interpretability of the factors, relation to prior findings, etc.). The case for choosing the simple structure rotation from among the many possible rotations is strengthened by arguments based on prior knowledge of the data and its presumed latent influences, and in this way the factors gain in explanatory plausibility. But as with any choice of factors based partly on prior expectations and on "meaningfulness" of the solution, the stronger one can make such an argument for explanatory validity, the more one must know about the presumed latent structure *before* the analysis, so the less one stands to learn from a factor analysis.

Because of the tediousness of graphical rotation, and because some researchers object to its alleged "subjectivity", most simple structure solutions are now discovered through one of the analytical computer rotation procedures. But although the various computer rotation procedures give somewhat similar solutions, the results are sufficiently different to affect interpretation of the rotated factors. For example, "normal" Varimax rotation (Kaiser, 1958) tends to distribute the variance evenly across many factors, while the Quartimax method (see Harman, 1960, Chapter 14) tends to concentrate it on a general factor, with the remaining factors having smaller influences (Harman, 1960, p. 302).

An additional uncertainty arises from the controversy over orthogonal vs. oblique rotations. Restricting the factors to orthogonal positions is equivalent to requiring that all factors be uncorrelated with one another. But this complete independence of underlying influences seems unlikely to hold in many real situations. On the other hand, some workers have pointed out weaknesses in the principles of oblique rotation (Comrey, 1967; Harman, 1960).

Even the uniqueness of simple structure solutions is sometimes questionable. Cattell comments that more than one simple structure position is sometimes found for the same set of data (using graphical rotation) "and in absence of any generally acceptable criterion we cannot decide, from the given matrix alone and without extraneous evidence, which is preferable" (Cattell, 1955, p. 251). It has recently been demonstrated that Varimax is not always going to give a unique rotation, but it does not seem to diverge too seriously from this ideal (Gebhardt, 1968).

The basic question, of course, remains; what grounds do we have for believing that a given rotation corresponds to the true underlying structure of influence that generated the data? Comrey questions the reasonableness of simple structure for many applications:

If we sample at random from the entire universe of factors and use predominately factor-pure measures, simple structure will no doubt give results which are reasonably satisfactory. In many real-life factor analyses, however, where selection of variables is anything but random, and measures of considerable factor complexity sometimes predominate, one can

only hope that the simple structure criterion is approximately applicable. Simple structure pursued to its logical conclusion, i.e. by increasing to a maximum the number of near zero loadings per factor while allowing the axes to go oblique, can give results which may be misleading. The varimax criterion can also present problems. There is no particular reason why the variance of the squared factor loadings must be maximized, except that loadings are more easily interpreted if they are either high or low rather than medium in absolute value. It may well be, however, that they should be in medium range, rather than high or low, depending on the data being analysed.

(Comrey, 1967, p. 143)

Comrey presents an alternative pair of rotation criteria for factor analysis, namely that variables which appear on the same factor should be correlated, and uncorrelated variables should not appear on the same factor. This principle, Comrey admits, would not always be valid in real life, but "the principle will be substantially appropriate for most correlation matrices taken in their entirety". The argument for this principle relies on expected probabilities. As with the simple structure criterion, we are making implicit assumptions about the probable latent structure (i.e. that it is not of the sort that would systematically violate the two principles). But perhaps the probabilities are more in our favor with Comrey's assumption than with an assumption of simple structure. (Until one is more familiar with the significance of the conditions wherein the principle might be expected to be *systematically* violated, it would seem hard to judge.)

In view of all of the questions raised in this section, there does not appear to exist a principle of rotation from which one could make a strong and general argument for explanatory validity. One wants to find conditions where the rotation selected necessarily corresponds most closely to the latent structure that created the data relationships (except for deviations due to error variance). Such conditions would presumably be based on logical or mathematical considerations deriving from the factor model itself. Many factor theorists have given up hope of such a principle. Yet in 1944 Cattell came tantalizingly close to a solution to the problem when he formulated his principle of Parallel Proportional Profiles (Cattell, 1944). His approach is the foundation for the solution to be proposed in this article.

## II. THE PRINCIPLE OF PARALLEL PROPORTIONAL PROFILES

In 1944 Cattell published a review of principles which might be used for determining the desired rotation of a set of factors (Cattell, 1944). His objective was to discover the principle of rotation which had the strongest likelihood of explanatory validity. He rejected the notion that psychologists should be satisfied with a "convenient description" and nothing more. "Psychological research, as such, cannot ... be content with this limited goal. It strives toward psychologically meaningful functional unities ... 'How can one decide which one, among many possible sets of factors, alone corresponds to the real functional unities in the psychological situation?'" (Cattell, 1944, p. 267).

Cattell reviewed seven principles for choice of rotation. These were: (1) Rotation to agree with clinical and general psychological findings; (2) Rotation to agree with factors from past factor analyses; (3) Rotation to put factors through the center of clusters; (4) Rotation for simple structure; (5) The principle of orthogonal additions: rotation to agree with successively established factors; (6) The principle of expected profiles: rotation to produce loading profiles congruent with general psychological expectations; (7) The principle of parallel proportional profiles.

After discussing and finding objections to principles 1-6, Cattell focused his attention on (7), which he called "the most fundamental principle". He starts with the notion that "if real psychological functional unities exist they are bound to appear as possible mathematical factors in many different kinds of situations, whereas mathematical factors which are artifacts will stand only the test of fitting the particular matrix in which they happen to appear and may not be reproducible elsewhere" (Cattell, 1944, p. 274). But agreement across two different factorizations is not, itself, sufficient.

... it is clear that to require agreement in factors and factor loadings among correlation matrices derived from the same or similar test variables on the same or similar population samples, is an empty challenge. No new source of rotation determination is introduced, for such matrices will differ only by sampling errors and there will be an infinite series of possible parallel rotations in the two or more analyses. The special and novel required condition is that any two matrices should contain the same factors but *that in the second matrix each factor should be accentuated or reduced in influence by the experimental or situational design*, so that all its loadings are proportionately changed, thereby producing, from the beginning, an actual correlation matrix different from the first.

(Cattell, 1944, p. 274) (his italics)

The most concise statement of the logic of the parallel proportional profiles criterion occurs in a later article:

The basic assumption is that, *if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study.* No inorganic factor, a mere mathematical abstraction, would behave in this way ... This principle suggests that every factor analytic investigation should be carried out on at least two samples, under conditions differing in the *extent* to which the same psychological factors ... might be expected to be involved. We could then anticipate finding the 'true' factors by locating the unique rotational position (simultaneously in both studies) in which each factor in the first study is found to have loadings which are proportional to (or some simple function of) those in the second: that is to say, a position should be discoverable in which the factor in the second study will have a pattern which is the same as in the first, but is stepped up or down.

(Cattell, 1955, p. 84) (his italics)

This rotation criterion overcomes the objections leveled at the other techniques. (1) It is completely general, in the sense that it makes no assumptions about the size or distribution of factor loadings in a given study, but only asks that they be proportional to the loadings for that same factor in another study. (Although for philosophical reasons Cattell also called this principle "simultaneous simple structure", it in fact bears no relation to the mathematical assumptions and procedures of simple structure rotation criteria.) (2) It is objective, in that it does not require judgments of "reasonableness" based on other knowledge of the data under study. (3) Finally, the criterion has explanatory validity, since we would expect the parallel proportionality conditions to be fulfilled with real latent factors and *only* with real latent factors.

But there was a basic problem with the implementation of the Proportional Profiles criterion. Cattell was only able to establish a procedure for discovering proportional profiles with orthogonal factors (Cattell, 1955). If the true underlying factors were not in fact orthogonal, then Cattell's technique would not be able to discover them. Since it is not normally possible to be sure, in advance, that the factors underlying a given data set are orthogonal, the explanatory validity of factors selected by Cattell's PP procedure would be questionable, at best.

Demonstrations were made of the successful application of parallel proportional profiles with the orthogonal restriction to synthetic data



of known (orthogonal) factor composition. But the failure of a study using real data raised serious questions about the principle's effectiveness:

Still seeking for living and organic (rather than merely physical) examples that would nevertheless be controllable we were encouraged by the recent discovery of drives as factors to suggest that hunger and thirst be factored from the usual learning and deprivation variables on two populations of rats. This work, by Haverland (Haverland, 1954) had larger variation of food deprivation in one sample and larger variations of water deprivation in the other. The many interesting findings cannot be described here; but it is questionable whether proportional profiles gave a meaningful solution, whereas simple structure gave easily recognisable drive patterns (best oblique, but acceptable even when orthogonal).

(Cattell, 1955, p. 85)

Because of the use of correlations as the base for analysis, Cattell encountered problems with his simple proportionality criterion. "If a variable is loaded by a factor 0.5 and the variance of that factor (still orthogonal) is doubled relative to other factors the loading will not equal 1.0. If we apply the formulae of Thompson and Ledermann (1939) and Thurstone (1947) for predicting the correlations to be expected after the variance of the selection variable has been changed (counting the factor as the variable on which the selection is made), it will be found that not all the loadings of a factor will change in the same proportion when the factor variance is changed" (Cattell, 1955, p. 86).

At the fundamental philosophical level, parallel proportional profiles seemed clearly superior to all other rotation criteria.\* It held out promise, in 1944, of answering the fundamental rotational problem of factor analysis in a satisfactorily general, objective, and persuasive fashion. It would have presented a strong argument of explanatory validity, since it sought to select a factor solution based on a form of invariance which one could expect only in "true, underlying factors".

But parallel proportional profiles has not been developed and successfully used as a principle for factor rotation. It has languished, apparently bogged down in mathematical difficulties and restricted to orthogonal factors. Thus explanatory factor analysis has remained in the unsatisfactory condition described earlier. But (as subsequent sections of this article will show), when Cattell's principle of proportional profiles is incorporated into a revision and generalization of the factor model itself, techniques can be developed for arriving at unique explanatory factor solutions which will allow not only orthogonal and oblique factors, but non-linear ones as well.

---

\* for data showing system variation. This qualification will be developed in Section III.

### III. DEVELOPMENT OF THE THREE-MODE PP MODEL AND OTHER GENERALIZED FACTOR MODELS

Cattell developed his principle of proportional profiles (PP) simply as a criterion for selection among the possible rotations of a pair of classical factor analyses. In fact, however, it provides a fundamental clarification or enrichment of the notion of a "true underlying" factor. As such, it can be considered as a modification of the classical factor model.

Whereas the classical factor model was concerned only with the behavior of factors within a given study, Cattell's contribution was to consider also the expected behavior of factors in a pair of related but differing studies. Proportional profiles is a precise definition\* of the expected behavior of the loadings of "true underlying" factors in a pair of studies where the relative influence of the factors changes.

To use Tucker's terminology (Tucker, 1964) classical factor analysis only considered two "modes of measurement" at a time. Typical of this would be the study of the scores of a number of persons, on a number of measures. The resulting data consisted of a two-way matrix of persons against measures. But the principle of proportional profiles is concerned with persons, on measures, *on two different occasions*. It is fundamentally a three-mode concept.

#### The Three-Mode PP Model

The classical equation of factor analysis could be written, for  $m$  factors, as:

$$(1) \quad x_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \dots + a_{jm}F_{mi} + U_{ji}$$

Here  $x_{ji}$  is a score on measure  $j$  for person  $i$ . Contribution of any factor (e.g. factor 2) to that score, is given by taking the product of  $a_{j2}$  the loading of the  $j^{\text{th}}$  measure on that factor, and  $F_{2i}$  the score of the  $i^{\text{th}}$  person on that factor. These values represent the relative influence of that factor on the  $j^{\text{th}}$  measure and in the  $i^{\text{th}}$  person. The last term,  $U_{ji}$  represents the contribution of a "unique factor" for each data value, and consists of the "specific factor" contribution plus error.

---

\* Assuming, as shall be done here, that conventions for expressing loadings are revised so that proportionality is retained when factors change relative influence (e.g. change from correlations to functions which allow loadings to exceed 1.0). This will be developed later.

The three-mode interpretation of parallel proportional profiles gives a straightforward generalization of this classical factor equation. The generalized equation would predict scores in a three-way data matrix, composed of elements  $x_{kji}$  representing the score for the  $i^{\text{th}}$  person on the  $j^{\text{th}}$  measure on the  $k^{\text{th}}$  occasion of measurement. The contribution of a given factor (e.g. factor 2) to a given score is computed in part as it was before, by multiplying the degree of that factor's influence for that individual,  $F_{2i}$ , times the degree of that factor's contribution to that variable,  $a_{j2}$ . But this estimate of the factor's contribution has to be modified by a coefficient describing the relative influence of that factor on that occasion:  $o_{k2}$ . Thus all the contributions of factor 2 on occasion  $k$  are stepped up or down in a parallel proportional fashion by the coefficient  $o_{k2}$ . The classical factor model becomes generalized as follows:

$$(2) \quad x_{kji} = o_{k1}a_{j1}F_{1i} + o_{k2}a_{j2}F_{2i} + \dots + o_{km}a_{jm}F_{mi} + U_{kji}$$

In this formulation, the principle of proportional profiles applies to any number of occasions, rather than just two. It is no longer a principle for rotation of a classical solution, but instead it has become incorporated into a revision of the basic factor equation itself. Not only will this formulation play a basic role in the development of a PP analysis procedure, but its incorporation of more than two occasions will prove crucial for the successful practical application of PP analysis to data sets involving more than a very few factors (in Section IV, below).

However, this form of the factor model is not completely general. It is not always true that the contributions of the factors from one occasion to the next would vary systematically in this manner. Cattell was aware that the proportionality would not always hold, and he discussed experimental designs in which one might ensure the desired systematic variation in factor influence. His unsuccessful experiment with drives in rats (Haverland, 1954, as cited in Cattell, 1955, p. 85) was an example of an attempt to ensure that the relative influences of the presumed factors of hunger and thirst were systematically different from one group of rats to the other. But, perhaps because of his close association with personality-measurement applications, Cattell did not fully develop the theory of factor analysis to which the parallel proportional model could be applied, and distinguish it from the theory of a different type of factor for which an entirely different type of model would be necessary. This theoretical basis will be developed here in terms of two corresponding conceptual models for factor analysis.

## Two Conceptual Models for Factor Analysis

Behind the *mathematical* model of factor analysis (the fundamental equation describing scores in terms of contributions from factors)

there must lie a *conceptual* model, a logical or semantic interpretation of the terms in such an equation by virtue of which its application to real data is deemed reasonable. Only when factor analysis has its foundation in some conceptual model can one hope that it will tell us something of explanatory value about the world. It would seem, therefore, that one logical approach to a three-mode factor model would be to start at the conceptual level, to apply the classical conceptual model of factors and factor analysis to three modes of measurement, and then from this enlarged *conceptual* model to develop an appropriate generalization of the *mathematical* model. Unfortunately, this is not possible.

It soon becomes apparent that the classic two-way *mathematical* model of factor analysis is, in fact, an ambiguous representation of at least *two* fundamentally different *conceptual* models. Although these conceptual models have equivalent algebraic representation in the two-way case, their algebraic generalizations to three-way and higher order data relationships are substantially different.

For the purposes of discussion, the two basic models will be called (1) the "system-variation" model and (2) the "object-variation" model of latent data influences. In the "system-variation" model, unitary latent influences act across many objects, and thus the factor scores of the objects will vary proportionally across any third mode, (e.g. across occasions of measurement). In the "object-variation" model, latent influences are present separately in each object, and "factors" represent common *types* of these individual influences. Here the factor scores of each object will vary independently of the variations of those scores in other objects, when measured across a third mode. It will be useful, before proceeding further, to develop a clear picture of these two conceptual models.

#### (1) The System-Variation Model

Consider the example of an economic system. For a set of data, different individual businesses might be measured on a number of variables (e.g. production, investment, salaries, layoffs, etc.) on each month throughout a five year period. In such a set of data it would be reasonable to look for factors present in the system as a whole, operating across a number of individual businesses. One such system-factor might be described as "inflationary pressure". When this factor increased from one month to the next, its contribution to individual scores in the data matrix would rise in different businesses in proportion to their loadings on that factor. It would appear (in terms of the classical factor model) that their factor scores had all increased proportionately. The loadings of the individual businesses ( $F_{1i}$ ,  $F_{2i}$ , ...) on that factor (classically called factor scores) would here represent the relative susceptibility of each business to the influence of that factor. The loadings of the different months ( $\alpha_{k1}$ ,  $\alpha_{k2}$ , ...) would represent the relative degree of activity or influence of that factor in the system as a whole for that month (in this case the fluctuations in inflationary pressure in the

economy from month to month). The loadings on the variables ( $a_{j1}$ ,  $a_{j2}$ , ...) would represent the degree to which each type of measurement was influenced by (or reflected) changes in that factor (in this case, the degree to which layoffs, salaries, etc. are influenced by "inflationary pressure"). The system-variation conceptual model would be appropriate for such data, and thus it would be appropriate to apply the parallel proportional profiles mathematical model for three-mode factors which was developed earlier in this section.

Another source of data which could be expected to show system-variation of factors would be any biological system, for example the brain of a laboratory animal. Electrodes could be placed in different brain centers, and the EEG activity recorded from each electrode could be subject to spectral analysis. Sets of these spectra could be measured for each of a number of different occasions, yielding a three-way data set of measured spectral intensities. Its modes would be frequency by brain center by occasion. The system-factors sought in such an experiment would be the hypothetical "generators" of EEG activity within the brain. Each factor might correspond to one such generator. It would have large loadings on those spectral frequencies corresponding to the type of EEG "wave" it generated. Its factor loadings in different brain centers would describe the relative influence of that generator in the brain activity at those locations: the "effective proximity" of those brain centers to the generator. The loadings on different occasions would represent the relative "activity" of that generator across those occasions. Certain EEG generators could be expected to be much more active during certain stages of sleep, others during alert goal-oriented behavior, etc. Again, the factors are acting *in* the system, *across* "individuals". The loadings of "individuals" (e.g. brain centers) represent their different susceptibilities to these common latent influences. Therefore it is reasonable, with this conception of brain activity, to expect parallel proportional changes in factor influence across a third mode of measurement as the latent factors change in their relative degree of activity or influence.

Other likely examples of system-variation data might be found in studies of languages, weather patterns, ecological systems such as a forest, and social systems. The key feature of the factor model in all these systems is that the factors are interpreted as single, unitary influences, existing in the system as a whole, and through the system influencing individuals in a coordinated fashion.

The system variation concept will immediately generalize to cases of four-mode and higher order relationships. A four-mode application might arise, for example, if an attempt were made to analyze (1) a number of measurements of (2) a number of different speech sounds, as made by (3) a number of different individuals, and as occurring in (4) a number of different languages. If  $c_{pn}$  is taken to represent the loading of the  $n^{\text{th}}$  factor on the  $p^{\text{th}}$  condition of measurement, then a four-mode linear model for system-variation would be written:

$$(3) \quad x_{pkji} = c_{p1} a_{k1} a_{j1} F_{1i} + c_{p2} a_{k2} a_{j2} F_{2i} + \dots + c_{pm} a_{km} a_{jm} F_{mi} + U_{pkji}$$

## (2) The Object-Variation Model

Now consider, by contrast, the example of personality or aptitude measurement. In a typical personality study a group of several hundred persons are administered questionnaires containing items like "When I get angry, I throw things." and are asked to rate how much each item applies to them. The result is a two-way data set of answers by persons. If one re-administers the questionnaires to the same individuals on a number of different occasions, a three-way data set (answers by persons by occasions) will be generated. But the properties of such a data set are fundamentally different from the system-variation examples previously described, and therefore require a different conception of "factors" and their relation to the data.

In cases such as personality or aptitude measurement, the "factors" are not conceived of as residing in the *system*, but instead in the *objects* or individuals being measured. Such "factors" are not unitary influences across individuals, but are merely "unitary" across a set of measures within a given individual. The meaning of extracting factors from a whole population of individuals lies in identifying characteristic *types* of latent influences within individuals or objects. When a researcher says "a factor of hostility clearly emerged in this population," he means that a characteristic *type* of influence was identified. The factor of "hostility" occurs in a distinct, separate way in each individual. There is no reason to suppose that the "hostility" factor scores of two different persons would necessarily vary in parallel from one occasion of measurement to the next. This concept of the data and its variation is here called the object-variation model for factor analysis.

For two-mode factor analysis, the *mathematical* model is the same, whether the underlying conception is of a system-variation or object-variation case. In both cases the loadings of the individuals on the factors (the factor scores) represent the influence of the given factor for that particular individual. It does not matter whether that factor is conceived of as a unitary influence acting simultaneously on all the individuals of the system, or whether the "factor" is in fact conceived of as a type of influence which occurs separately within each individual. In both cases the  $(F_{1i}, F_{2i}, \dots)$  coefficients suffice for expressing in the two-way model the degree of the factor's influence in a given individual.

This mathematical equivalence of the two conceptual models breaks down for three-way and higher order data variation. As noted earlier, this is because the system-variation model predicts a coordinated variation of factor influence across a third mode (and any fourth or higher mode), whereas the object-variation model predicts that the changes in

the degree of expression of a given factor across any third mode should be uncorrelated from one individual to the next.

Equation 2 gives a mathematical model appropriate for a linear three-mode application of the system-variation concept. The corresponding linear three-mode model can be constructed to express the object-variation concept, as follows: the difference between the influence of a given factor on one occasion and any other consists of a number of uncoordinated shifts (generally small) in the amount of this factor in some of the individuals; this small shift or variation "V" in the factor score of a given factor for individual *i* on occasion *k* is symbolized  $V_{kli}$  for factor 1,  $V_{k2i}$  for factor 2, etc. The general object-variation expression for a given data value then becomes (in the linear, three-mode case):

$$(4) \quad x_{kji} = a_{j1}(F_{1i} + V_{k1i}) + a_{j2}(F_{2i} + V_{k2i}) + \dots + a_{jm}(F_{mi} + V_{kmi}) + U_{kji}$$

This model resembles a set of independent two-way factor models with respect to the factor scores of individuals, but it is a common model with respect to the factor loadings on variables. In this weak general form it will have no unique solution determined by the data. To try to secure a unique, "explanatory" solution, an additional constraint must be placed on the *V* values. Is there a criterion which can be applied to take advantage of the additional information supplied by the third mode, without demanding parallel proportional relationships which would only be found in system-variation data?

One possible constraint, suitable for many sets of object-variation data (including personality questionnaire data) can be derived from considerations of continuity. Even when an individual's variations in factor scores are not related to those of *other* individuals, they are still related to his *own* previous scores, if the occasions of measurement are "sufficiently close" to one another in the third mode (e.g. sufficiently close in time if the third mode covers successive occasions of measurement). Just how close is "sufficiently close" will depend, of course, on the rates of change expected in the factors underlying the particular data. Perhaps this could be estimated from the rates of change of the variables themselves. If the occasions are "sufficiently close" to one another, then the shifts or *V* values, measured from one occasion to the immediately prior or succeeding one, should always be small, even when the cumulative shifts over many occasions might be large for some individuals. We could expect most types of underlying factors to show this sort of continuity over samples taken at sufficiently short intervals. Therefore a reasonable additional constraint upon the *V* values of equation 4 would be the condition:

$$(5) \quad \text{minimize} \quad \sum (V_{kfi} - V_{(k-1)fi})^2$$

This elementary continuity measure has many shortcomings, and we shall immediately take advantage of an insightful and highly related paper by Shepard and Carroll (Shepard and Carroll, 1966, p. 581) and borrow their adjusted criterion (translated into our notation):

$$(6) \quad \text{minimize} \quad \sum_k \sum_{k'} \left[ \frac{V_{kfi} - V_{k'fi}}{k - k'} \right]^2 w_{kk'}, \quad k \neq k'$$

where  $w_{kk'}$  is some monotone decreasing function of  $k - k'$ . In effect, this considers not only immediately adjacent factor scores but (to a lesser degree) scores in the vicinity, as determined by the  $w_{kk'}$  weighting function.

Whereas the proportional profile system-variation model will be studied in detail in Section IV, the actual properties of the object-variation model have not yet been investigated. The model is developed here because of its great theoretical interest and its relations to the system-variation case. As an initial comment, however, it seems clear that the factor scores of a number of individuals would have to shift significantly and in independent fashion in order for a unique solution to be obtained with the proposed continuity criterion. Since no tests have been conducted, it is not known how effective such a criterion would be, and under what conditions it could be expected to give an "explanatory" unique solution. But it seems definitely to merit further investigation.

#### Mixed Types of Variation

With some types of data, one might expect significant effects involving both system- and object-variation. Consider, for example, studies of the development of human abilities. As children grow older, general systematic changes can be expected to occur in the relative influence and expression of different ability factors. There will also, however, be variations in these factors for each individual due to the unique circumstances, history, and genetic complement of that child. A similar interplay of system-variation and object-variation would be expected in measurements of individuals going through a series of learning trials, mastering some complex skill.

One simple way of analyzing data containing such mixed variation patterns might be to ignore the object-variation and merely try to "fit" the system-varying factors. If, in a given case, all factors of interest could be expected to show sufficient system-variation, then the object-variation could be considered part of the "noise" or error of the model's fit to the data. On the other hand, if techniques for analysis of object-variation were successfully developed (along the lines suggested earlier



in this section), and if such an analysis were performed on mixed type data, then both the object- and the system-variation would be described by the tables of "V" values for each factor. An analysis of these tables would reveal precisely how much variation of each type was present.

The concept of analysis of the "V" tables needs to be explored more fully. When considering the possible structures of these tables one becomes aware of possible complex patterns which could not be attributed to either system-variation or object-variation, or even to any "mixture" of these two cases. These complex new types of possible data variation pose difficult questions of interpretation, apparently requiring new conceptual models. But since the appropriate models for such situations appear to be strongly related to Tucker's concepts developed for his three-mode factor analysis (Tucker, 1963, 1964, 1966), further discussion of these patterns will therefore be deferred until Section VII, where they will be discussed in relation to Tucker's approach.

The object-variation model given in equation 4 allows for change in the amount of a factor in any individual or object, but retains constant factor loadings on the variables. Even this restriction will sometimes seem inappropriate. It might sometimes be the case (e.g. in the child development example given above) that one could expect shifts in the loadings of the factors on the measures, as well as on the individuals. As children grow, they might alter their interpretation of a given question, or there might be shifts in the degree to which certain abilities are used to solve a given problem. A more general model of object-variation might therefore be written:

$$(7) \quad x_{kji} = (a_{j1} + V'_{klj})(F_{1i} + V_{kli}) + (a_{j2} + V'_{k2j})(F_{2i} + V_{k2i}) + \dots \\ \dots + (a_{jm} + V'_{kmj})(F_{mi} + V_{kmi}) + U_{kji}$$

With this generalized object-variation model, the continuity constraint would have to apply to both the  $V$  values of factor score variations, and the  $V'$  values of factor loading variations.

The generalized object-variation model of equation 7 suggests that a classic factor analysis might be performed on each of the successive occasions of a three-way data matrix, and then an attempt should be made to rotate these separate solutions into the greatest degree of agreement with one another. Techniques have been developed for such rotation of two sets of factors (Wrigley and Neuhaus, 1955; Horst, 1965) and all that might be needed for their application here is a generalization of the rotation criterion from that of best agreement between *two* sets of factors, to the most continuous agreement of factors over the entire series of occasions ("most continuous" defined as in the continuity criterion in expression 6).

### Extending the Application of the PP Model

The distinction developed in this section between system-variation and object-variation would seem to rule out applying the PP model to many interesting types of data sets, for example personality test data. The same apparent restriction confronted Cattell in his development of the original proportional profile rotation criterion. One of his primary interests was the application to personality test data, and so he devised ways in which one could deliberately impose the required systematic variation on what would otherwise be object-variation type data (Cattell, 1944, p. 275). Though not all of these techniques apply to our generalized model, several of his suggestions are quite relevant. The basic idea is simply to alter the conditions under which measurements are made on the successive occasions in such a way that the relative influence of different factors will change for all individuals. Administering a test under speeded vs. unspeeded conditions, under different "sets", using different instructions, etc. might all be used to impose overall enhancement or suppression of particular factors. For further discussion the reader is referred to Cattell (1944, p. 275).

#### IV. TESTING THE THREE-MODE PROPORTIONAL PROFILES MODEL

The PP model has been tested in two basic ways: (1) by using it to analyze synthetic data of known factor composition; (2) by applying it to real data. The tests with synthetic data helped to determine the behavior of the model and of the analysis technique with different size data sets and different latent factor structures. The conditions in which uniqueness breaks down and the minimal requirements of an "adequate" data set were determined in a trial and error fashion. This knowledge then guided the applications of the model to real data.

##### Synthesis of the Data

In order to judge whether or not PP three-mode factor analysis can correctly extract the true latent structure of a set of data, it is of course necessary to know in advance what that latent structure is. To this end, a FORTRAN IV computer program was written to generate synthetic data. Input to the program was a description of the factors and the dimensions of the desired data set, output was the set of data generated by those factors. Exact error free data was created, since this simplified the recognition of optimal fit and allowed the initial tests to concentrate on those aspects of the solutions determined by the data structure rather than by error variations. Specification of the factors used to generate a given set of data consisted simply of specifying the loadings of those factors on each measure, on each individual or object, and on each occasion.

The term "loading" as applied to the PP model, should not be taken to mean more than simply a weighting coefficient, describing the variations of relative influence of a given factor from one occasion to the next (or one person or variable to the next). This use is a generalization of the term "loading" as it normally occurs in factor analysis. Since the analysis procedure developed below completely bypasses the computation of correlation coefficients, either among components of the data or among data and factors, the mathematical properties of the "loadings" developed by the PP analysis are not the same as those of conventional loadings. This is particularly significant since (as Cattell discovered) strict proportionality does not hold across conventional loadings when a factor increases its relative influence (Cattell, 1955, p. 86). In contrast, the "loadings" used in the PP analysis do maintain proportionality as a factor changes its relative influence. These "loadings" are simply the weights which would occur in equations, such as 8, predicting the raw data from the factors.

A set of factor loadings input to the program generating synthetic data is given in Table 1, and the resulting data in Table 2. The data was synthesized from the input loadings as follows:

TABLE 1

## Typical Loadings Used to Generate Synthetic Data

Occasion Loadings

	Factor 1	Factor 2	Factor 3	Factor 4
1	1.000	1.000	1.000	1.000
2	1.500	2.150	0.0	0.100
3	1.000	0.350	0.300	0.400
4	0.500	0.900	1.200	0.300
5	0.900	0.500	1.300	1.100
6	1.000	1.100	2.800	2.300
7	1.100	1.000	0.200	1.800
8	0.0	1.000	1.100	0.500
9	0.900	0.0	0.600	0.700
10	2.100	2.000	1.500	1.800

Person Loadings

	Factor 1	Factor 2	Factor 3	Factor 4
1	1.000	0.600	1.000	2.000
2	1.100	0.700	1.400	0.600
3	0.900	0.800	1.300	0.700
4	1.200	0.900	1.200	0.800
5	0.800	1.000	1.100	0.900
6	1.300	1.100	1.000	2.000
7	0.700	1.200	0.900	0.900
8	1.400	1.300	0.800	0.800
9	0.600	1.400	0.700	0.700
10	1.000	1.000	0.600	0.600

Test Loadings

	Factor 1	Factor 2	Factor 3	Factor 4
1	-1.100	0.010	0.120	0.110
2	1.000	0.030	0.210	0.070
3	4.000	0.030	0.120	0.030
4	0.190	3.000	0.320	0.600
5	0.230	7.190	0.500	0.110
6	0.010	0.740	5.250	0.130
7	0.500	1.500	9.500	1.500
8	-0.900	0.200	2.100	0.300
9	0.010	0.100	0.030	9.990
10	0.110	-0.200	-0.800	7.640

A typical set of factor loadings used to generate synthetic data. Note the absence of orthogonality and simple structure.

TABLE 2

## Data Computed for Occasion 1

1	-0.754	1.368	4.198	3.510	5.264	5.964	13.900	1.920	20.080	14.470
2	-0.969	1.457	4.607	3.117	6.052	7.957	15.800	2.270	6.117	3.445
3	-0.749	1.246	3.801	3.407	6.686	7.517	15.050	2.290	7.121	4.247
4	-1.079	1.535	4.995	3.792	7.435	7.082	14.550	1.860	8.130	5.104
5	-0.639	1.124	3.389	4.044	8.023	6.640	13.700	2.060	9.132	5.884
6	-1.079	1.683	5.413	5.067	8.928	6.337	14.800	1.750	20.133	14.403
7	-0.551	0.988	2.971	4.561	9.338	5.737	12.050	1.770	9.145	5.993
8	-1.343	1.663	5.759	4.902	10.157	5.280	11.450	0.920	8.160	5.366
9	-0.485	0.838	2.547	4.958	10.631	4.808	10.100	1.420	7.160	4.574
10	-0.952	1.198	4.120	3.742	7.786	3.978	8.600	0.740	6.122	4.014

## Data Computed for Occasion 2

1	-1.622	1.532	6.024	2.205	4.681	0.485	1.950	-1.170	2.073	1.573
2	-1.801	1.675	6.623	2.449	5.419	0.542	1.965	-1.327	0.686	0.500
3	-1.469	1.379	5.426	2.698	6.070	0.615	1.980	-1.034	0.793	0.523
4	-1.962	1.833	7.229	3.090	6.894	0.694	2.370	-1.416	0.907	0.629
5	-1.300	1.236	4.833	3.282	7.476	0.764	2.235	-0.853	1.011	0.620
6	-2.112	1.997	7.839	3.790	8.379	0.859	2.925	-1.475	2.127	1.522
7	-1.133	1.092	4.239	3.853	8.879	0.910	2.460	-0.678	1.030	0.563
8	-2.288	2.145	8.441	4.347	9.839	0.993	3.120	-1.606	0.950	0.582
9	-0.968	0.947	3.644	4.413	10.281	1.054	2.655	-0.509	0.848	0.354
10	-1.633	1.534	6.032	3.321	7.542	0.763	2.340	-1.132	0.714	0.423

.

.

.

## Data Computed for Occasion 9

1	-0.764	1.124	3.714	1.203	0.661	3.341	8.250	0.870	14.013	10.315
2	-0.942	1.196	4.073	0.709	0.694	4.474	9.105	0.999	4.231	2.646
3	-0.743	1.008	3.348	0.697	0.630	4.167	8.550	1.056	4.927	3.209
4	-1.040	1.270	4.423	0.772	0.670	3.864	8.220	0.708	5.627	3.821
5	-0.643	0.903	2.978	0.726	0.565	3.554	7.575	0.927	6.321	4.364
6	-1.061	1.394	4.794	1.254	0.723	3.344	8.385	0.627	14.016	10.345
7	-0.559	0.787	2.604	0.670	0.484	2.923	6.390	0.756	6.316	4.450
8	-1.267	1.400	5.114	0.729	0.591	2.605	6.030	0.042	5.621	4.033
9	-0.490	0.662	2.225	0.531	0.388	2.274	4.995	0.543	4.913	3.467
10	-0.901	1.005	3.656	0.538	0.433	1.954	4.500	0.072	4.216	3.020

## Data Computed for Occasion 10

1	-1.722	2.703	8.724	6.639	10.257	9.252	22.500	2.580	36.150	26.295
2	-2.156	2.869	9.566	5.959	11.766	12.224	24.825	2.935	11.015	6.545
3	-1.690	2.436	7.880	6.539	13.052	11.604	23.760	3.092	12.825	7.954
4	-2.380	3.053	10.393	7.319	14.580	10.994	23.220	2.304	14.645	9.479
5	-1.452	2.200	7.027	7.819	15.770	10.370	21.945	2.839	16.450	10.842
6	-2.405	3.363	11.274	9.759	17.592	9.998	24.315	2.213	36.256	26.164
7	-1.253	1.939	6.163	8.883	18.447	9.089	19.590	2.478	16.479	10.978
8	-2.906	3.371	12.025	9.607	20.129	8.441	18.930	0.826	14.711	9.845
9	-1.093	1.653	5.288	9.731	21.085	7.761	16.695	2.009	12.911	8.365
10	-2.063	2.425	8.600	7.335	15.432	6.366	14.220	0.724	11.037	7.362

Data synthesized with the latent structure defined by Table 1.

$$(8) \quad x_{kji} = O_{k1} I_{j1} M_{i1} + O_{k2} I_{j2} M_{i2} + \dots + O_{km} I_{jm} M_{im}$$

where the "O" values represent weights for the different occasions, the "I" values represent weights for the different individuals, and the "M" values represent weights for the different measures.

A relativity of scale is apparent in the PP model. Consider equation 8, which generates data according to that model. If all of the "O" values for any factor are doubled, and all the "I" values are halved, the resulting data is unchanged. In general, any factor could have all of its loadings for one mode multiplied by some constant, and so long as all its loadings on one of the other modes were divided by the same constant, the results would be equivalent. Such an operation changes the factor's absolute scale in any given mode, but it does not alter the relationships among the loadings within any mode. Therefore, all these changes of scale yield descriptions of the factors which are completely equivalent with respect to interpretation.

It was decided to adopt a scaling convention so that solutions with the PP model would not be subject to this trivial type of indeterminacy. For this purpose, the average value of the influence of any factor, across all "occasions", was taken to be 1.0. Similarly, the average amount of influence of any factor, taken across all "individuals", was also set to 1.0. Thus a loading of 2.0 on either of these two modes meant that a factor was twice as active as average on that occasion or in that individual, 0.5 meant half as active as average, etc. Loadings on the third mode, measures, would therefore give directly the average contribution of that factor to the predicted score of that measure. The input specifications of factors for the data synthesis program (Table 1) were selected to follow these conventions.

#### PARAFAC: The Computational Technique for Data Analysis

The PP three-mode factor model (2) is sufficiently different from the classical factor model (1) to make conventional factor analysis algorithms inapplicable. An analysis procedure called PARAFAC (for *parallel factors*) was developed to accomplish the required analysis. One objective of this work has been to develop a technique which is general enough to fit both the PP model and a family of related models to real data. For this reason, an approach was adopted which involved direct optimization of the parameters of the model. As Green points (Green, 1966, p. 440), "Nearly all psychometric procedures can be looked upon as attempts to optimize some criterion ... Factor analysis seeks the minimum number of parameters to describe the maximum amount of inter-correlation among the variables. Whenever one is fitting a model to data, one seeks parameters of the model that fit the data as closely as possible -- parameters that optimize some measure of fit." In this case, the PARAFAC analysis procedure tries to fit the mathematical model of (2) to the data as closely as possible, using the least-squares criterion of fit. This is similar to the approach recently

developed by Harman and Jones (1966) for conventional factor analysis models. The  $U$  term of equation 2 is here taken as equivalent to the error of estimate, i.e.  $(x - \hat{x})$ . In a sense, then, it "drops out" of the model. The equation used for computation of the estimated data values was

$$(9) \quad \hat{x}_{kji} = \sum_m O_{km} I_{jm} M_{im}$$

and it was required that the parameters (the  $I$ ,  $O$ , and  $M$  values) minimize the sum of the  $(x - \hat{x})^2$  residuals. Unlike Harman and Jones (1966), this technique estimates three sets of parameters (or factor "loadings"), rather than just one. As a result, a large number of parameters have to be "fit" simultaneously to the data. For a matrix of 10 measures, on 10 individuals, on 10 occasions, there are 1000 data points, and for each factor fit to such a data set there are 30 loadings to be estimated (10 weights for occasions, 10 weights on individuals, and 10 weights on measures). For four factors, a total of 120 parameters have to be estimated simultaneously. Such large problems can bog down conventional optimization techniques based on gradients.

Fortunately, a powerful optimization routine was made available for this work by Hans Reichenbach. The program is called Routine TBU, and is thoroughly described in a 400+ page manual and theoretical article (Reichenbach, 1969) of which only an earlier and more primitive version has been published (Reichenbach, 1962). A common cause of convergence difficulties with gradient techniques occurs when the routine becomes trapped on a "thin ridge" in its upward path on the multi-dimensional hypersurface. In order to avoid this difficulty, Routine TBU uses trends in the changes of the variables rather than gradients. A complete iteration of this routine contains a section where each variable is changed individually as a function of its current short-range and long-range trends, and another section where all variables are changed simultaneously according to some optimal proportion of their trends. The routine alternates between these two methods of changing variables. Both Harman and Jones (1966) and Joreskog (1966) adopted modifications of the steepest descent technique to improve convergence, and these modifications are related in different ways to features of TBU. The reader interested in further details on TBU is referred to Reichenbach (1962, 1969).

TBU proved successful in optimizing 30-120 variable models involving one to four factors, but convergence became markedly slower as the number of factors increased. On the IBM 360/91 computer, one factor could be fitted to a 10 by 10 by 10 data set in a few seconds. Two factors would often require twenty seconds, three factors one hundred seconds, and four factors several hundred seconds or more, depending on the characteristics of the data set. Beyond four factors, convergence took too long to be practical. This working range of 1-4 factors provided sufficient leeway to explore many of the theoretical properties of the model, such as the

conditions for a unique solution and the effects of extracting more or fewer factors than were used to create the data. But for practical application and for more general theoretical studies, an improvement of the algorithm was clearly necessary.

Dr. Robert Jennrich of the UCLA Department of Biomathematics proposed a special purpose computation algorithm for the solution of the *linear* PP three-mode model which proved to be much simpler and faster than the generalized approach which was being used. His development of this algorithm was as follows:

Given the representation

$$(10) \quad x_{ijk} = \sum_l O_{il} I_{jl} M_{kl} + E_{ijk}$$

we seek values of  $O_{il}$ ,  $I_{jl}$ , and  $M_{kl}$  which minimize the sum of squared errors  $\sum_{ijk} E_{ijk}^2$ . For fixed values of  $O_{il}$  and  $I_{jl}$  the selection of values  $M_{kl}$  which minimize this sum is a simple multiple linear regression problem whose solution is:

$$(11) \quad M_{kl} = \sum_m x^{lm} y_{mk}$$

where

$$(12) \quad x_{lm} = \left( \sum_i O_{il} O_{im} \right) \left( \sum_j I_{jl} I_{jm} \right) ,$$

$$(13) \quad y_{lk} = \sum_{ij} O_{il} I_{jl} x_{ijk} ,$$

and the matrix  $(x^{lm})$  is the inverse of the matrix  $(x_{lm})$ . Similar formulas give  $I_{jl}$  in terms of  $O_{il}$  and  $M_{kl}$  and  $O_{il}$  in terms of  $I_{jl}$  and  $M_{kl}$ . A complete algorithm consists in solving for the matrices  $(O_{il})$ ,  $(I_{jl})$  and  $(M_{kl})$  in order and repeating until convergence is obtained.

This algorithm has been found to converge in less than ten seconds for many data sets involving four to six factors. But in other cases (particularly certain types of data sets to be described below) convergence would proceed only to a certain point, beyond which a large number of iterations would produce little improvement of fit to the data.

To cope with this difficulty, a relaxation factor was added to the computation procedure. Examination of the behavior of the factor loadings after convergence slows down reveals that they continue to move in small, steady steps, usually in the direction of the desired unique solution.



In order to increase the size of these steps, a modification is made in the computation given in equation 11. An adjusted estimate  $M^*$  is computed for each factor loading according to the relation

$$(14) \quad M^*_{kZ} = M'_{kZ} + R(M_{kZ} - M'_{kZ})$$

where  $M_{kZ}$  is the value for the factor loading which would be computed for this step according to the unmodified procedure of equations 10 through 13, and  $M'_{kZ}$  is the old value for the loading computed during the last iteration.  $R$ , the relaxation factor, is assigned a value between 1.0 and 2.0.

The effect of this modification is to push each loading further in the direction that it has been moving, beyond the value which is optimum simply for this cycle, and hopefully towards the final overall optimum value. Although a relaxation factor seems to provide a small gain, since it increases any step size by less than 2.0, under the proper conditions it has a "snowballing" effect, since the resulting adjusted loadings are used as the basis for the unadjusted values computed in the next cycle, and the values of that cycle are then themselves adjusted, thus moving even further in the desired direction. Thus the effect of a small relaxation factor can continue to multiply itself on successive iterations.

Modifications to this algorithm are still being explored, and the results are as yet tentative and incomplete. The effectiveness of the procedure depends critically on both the size of the relaxation factor and the characteristics of the data set. To date, the best overall value for the relaxation factor seems to be in the range 1.2-1.3. Using this modified algorithm, it often becomes possible to push the residuals to very low limits until a point is reached where the factor loadings themselves cease to change. The ability to push analysis much closer to true convergence and a perfect fit to the data has led to the discovery of unique solutions where they were previously not believed to exist. This has forced a reevaluation of the minimal necessary conditions for uniqueness (to be discussed below). Therefore, even though these new algorithms are still in the process of development and only incomplete results are available, these results are significant enough to warrant including them in the discussion which follows.

#### The Analysis of Synthetic Data

Approximately 30 tests were performed by analyzing synthetic data using the optimization technique of Routine TBU. In these tests the number of factors in the data was varied from two to four, and the number of factors extracted ("fit" to the data) was varied from 1 to 5, ranging both below and above the "true" number of factors used to construct the data. The size of the data set was varied systematically, the number

of intervals in any mode ranging from 1 to 10. The pattern of factor loadings was initially all positive, but negative and zero loadings were subsequently introduced to ensure generality. As can be seen from inspection of a typical example given in Table 1, the factors were not generally orthogonal and did not display simple structure. They were selected to conform to the scale convention discussed earlier, and were varied so that a row of factor loadings would not repeat the relations of any other row in that mode.

After the development of the quick algorithm for the linear case, approximately 50 additional tests were performed with synthetic data. In these experiments the number of factors in the data ranged as high as 10, and the dimensions of the data set were expanded to a maximum of 20 by 20 by 20. In most of these experiments the factor loadings were generated from a sequence of uniform pseudo-random numbers\* between 0.0 and 2.0 (before normalization). Since this left the factors in general uncorrelated, specific tests were performed with correlated factors. These factors were constructed by adding a percentage of each loading of a given factor to the loading of the factor with which it was to be correlated.

Because error free data was used, classification of the resulting solutions was usually clear cut. Complete fit was recognizable by near-zero residuals (from 1% to .01%, depending on the amount of computer time used). A solution matched the "true" latent structure if its factors had the same loadings within 0.1% (or, when convergence was carried further, to within a few hundredths of a percent). The columnar order of factors might differ from that used on input to create the data, but factors were easily "identified" regardless of their column. A "match" was not expected to exhibit the same columnar order as the factors used to create the data.

Initially, the test of uniqueness was the consistent finding of a matching solution regardless of the starting values for the loadings. This required repeating each test from several different random starting configurations. This condition was eventually dropped and the discovery of a perfect matching solution with one starting configuration was taken to demonstrate that this match would be found with any starting configuration (since the likelihood of stumbling on the one perfect match by chance when the solution was not unique seemed quite small, and earlier results showed that when any solution matched, all solutions with that data and that number of factors would match, regardless of the starting configuration). Non-unique solutions were always found to clearly diverge from the "true" loadings.

---

\* This was accomplished by using the RANDM function included as an addition to the FORTRAN IV level G compiler by the UCLA Campus Computing Network.

Since decisions of uniqueness and good fit were non-controversial, individual sets of "true" and "reproduced" loadings will not, in general, be reproduced here. Instead, results from the 80 experiments with synthetic data will be summarized according to those conditions which were found to affect the explanatory validity of the solutions. Three things should be considered in this regard: (1) the effect on the solution of the number of factors chosen to be extracted from a given data set; (2) the data characteristics necessary to adequately determine the correct solution for a given latent factor structure; (3) the data characteristics which affect the ability of the analysis procedure to reach convergence, and the properties of different types of solutions prior to convergence.

#### A. Effects of extracting different numbers of factors

Before performing any analysis, one must decide how many factors to extract from the data, i.e. how many factors to specify in the model which is to be fit to the data. An understanding of the consequences of extracting the correct number of factors, versus too few or too many factors, is essential for interpretation of the solutions obtained. On the basis of such information one can then revise the estimate of the number of factors present, and reanalyze the data.

(1) If the number of factors extracted from the data is the same as the number of factors actually present (i.e. used on input to create the data) then the solution is unique and "true" in that it exactly discovers the latent factor structure built into the data. (This uniqueness holds, however, only if the data set is "adequate". The conditions of data adequacy will be discussed below.) The factor loadings need not fall into any simple structure or other natural pattern. The factors may be orthogonal or oblique. An example of a set of factors which are neither orthogonal nor displaying simple structure, yet which were correctly recovered from both a 1000 and a 64 point data set is given in Table 1. (The 64 point solution was for the first 4 intervals in each mode.)

(2) If fewer factors are extracted than are present in the data, and the data is adequate, then the solution is unique, but (of course) incorrect. The resemblance between the recovered solution and the true factor pattern will depend on how "wrong" the estimate of the number of factors was. In a limiting case, if the true number of factors was only 1 larger than the number extracted, and one of the factors used to create the data made only small contributions to the data values, then the resulting solution will closely approximate the true solution except that the small factor will be omitted. In less ideal cases, the extracted factor structure will be proportionately more distorted, as it tries to account for the contributions of the additional (missing) factors within the fewer factors allowed on analysis. Examples for three and four factors are reproduced in Tables 3A and 3B.

(3) If more factors are extracted than are present in the data (i.e. used to create the data), the solution which is extracted is not

TABLE 3A

## Distortions Due to Extracting Too Few Factors

## Occasion Loadings (Input)

	Factor 4	Factor 3	Factor 1	Factor 2
1	1.724	1.315	1.020	1.020
2	0.172	0.0	1.531	2.194
3	0.690	0.395	1.020	0.357
4	0.517	1.579	0.510	0.918
5	1.897	1.710	0.918	0.510

## Person Loadings (Input)

	Factor 4	Factor 3	Factor 1	Factor 2
1	1.620	2.258	0.980	1.078
2	1.268	0.161	1.078	0.980
3	0.352	0.887	0.0	0.980
4	0.493	0.484	0.882	0.0
5	1.268	1.210	2.059	1.961

## Test Loadings (Input)

	Factor 4	Factor 3	Factor 1	Factor 2
1	1.647	0.942	1.000	0.600
2	0.494	1.319	1.100	0.700
3	0.577	1.225	0.900	0.800
4	0.659	1.131	1.200	0.900
5	0.741	1.037	0.800	1.000

\* \* \*

## Occasion Loadings (Extracted)

	Factor 1	Factor 2	Factor 3
1	2.068	1.424	0.961
2	0.054	-0.340	2.048
3	0.234	0.124	0.720
4	0.335	1.869	0.668
5	2.309	1.922	0.603

## Person Loadings (Extracted)

	Factor 1	Factor 2	Factor 3
1	1.621	3.424	1.398
2	1.268	-0.618	0.906
3	0.176	0.731	-0.079
4	0.575	-0.147	-0.285
5	1.359	1.610	3.061

## Test Loadings (Extracted)

	Factor 1	Factor 2	Factor 3
1	1.285	0.409	0.739
2	0.518	0.549	0.800
3	0.504	0.521	0.763
4	0.632	0.462	0.993
5	0.561	0.447	0.849

Three factors extracted from data generated by four input factors. Many of the loadings are similar despite the too few factors extracted. Note, for example, the correspondence between the input loadings for factor 4, and the extracted loadings for factor 1 (factors are labeled according to their position in the extracted matrix, but their columnar position has been rearranged to facilitate comparison with the input factors).

TABLE 3B

## Distortions Due to Extracting Too Few Factors

Occasion Loadings (Input)

	Factor 3	Factor 2	Factor 1
1	1.316	1.020	1.020
2	0.0	2.194	1.531
3	0.395	0.357	1.020
4	1.579	0.918	0.510
5	1.710	0.510	0.918

Person Loadings (Input)

	Factor 3	Factor 2	Factor 1
1	2.258	1.078	0.980
2	0.161	0.980	1.078
3	0.887	0.980	0.0
4	0.484	0.0	0.882
5	1.210	1.961	2.059

Test Loadings (Input)

	Factor 3	Factor 2	Factor 1
1	0.942	0.600	1.000
2	1.319	0.700	1.100
3	1.225	0.800	0.900
4	1.131	0.900	1.200
5	1.037	1.000	0.800

\* \* \*

Occasion Loadings (Extracted)

	Factor 1	Factor 2
1	1.509	1.050
2	-0.306	2.155
3	-0.103	0.580
4	1.876	0.551
5	2.023	0.664

Person Loadings (Extracted)

	Factor 1	Factor 2
1	3.924	1.511
2	-0.927	0.808
3	0.425	-0.592
4	-0.314	-0.863
5	1.891	4.136

Test Loadings (Extracted)

	Factor 1	Factor 2
1	0.287	0.417
2	0.466	0.525
3	0.419	0.486
4	0.391	0.646
5	0.337	0.527

Two factors extracted from data created by three factors. The extracted factors show less correspondence to the input factors than in Table 3a.

TABLE 3C

## Non-uniqueness Due to Extracting Too Many Factors

## Occasion Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	0.306	0.066	1.700	1.772
2	1.835	2.103	0.142	2.048
3	0.143	0.994	0.622	1.100
4	0.572	0.281	2.453	0.066
5	2.144	1.557	0.084	0.015

## Person Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	1.817	0.354	1.187	0.988
2	0.629	0.741	1.110	0.854
3	0.791	1.255	0.223	1.235
4	0.996	0.867	1.247	0.969
5	0.766	1.783	1.233	0.954

## Test Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	0.368	0.677	0.556	1.623
2	0.345	0.312	1.751	2.033
3	0.309	0.382	1.746	0.277
4	0.294	0.408	0.346	2.186
5	0.013	0.853	1.378	2.559

\* \* \*

## Occasion Loadings (Extracted)

	Factor 5	Factor 3	Factor 4	Factor 1	Factor 2
1	0.306	0.065	1.663	1.781	1.738
2	1.835	2.103	-0.824	2.291	1.149
3	0.143	0.995	0.380	1.161	0.874
4	0.572	0.280	3.664	-0.239	1.192
5	2.144	1.557	0.118	0.007	0.048

## Person Loadings (Extracted)

	Factor 5	Factor 3	Factor 4	Factor 1	Factor 2
1	1.817	0.354	1.338	0.969	1.070
2	0.629	0.741	1.304	0.830	0.959
3	0.791	1.255	-0.544	1.333	0.819
4	0.996	0.867	1.458	0.942	1.083
5	0.766	1.783	1.445	0.927	1.069

## Test Loadings (Extracted)

	Factor 5	Factor 3	Factor 4	Factor 1	Factor 2
1	0.368	0.676	0.129	1.010	1.040
2	0.345	0.311	0.480	0.885	2.419
3	0.309	0.382	0.520	-0.423	1.925
4	0.294	0.408	0.052	1.505	0.974
5	0.013	0.852	0.355	1.412	2.170

Five factors extracted from data created by four factors. Since the ratio of factors extracted to the true number input is not much greater than one, some factors (in this solution factors 1 and 2) were exactly recovered. Other extracted factors show various degrees of resemblance to input factors.

unique. Its resemblance to the true solution depends on the ratio of the number of factors extracted to the number of factors actually present. If this ratio is only slightly greater than one, then some extracted factors will match true factors, although which of the true factors will be recovered will depend on the starting configuration for a given analysis. Given the following ratios of the number extracted to the true number of factors, the number of matching factors might typically be 2 for  $5/4$ , 3 for  $6/5$ , 0 for  $3/2$ , and 0 for  $6/3$ . The non-matching factors will usually include both some which bear a rough resemblance to true factors and some which are completely different from any of the true factors. An example displaying matching and both kinds of non-matching factors is given in Table 3C.

Because of the characteristics of PP analysis just described, the investigator has two interlocking criteria for determining the correct number of factors to extract from adequate data: (1) the largest number of factors that still gives a unique solution; and (2) the traditional criterion of the largest number of factors which still significantly improves "fit".

#### B. Conditions of an adequate data set

According to the fundamental theory of PP factor analysis (Sections II and III, above) a unique solution requires data for a number of measures on a number of persons collected on at least two occasions, with system-variation of the factors present across these two occasions. This basic requirement was confirmed in experiments with synthetic data. Data sets containing only one occasion of measurement (in effect, two-way data sets) failed to yield unique solutions when factors were fitted to them by TBU or the quick algorithm. Similarly, three-way data sets in which all the factors had constant loadings across the occasions (i.e. with no system-variation) did not yield unique solutions regardless of the number of occasions.

On the other hand, it was possible for all factors to be uniquely determined and correctly recovered if their columns of loadings on the different occasions were distinct. (For two columns to be distinct, it is apparently sufficient that they be linearly independent.) Thus, even when one factor did not change its influence across occasions, it could be correctly recovered in the analysis so long as no other factor was similarly constant.

If, in a given data set, only some factors had occasion loading patterns distinct from all other factors (e.g. see Table 4A), then only the distinct factors would be uniquely determined in the solution. For example, a data set was created in which three of the four latent factors had the same increasing pattern of loadings across occasions. Analysis of this data gave non-unique solutions for three of the four factors extracted. Further, the one factor which was uniquely determined in the otherwise different solutions (as evidenced by its occurrence in

the same form in each solution) was found to be an exact match of the one factor which had a loading pattern across occasions which was distinct from the other three factors. This type of result is encouraging, since it indicates that the conditions underlying the data need not be completely adequate in order to recover some valid factors. When there is inadequacy for some factors, the solution does not collapse as a whole, but rather one factor at a time. Any factors which do show the necessary distinctness across occasions of measurement can be uniquely recovered in the solution.

Two features of the non-unique solutions obtained with partially or completely inadequate data sets serve to distinguish them from the non-unique solutions obtained by too many factors. First, even when the ratio of estimated to true number of factors is small enough that some factors will be correctly recovered, it is not always the *same* factors which are recovered when different starting configurations are used. With partially inadequate data, however, the same factors will always be uniquely determined and recovered in the solution, regardless of the starting configuration used. Secondly, in these experiments the inadequate *occasion* loading patterns still seemed to be correctly recovered, even though the loadings for the other modes could not be correctly recovered. Thus the occasion loading patterns always displayed that lack of distinctness between the non-unique factors which was the cause of their non-uniqueness. By contrast, when too many factors are extracted, the loading patterns are not uniquely determined for any mode except for those factors correctly recovered as a whole. Considerations such as these provide important guides to interpretation of solutions obtained with real data. An example of loadings used to generate a partially inadequate data set and one solution obtained from data so generated is given in Table 4A.

Although distinct loading patterns across two occasions might be theoretically sufficient to uniquely determine any number of factors, this sufficiency could not be established in these experiments because of practical convergence difficulties. These difficulties made it impossible to solve for more than 4 or 5 factors using a data set with only two occasions. Two occasions were never found *insufficient*, since convergence on a non-unique solution was never obtained with otherwise adequate data when there were only 2 occasions. However, since convergence could not be obtained with two-occasion data sets which had a large number of latent factors, the adequacy of a two-occasion data set for uniquely determining these large numbers of factors could not be empirically confirmed.

Apparently, when the number of occasions is very small, specifically two or three, the properties of the hypersurface where the routine is seeking an optimum become such that the upward paths are narrow and twisted and convergence is difficult or impossible within a few thousand iterations. Before the development of the quick algorithm, these problems led to the conclusion that there must be at



TABLE 4A

## Factors Extracted from "Partially Inadequate" Data

Occasion Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	0.333	0.333	0.333	1.772
2	0.667	0.667	0.667	2.048
3	1.000	1.000	1.000	1.100
4	1.333	1.333	1.333	0.066
5	1.667	1.667	1.667	0.015

Person Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	1.817	0.354	1.187	0.988
2	0.629	0.741	1.110	0.854
3	0.791	1.255	0.223	1.235
4	0.996	0.867	1.247	0.969
5	0.766	1.783	1.233	0.954

Test Loadings (Input)

	Factor 1	Factor 2	Factor 3	Factor 4
1	4.478	4.478	2.515	1.623
2	4.204	2.062	7.915	2.033
3	3.761	2.529	7.892	0.277
4	3.574	2.703	1.565	2.186
5	0.157	5.642	6.228	2.559

\* \* \*

Occasion Loadings (Extracted)

	Factor 1	Factor 3	Factor 4	Factor 2
1	0.333	0.333	0.333	1.772
2	0.667	0.667	0.667	2.048
3	1.000	1.000	1.000	1.100
4	1.333	1.333	1.333	0.066
5	1.667	1.667	1.667	0.015

Person Loadings (Extracted)

	Factor 1	Factor 3	Factor 4	Factor 2
1	1.043	5.207	3.931	0.988
2	0.861	0.877	1.704	0.854
3	0.731	-1.183	-2.072	1.235
4	1.050	1.654	2.058	0.969
5	1.316	-1.556	-0.621	0.954

Test Loadings (Extracted)

	Factor 1	Factor 3	Factor 4	Factor 2
1	11.641	0.971	-1.141	1.623
2	13.192	0.095	0.893	2.033
3	13.337	-0.101	0.946	0.277
4	7.851	0.910	-0.920	2.186
5	12.401	-1.399	1.025	2.559

Loadings used to generate "inadequate" data, and one resulting solution extracted from such data. The latent structure used on input was "inadequate" to distinguish factors 1, 2, and 3, since they had the same pattern of loadings across occasions. A typical solution extracted from this data set shows that only factor 4 was correctly recovered.

least as many intervals in each mode of an adequate data set as there are factors being extracted from the data set. For small numbers of factors this still seems to be a good rule. For large numbers of factors, 4 to 8 occasions might provide a good base for reasonably quick convergence, depending on the particular characteristics of the data set. (To date, experiments have been limited to a maximum of 10 factors because of trivial array size limitations, but the program is being modified for larger arrays in future experiments.) Table 4B gives some approximate convergence times for data sets of different sizes and containing differing numbers of factors. It should be noted that a number of other characteristics of the data set will affect its convergence properties, as will be discussed below. Therefore, the iteration counts listed in Table 4B are only indicative of reasonable but not necessarily average or minimal values. They all reflect values obtained with the quick algorithm using a relaxation factor of 1.275, starting from a random configuration of loadings.

The discussion of data adequacy has focused, up to this point, on the conditions of only one mode. For the purposes of discussion this mode has been called "occasions", although the results would apply by symmetry to any single mode of a three mode matrix. The assumption of the adequacy of the other two modes has been implicit in the discussion of the effect of changes in the "occasions" mode. The experiments listed in Table 4B (with the exception of the last one listed) were consistent with this assumption. They each had at least as many intervals in the second and third mode as there were factors underlying the data. Further, in these two modes the pattern of loadings was distinct for each factor (i.e. was always linearly independent of the loading patterns for all other factors).

The case of a single inadequate mode was studied most intensively both because it was a simplified approach to the general question of data adequacy, and because it was thought to represent the most common way in which data would be found to be inadequate. When there are difficulties with obtaining adequate data, these difficulties will frequently involve only one of the three modes. For example, it is sometimes easy to secure many measurements, and many individuals to measure, but more difficult to establish more than perhaps two or three occasions or conditions of measurement where there will be systematic differences in relative factor influence.

A few experiments have been performed, however, where more than one mode was restricted (e.g. had fewer intervals than there were factors to be extracted). The results dramatically demonstrate the differences between the theoretical capabilities and properties of classical two-mode and the new three-mode PP factor analysis. In one experiment, 10 factors were correctly recovered from an 8 by 8 by 8 data matrix (this experiment corresponds to the last entry of Table 4B). Clearly, the mathematical rank of any two-way slice of this data matrix cannot be greater

TABLE 4B

Computation Times for Some Data Sets Having  
Fewer Occasions Than Factors

	Number of Factors	Dimensions of Data Set	Number of Iterations to Convergence
a.	3	2 x 4 x 5	1000
b.	4	2 x 6 x 7	1500
c.	6	2 x 7 x 8	not close to conver- gence after 2000 iterations
d.	4	3 x 8 x 10	400
e.	4	3 x 8 x 10	300
f.	5	3 x 10 x 10	600
g.	6	3 x 11 x 10	500
h.	6	4 x 8 x 10	300
i.	10	8 x 8 x 8	800

Convergence times for 9 selected experiments. The number of iterations required to reach convergence increases with fewer intervals in the smallest mode of the data matrix, particularly as the number of factors becomes larger.

than 8, and classical two-mode factor analysis of any such slice could only reveal at most 8 factors. Yet when all the "slices" are assembled into a three-way matrix and subjected to three-mode PP factor analysis, 10 factors are correctly and uniquely recovered.

The complete set of rules defining the *minimal* conditions of an adequate data set have not yet been determined. Non-minimal *sufficient* conditions can be described with some assurance, however. It appears that in order to uniquely recover N factors, it is sufficient that the data set have N different intervals in each mode (except where this would provide fewer data points than factor loadings, as in the case of 2 factors). This empirically derived rule of sufficiency will be supported by a mathematical proof in Section V. In addition, some weaker sufficiency rules have been described above for certain selected cases, and a type of sufficiency for the correct recovery of a given factor from an otherwise inadequate data set has also been defined. These results will have direct application to the analysis of real data, but further mathematical and experimental work needs to be done in order to rigorously define the minimal necessary conditions of an adequate data set, as a function of the number of latent factors present.

### C. Conditions affecting convergence

The ratio of the number of factors to the number of occasions (or to the number of intervals in whichever is the smallest of the three modes) is one characteristic of the data set which strongly affects convergence rate (as was discussed in B above). This points out the practical necessity of the generalization of the PP criterion from two to any number of "parallel" occasions. Additional conditions which affect convergence include: the presence of positively correlated factors; the presence of factors with highly unequal contributions to the data; the number of near zero loadings in the latent structure; non-uniqueness of the solution; and the starting configuration from which an attempt at optimization was begun.

If two of the latent factors have a strong positive correlation (e.g. +.75 or higher) the similarity of their loading patterns apparently begins to pose problems for the optimization routine. Faced with two similar but not identical loading patterns across a given mode, the PP analysis procedure usually reaches a rough fit and then struggles across a plateau where it tries to disentangle the influence of the two correlated factors. Convergence can be slowed in such circumstances as much as 4 to 10 times.

Similar difficulties arise when one or more factors make only very small contributions to the data values relative to the other factors. Here, too, the program first finds a rough fit to the data and then has to spend a great deal of time transforming this "first guess" into the often very different solution required for a final good fit. The rough "first guess" gives results similar to extracting too many factors, since the small factors haven't exerted their influence effectively.

Often values are very different from the correct final values, and if one stopped at this point the solution would appear non-unique from different starting configurations. After this "first guess" the program goes through a long series of successive corrections to achieve small improvements in fit, as it tries to respond to the feeble influence of the small factors. One possible way to handle such cases may be to first solve for fewer factors. The program should converge fairly quickly on a unique solution. The factors in this solution would resemble the larger factors, only slightly distorted because of the influence of the unrepresented small factors. The loading values for these large factors can be used as a starting point for the solution involving more factors. Starting values for the small factors can be columns of zeros, or any other good first estimate. From such a starting configuration, convergence should proceed much faster than it otherwise would in the case of very unequal factors.

If a number of zero loadings are present in the latent structure, convergence can be facilitated. Apparently this simplifies the path up the hypersurface by reducing the interdependencies among the variables. This effect has been most clearly noticed for Routine TBU.

Convergence is greatly speeded when the solution is not unique, either because of data inadequacy or because of the extraction of too many factors. This might have been expected, since it should be easier to climb to the optimum by finding and following one of the many paths that would end up on that hyperplane defined by an infinity of alternative optimum solutions, than it is to find and follow one of the fewer paths that would converge on a unique point. The "more non-unique" the solution (e.g. the more the number of factors extracted exceeds the true number of factors, and thus the more divergent the possible loadings in the alternative solutions), the more the convergence is accelerated. Thus convergence speed might provide an additional confirmatory guide to the choice of the number of factors.

With analyses having unique solutions, the particular starting configuration used was often found to have a very significant effect on convergence time, sometimes causing variations in convergence time as great as 500%. But with the error free synthetic data, there appeared to be no problems with local optima. Although both TBU and the quick algorithm sometimes experienced slowdowns where they would display steady but very gradual progress up the hypersurface, the values of the parameters would never stop changing completely unless the program had achieved a near perfect fit to the data.

In general, it seems that although difficulties are encountered with some types of data sets, the quick algorithm with the relaxation factor is adequate to allow analysis of most real data to which the PP model might apply. To cope with the remaining difficult cases, further development of the algorithm is being explored, in part by combining features of the quick algorithm with additional trend estimation techniques such as are found in Routine TBU.

## The Analysis of Real Data from Vowel Sounds

Although the experiments with synthetic data provide the only certain way to demonstrate the discovery of a true latent structure, they may not adequately simulate some other properties of real data which could affect performance of the analysis procedure. Therefore a set of real data was selected for analysis.

The data consists of a set of measurements of formant frequencies (converted into pitch values in mels) for eight vowel sounds, as uttered by 11 individuals. The resulting data was a 4 by 8 by 11 matrix of pitch values whose dimensions are formant, by vowel, by individual (Table 5). This data is a subset of a four-way matrix of formant by vowel by individual by occasion, originally published by Ladefoged (1967, pp. 88-89).

The theoretical interest in the data lies in the question of the nature of vowel quality. Phoneticians have been seeking an objective description of the features in the acoustic signal which are characteristic of the different vowels. It is important in this respect to distinguish between those differences in vowel quality that do not serve to identify the vowel (e.g. the different quality of different voices) and those characteristic qualities which do serve to identify the vowel. The second type of difference is of primary interest here.

A set of dimensions for vowels has been developed by phoneticians through consideration of the positions of the tongue and lips, and through evaluation of subjective sound quality. But the systematic classification of vowels on this basis has not been clearly related to the particular objective characteristics of the acoustic signal which (presumably) the brain uses in recognizing vowels in speech. It is, however, generally agreed that the most important acoustic features are certain characteristic peaks in the spectrum of vowel sounds (Fant, 1960). These peaks are called "formants", and they are closely related to resonances of the vocal tract, as determined by the shape of the tongue, lips, etc. In the production of different vowels, the shape of these vocal organs is changed and thus the location of the formants in the sound spectrum alters.

Careful study has failed to find a satisfactory relationship between the formant frequencies (or corresponding pitch values in mels) and the presumed dimensions of vowel quality. In fact, phoneticians have been unable to discover any physical dimensions which would serve to define the perceptual qualities which distinguish vowels (Ladefoged, 1967, pp. 100-103; Ladefoged, 1970). It was in hopes of discovering such relationships and giving objective acoustic definition to the dimensions of vowel quality that factor analysis of the formant data was carried out, using the proportional profile three-mode factor model.

Since, as has already been noted, the PP model is not completely general, its application to this three-way data set must be justified.

TABLE 5

## Data Input for Formant 0

1	215.000	215.000	215.000	215.000	215.000	215.000	215.000	215.000
2	255.000	255.000	255.000	255.000	255.000	255.000	255.000	255.000
3	135.000	135.000	135.000	135.000	135.000	135.000	135.000	135.000
4	205.000	205.000	205.000	205.000	205.000	205.000	205.000	205.000
5	350.000	350.000	350.000	350.000	350.000	350.000	350.000	350.000
6	260.000	260.000	260.000	260.000	260.000	260.000	260.000	260.000
7	300.000	300.000	300.000	300.000	300.000	300.000	300.000	300.000
8	245.000	245.000	245.000	245.000	245.000	245.000	245.000	245.000
9	325.000	325.000	325.000	325.000	325.000	325.000	325.000	325.000
10	300.000	300.000	300.000	300.000	300.000	300.000	300.000	300.000
11	365.000	365.000	365.000	365.000	365.000	365.000	365.000	365.000

## Data Input for Formant 1

1	405.000	475.000	655.000	930.000	0.0	590.000	465.000	420.000
2	275.000	455.000	695.000	855.000	760.000	555.000	460.000	370.000
3	480.000	535.000	710.000	915.000	765.000	615.000	505.000	400.000
4	395.000	505.000	745.000	995.000	810.000	590.000	535.000	340.000
5	345.000	540.000	635.000	920.000	715.000	715.000	495.000	350.000
6	290.000	415.000	730.000	945.000	0.0	725.000	430.000	290.000
7	340.000	505.000	690.000	980.000	0.0	665.000	515.000	360.000
8	335.000	455.000	715.000	910.000	715.000	720.000	415.000	360.000
9	355.000	550.000	715.000	1045.000	0.0	435.000	505.000	335.000
10	315.000	475.000	680.000	980.000	0.0	650.000	485.000	395.000
11	370.000	505.000	765.000	1115.000	760.000	595.000	595.000	365.000

## Data Input for Formant 2

1	1720.000	1620.000	1460.000	1370.000	0.0	0.0	0.0	0.0
2	1620.000	1605.000	1485.000	1260.000	980.000	825.000	825.000	0.0
3	1650.000	1620.000	1530.000	1305.000	960.000	795.000	695.000	695.000
4	1790.000	1660.000	1620.000	1400.000	1015.000	895.000	855.000	0.0
5	1765.000	1805.000	1620.000	1350.000	975.000	960.000	790.000	740.000
6	1760.000	1645.000	1510.000	1455.000	0.0	0.0	745.000	735.000
7	1670.000	1615.000	1510.000	1400.000	0.0	980.000	830.000	0.0
8	1770.000	1620.000	1470.000	1380.000	980.000	940.000	0.0	0.0
9	1880.000	1820.000	1700.000	1405.000	0.0	845.000	0.0	0.0
10	1725.000	1670.000	1585.000	1430.000	0.0	0.0	0.0	0.0
11	1775.000	1765.000	1645.000	1405.000	985.000	870.000	0.0	855.000

## Data Input for Formant 3

1	2030.000	1760.000	1760.000	1835.000	1835.000	1875.000	1965.000	2070.000
2	1895.000	1870.000	1905.000	1720.000	1785.000	1720.000	1705.000	0.0
3	1960.000	1885.000	1750.000	1855.000	1865.000	1825.000	1735.000	1735.000
4	2135.000	1960.000	1895.000	1855.000	1940.000	1900.000	1795.000	2100.000
5	2075.000	1955.000	2030.000	2050.000	2045.000	2040.000	2015.000	0.0
6	2145.000	1865.000	2135.000	1670.000	1775.000	1750.000	2115.000	2160.000
7	2040.000	1820.000	1820.000	1940.000	1800.000	1750.000	1680.000	1710.000
8	2045.000	1860.000	1715.000	1840.000	2035.000	1945.000	1855.000	0.0
9	2260.000	2060.000	2015.000	1995.000	2080.000	1985.000	2175.000	2240.000
10	2200.000	1900.000	1850.000	1895.000	1965.000	1905.000	2045.000	2145.000
11	2140.000	1950.000	1930.000	1985.000	2090.000	2095.000	2085.000	2175.000

Formant pitch values (in mels) for the fundamental frequency ("formant 0") and the first three formants of eight cardinal vowels as spoken by 11 persons. Zero values represent missing data points, which during analysis were simply skipped and not "fitted" by the optimization program.

The application is of particular interest since none of the three modes represents a time dimension. Although throughout the earlier discussion of the PP model, the third mode was called "occasions", it could in fact be any third set of circumstances in which systematic variations of relative contribution of the factors might be expected. In the case of this vowel data, the measurements (formant pitch values) for the individuals (persons producing the vowels) are taken for eight different vowels (i, e, ε, a, α, ɔ, o, u).

The PP or system-variation model is thought to be justified here because systematic general differences of factor influence are expected not only when one vowel is compared to another, but also from one person to the next, and from one formant to the next.

By definition, the individual vowels must differ in their relative positions on the underlying dimensions of vowel quality, since it is these dimensional differences which are presumed to distinguish them. But it can also be expected that different persons will not use each of the underlying dimensions to exactly the same degree in their vowel production. The speech of some will tend, in general, to be slightly more prominent in the use of a certain aspect of vowel quality, and on the other hand have perhaps slightly less of a different aspect, consistently across all vowels. These differences could arise because some speakers tend to use a greater amount of lip rounding on all vowels, other speakers tend to speak with the tongue slightly higher in the mouth than is usual, etc. In addition, physiological differences in the shape of each person's vocal tract, differences in tongue size, etc., could be expected to cause shifts in the quality of that person's vowels compared to those produced by other speakers. These shifts would presumably be of different degrees for each of the underlying dimensions, since alterations of vocal tract shape would probably not have precisely equal consequences for all dimensions of vowel quality.

The argument for system-variation seems less convincing with respect to the formant mode. There seems to be no obvious reason for expecting any particular pattern of factor loadings across the three formants. But of the various possibilities, the vast majority would show unequal and distinct loading patterns for each factor across the three formants. Therefore, it does not seem unlikely that the required system-variation would be present in the formant mode as well.

Before actually performing an analysis, one can only argue that this system variation model of vowel quality is a plausible one. One cannot really know whether the model is, in fact, correct. But after the analysis the appropriateness of the model can be confirmed or not by examining the results. If one or more of the factors do not show the postulated system-variation, this should become apparent in the results of the analysis. From the experiments with synthetic data it was learned that if system-variation of a given factor's influence is not present across all modes of a data set, then that data set would not be "adequate"



for that factor, and the solution for that factor would not be unique.

The analysis of this real data was performed by fitting one, two, and three factor PP solutions to the data using Routine TBU. The quick algorithm could not be applied in order to speed computation because of the many missing entries in the data matrix (the missing entries appear as zeros in Table 5). Although attempts are being made to overcome the difficulty, the quick algorithm has not yet been modified to adequately deal with missing data. The problem arises because the computation of new values for a given mode requires intermediate products which involve the complete data set (equation 13). Routine TBU, on the other hand, can deal easily with missing values, since it does not require such intermediate products but relies instead on changes in the sum of the squared errors resulting with each attempt to "fit", by estimation from factor loadings, the non-missing values of the data set. Any missing value can simply be ignored by the program.

After applying TBU to the vowel data, it became apparent that the data set possessed characteristics which substantially slowed the process of convergence. Since TBU is already slow compared to the special purpose quick algorithm, true convergence could not be obtained with many of the analyses. Nonetheless, by pushing convergence as far as computer time would allow, results have been obtained which are believed to represent reasonable approximations to what the values would have been at convergence. Because of the additional slowdown of TBU as the number of factors increases, solutions containing four factors were not attempted. This omission is less serious however, because the results with two and three factors show that any fourth factor, even if present, would only make a very small contribution to the values of the data points. And it has been demonstrated with synthetic data that omitting a small fourth factor will not substantially distort the loadings of the other three factors.

In light of these limitations, the results reported as the two and three factor solution for this vowel set must be viewed as likely approximations, to be refined further when the quick algorithm is adjusted to cope with missing values, and also to be confirmed by obtaining and analyzing different sets of vowel formant data which do not have missing values. Both lines of work are currently being pursued.

Nonetheless, PP factor analysis of this vowel data is not simply an exercise of the program, but has inherent interest from the viewpoint of linguistic phonetics. Further, it will be seen that the solution obtained possesses a highly meaningful form. The results therefore seem to demonstrate the usefulness of PP analysis, even when operating under handicaps.

#### Results of the Analysis of Vowel Data

The analysis was carried out first for the four vowels (i, e, ε, a).

In order to determine the number of real factors, comparisons were made of both the amount of error remaining after convergence slowed or stopped, and of the uniqueness of the solutions obtained as one, two, and three factors were extracted. These criteria had to be used flexibly, however, because of the approximations caused by imperfect convergence, and also because of the distortions which would be caused by the presence of error variation in the data. On the basis of 10 analyses of the restricted four vowel data set, an apparent cutoff was obtained at two factors. The mean squared error of estimation was consistent at around 6,600 mels both for four different two-factor and for four different three-factor analyses. The analyses were performed from different random starting configurations. There were variations of only around 10% in the final mean squared error values for all eight of these analyses, and the three-factor solutions did not tend to have lower scores. With regard to uniqueness, the results were less clear. Both for the two-factor and three-factor solutions, there were similarities among solutions, but no clear uniqueness which could be said to hold for the loadings in all three modes. The vowel mode seemed to give the most consistent loading patterns. In all two-factor solutions the vowels showed a systematically increasing set of loadings on one factor and a systematically decreasing set of loadings on the other factor. Loadings for the person mode showed some similarities in successive analyses but could not be said to be recognizably unique even for two factors. The same was true for the formant mode. Even the three-factor solutions for the four vowel data set showed some interesting similarities in the vowel mode, as can be seen in Table 7. It was concluded that the error variation in the data, the convergence difficulties, and the possible distorting presence of a third factor were making the analysis difficult.

In an attempt to minimize the inaccuracy due to incomplete convergence, a two-factor solution was carried to 350 iterations of Routine TBU (requiring 300 seconds of CPU time on the 360/91). This represented at least five times as many computations as on any of the other analyses of the four vowel data set. Yet convergence still had not been achieved at the end of this time. These results were therefore taken as the best approximation, short of true convergence, obtainable for the loadings of two factors on the four vowel data. The obtained loadings are given in Table 6, and the position of the four vowels in the plane determined by the two factors is given in Figure 1.

The data set was then expanded to include all eight vowels, and the analysis procedure repeated. An immediate improvement was noticeable in the stability of the solutions obtained for both two and three factors. Further, with this expanded data set the mean squared error of fit could be seen to drop from 7000 mels for two factors to 2800 mels for three factors. Although the precise values of the vowel loadings varied somewhat across analyses, they were roughly similar, and the patterns of the loadings was more clearly unique, for the three-factor as well as the two-factor solutions. In all cases there was one factor which would increase across the eight vowels, another factor which would decrease, and

TABLE 6A

## Two-Factor Analysis of Four Vowels

Formant Loadings		
Factor 1	Factor 2	
1	0.248	0.240
2	-0.036	1.227
3	1.836	0.996
4	1.952	1.536

Vowel Loadings		
Factor 1	Factor 2	
1	1.375	0.615
2	1.151	0.783
3	0.915	1.082
4	0.559	1.520

Person Loadings		
Factor 1	Factor 2	
1	552.258	508.999
2	553.917	489.476
3	540.112	524.531
4	587.279	533.850
5	600.630	545.492
6	597.261	486.751
7	546.270	546.178
8	567.719	502.570
9	631.319	554.582
10	589.327	526.453
11	578.573	576.265

TABLE 6B

## Two-Factor Analysis of Four Vowels

Formant Loadings		
Factor 1	Factor 2	
1	143.0430	126.4400
2	-20.7644	646.4220
3	1058.9800	524.7240
4	1125.8900	809.2130

Vowel Loadings		
Factor 1	Factor 2	
1	1.375	0.615
2	1.151	0.783
3	0.915	1.082
4	0.559	1.520

Person Loadings		
Factor 1	Factor 2	
1	0.9575	0.9662
2	0.9604	0.9291
3	0.9364	0.9956
4	1.0182	1.0133
5	1.0413	1.0354
6	1.0355	0.9239
7	0.9471	1.0367
8	0.9843	0.9540
9	1.0946	1.0527
10	1.0218	0.9993
11	1.0031	1.0938

Two factors extracted by extended analysis of the data for the first four vowels. Table 6b is the same as 6a, except that the normalization has been performed across persons so that formant loadings reflect true pitch values in mels.

TABLE 7

Three Attempts at Fitting Three Factors to Four Vowel Data

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	-0.578	-9.213	-1.154
2	0.442	10.930	4.140
3	2.081	0.972	0.000
4	2.055	1.312	1.014

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	0.054	0.222	2.015
2	0.260	2.031	-1.192
3	1.767	0.477	1.412
4	1.918	1.270	1.765

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	0.896	-3.276	4.764
2	-2.052	3.040	-2.174
3	2.605	2.952	-0.803
4	2.551	1.285	2.213

Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	1.111	0.983	0.811
2	1.025	0.942	0.862
3	0.973	0.998	1.038
4	0.892	1.076	1.289

Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	1.173	0.487	1.199
2	1.065	0.702	0.987
3	0.957	1.116	0.971
4	0.805	1.695	0.843

Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	2.237	0.921	0.894
2	1.610	0.930	0.901
3	0.733	1.015	1.017
4	-0.580	1.134	1.188

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	792.646	-120.157	380.405
2	782.005	-125.907	387.647
3	790.147	-109.530	363.755
4	836.736	-124.931	400.044
5	857.004	-147.075	443.417
6	820.388	-127.126	387.872
7	805.151	-136.716	427.242
8	802.046	-125.324	390.155
9	885.816	-147.557	455.140
10	834.089	-138.044	423.661
11	845.975	-151.325	474.186

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	752.396	241.872	59.777
2	720.434	240.489	83.052
3	788.191	236.796	20.412
4	805.078	253.468	53.670
5	757.301	275.162	121.408
6	760.334	236.580	88.424
7	716.428	279.801	100.674
8	750.628	243.421	73.744
9	805.671	281.166	104.401
10	755.095	264.342	100.735
11	735.898	310.031	120.937

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	82.893	559.428	415.109
2	86.003	544.676	414.191
3	77.347	567.430	409.648
4	91.828	588.885	432.536
5	86.034	600.636	471.115
6	97.677	561.921	426.369
7	77.982	574.820	441.968
8	88.603	560.100	419.356
9	99.474	617.177	471.446
10	91.351	581.856	444.990
11	87.519	602.929	465.261

Sets of loadings resulting from separate attempts to fit three factors to the first four vowels. Although the vowel loading patterns were fairly consistent, the formant loadings and person loadings showed fairly large variations, indicating relative lack of uniqueness or very high distortion by error variations in the data.

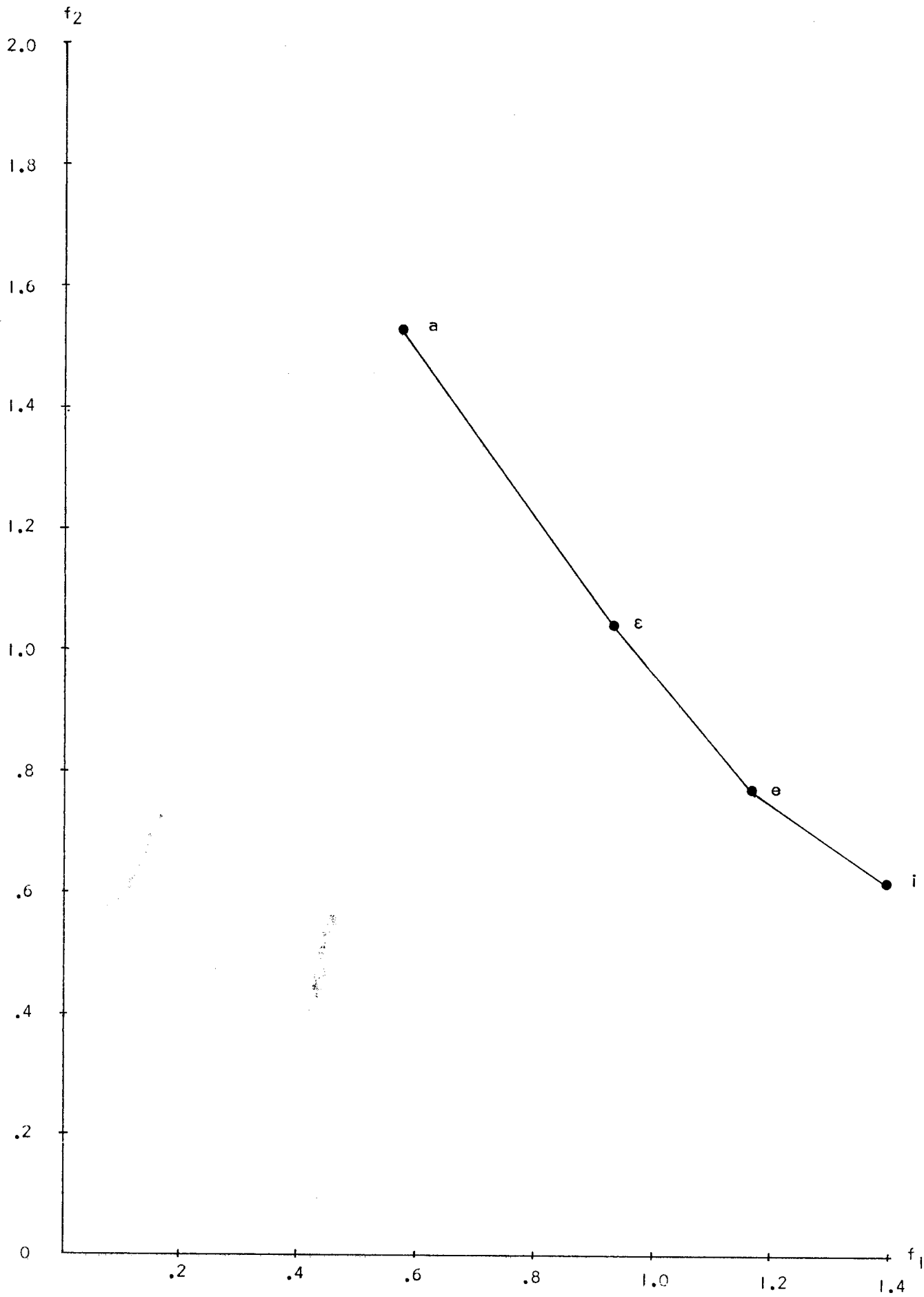


Figure 1. Plot of the position of the first four vowels in the two-factor plane extracted from the four vowel partial data set.

a third factor which would first decrease and then increase. In most of the solutions there was a tendency for the end of the steadily increasing or decreasing factors to show "flattening" or decreased step sizes, and sometimes even to "curl around" slightly. (The flattening is evident in Table 8 for the vowel loadings of factors 2 and 3).

Although much more stable than in the four vowel analyses, the loadings for the person and formant modes still showed somewhat greater variations than those of the vowel mode. In an attempt to reduce the variations due to lack of convergence, another prolonged analysis was performed. Iterations were allowed to proceed for 500 seconds on the 360/91. This provided 227 iterations of TBU, fewer than on the four vowel problem because of the greater number of parameters being optimized and the larger data set. At the end of this prolonged analysis, complete convergence still had not been obtained, but the rate of change of the parameters had slowed considerably. Since results of this analysis were not strikingly different from the results of the prior, shorter analyses, the values obtained in this long analysis were taken as the best approximations to a three-factor solution for the vowel data set. The final values of the obtained factor loadings are given in Table 8, and the vowels are plotted in the space determined by these three factors in Figure 2.

This three-factor solution seems to fit the data fairly well, since the average absolute residual is in the neighborhood of 50 mels, which represents perhaps 5% error across the data as a whole.

The most striking aspect of the analysis of all eight vowels is the correspondence between the factors which emerged and the theoretical dimensions of vowel quality deduced on subjective auditory and physiological grounds by phoneticians. A plot of the eight vowels as located in a space of these three theoretical dimensions is reproduced in Figure 3 (from Ladefoged, 1967, p. 140). Factor 1 of Table 8 would seem to correspond to the dimension of open-closed in Figure 3. Both in Ladefoged's diagram and in the results of the factor analysis (Table 8), the vowels seem to form a "U" on this dimension. Also striking is the correspondence of the other two dimensions. In both cases the vowels are consistently increasing on one dimension and consistently decreasing on the other dimension. Because of the symmetry of these two dimensions, however, it is difficult to identify which of the two factors (factor 2 or factor 3) corresponds to which of the two theoretical dimensions (roundedness or front-back) in Ladefoged's diagram. This problem could be solved by the addition of data on other vowels which, according to phonetic theory, do not have symmetrical relations on the two dimensions.

This striking correspondence between two descriptions derived by completely independent means is strong evidence in favor of both the reality of the phonetician's theoretical dimensions, and the reality of the latent structure revealed by the PP factor analysis.

TABLE 8A

## Three Factor Solution for All Eight Vowels

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	0.455	0.274	0.249
2	-1.458	0.869	0.647
3	1.066	0.960	1.343
4	3.937	1.896	1.762
Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	1.661	-2.370	2.629
2	0.576	-2.059	2.538
3	-0.190	-0.658	2.007
4	-1.197	1.318	1.233
5	0.607	2.899	0.143
6	1.224	2.877	-0.005
7	2.144	2.880	-0.156
8	3.176	3.113	-0.388

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	60.215	290.334	628.164
2	50.124	274.292	617.065
3	43.419	280.509	630.385
4	59.646	293.643	659.668
5	58.310	312.318	680.971
6	71.227	283.766	644.470
7	42.056	285.500	641.036
8	60.625	296.037	638.089
9	72.603	308.226	694.827
10	65.631	301.906	658.913
11	64.702	317.330	683.434

TABLE 8B

## Three Factor Solution for All Eight Vowels

Formant Loadings			
	Factor 1	Factor 2	Factor 3
1	26.8267	80.8016	162.4620
2	-85.9634	256.2650	422.1390
3	62.8511	283.1010	876.2490
4	232.1250	559.1230	1149.6300
Vowel Loadings			
	Factor 1	Factor 2	Factor 3
1	1.661	-2.370	2.629
2	0.576	-2.059	2.538
3	-0.190	-0.658	2.007
4	-1.197	1.318	1.233
5	0.607	2.899	0.143
6	1.224	2.877	-0.005
7	2.144	2.880	-0.156
8	3.176	3.113	-0.388

Person Loadings			
	Factor 1	Factor 2	Factor 3
1	1.0213	0.9845	0.9628
2	0.8501	0.9301	0.9458
3	0.7364	0.9512	0.9662
4	1.0116	0.9958	1.0111
5	0.9890	1.0591	1.0437
6	1.2081	0.9623	0.9878
7	0.7133	0.9681	0.9825
8	1.0283	1.0039	0.9780
9	1.2314	1.0452	1.0650
10	1.1132	1.0238	1.0099
11	1.0974	1.0761	1.0475

Three factors extracted by extended analysis of the full set of eight vowels. In Table 8b, data is normalized across persons rather than formants to give formant loadings in mels as in Table 8a.

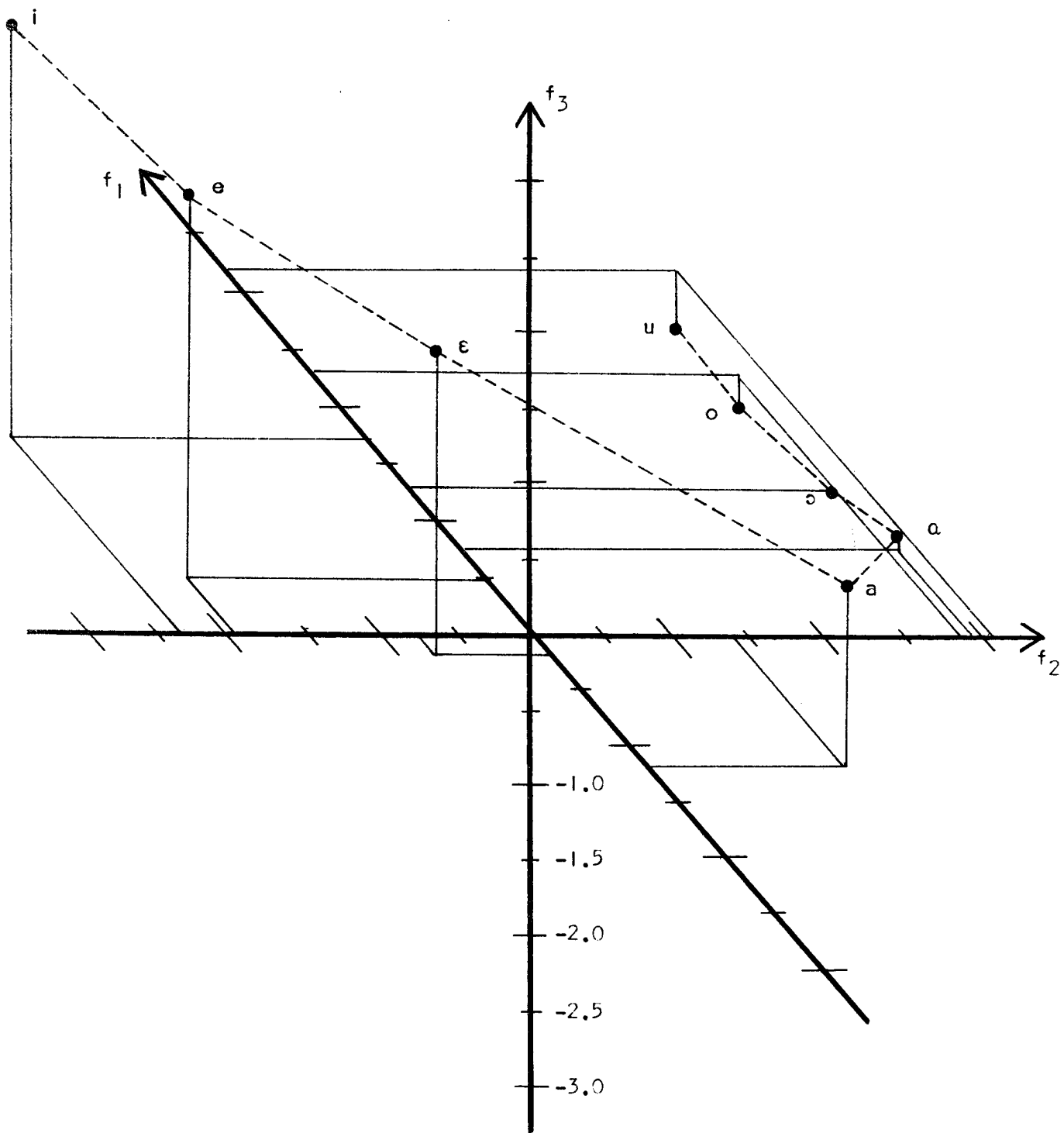


Figure 2a. The full set of eight vowels plotted on the three factors extracted from the complete data set.



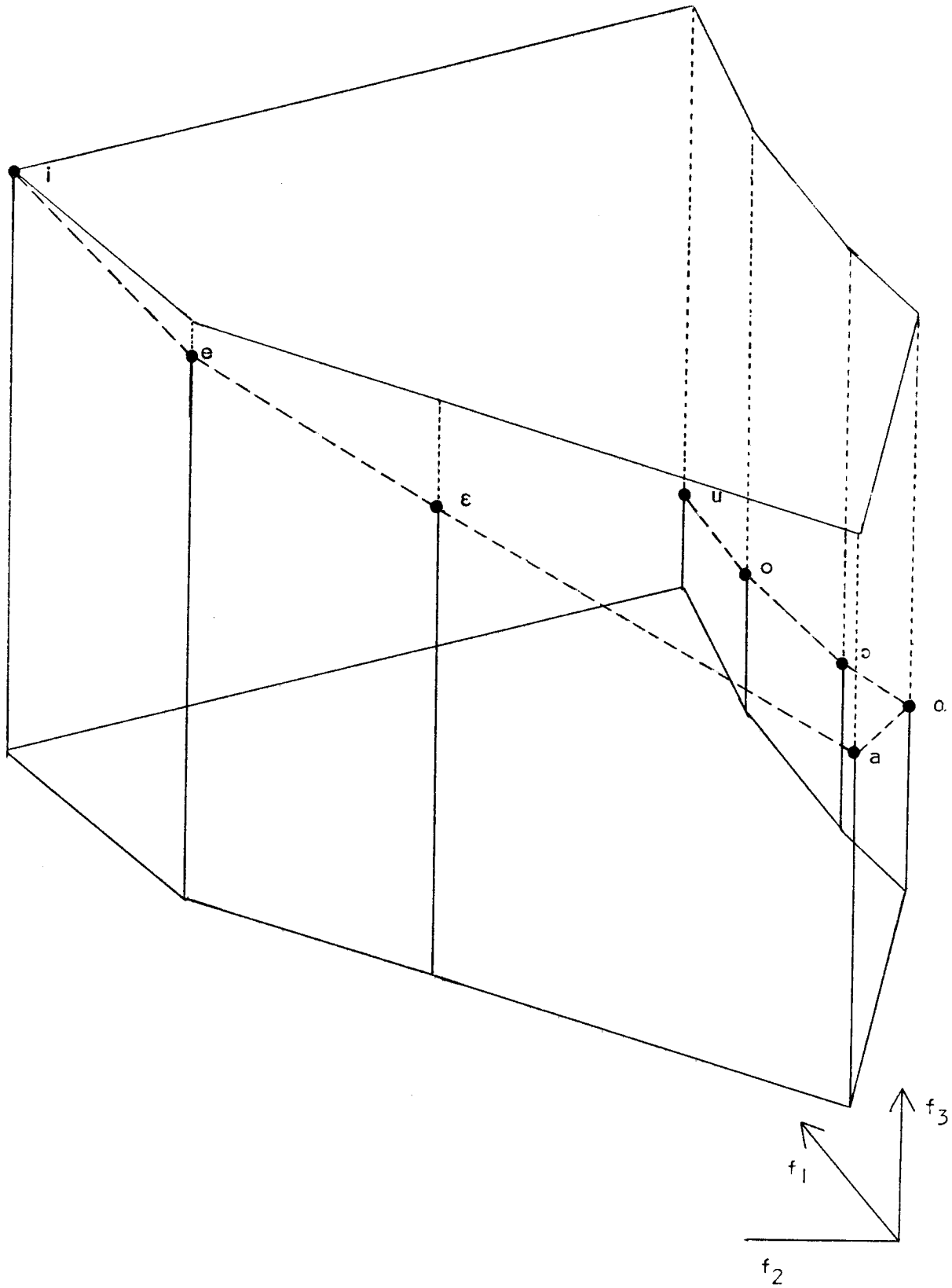


Figure 2b. The eight vowels plotted in the same space as Figure 2a, with the factor axes removed for clarity. A box is constructed around the vowel points in the same manner as was done in Ladefoged, 1967, to facilitate comparison with his figure (reproduced in Figure 3).

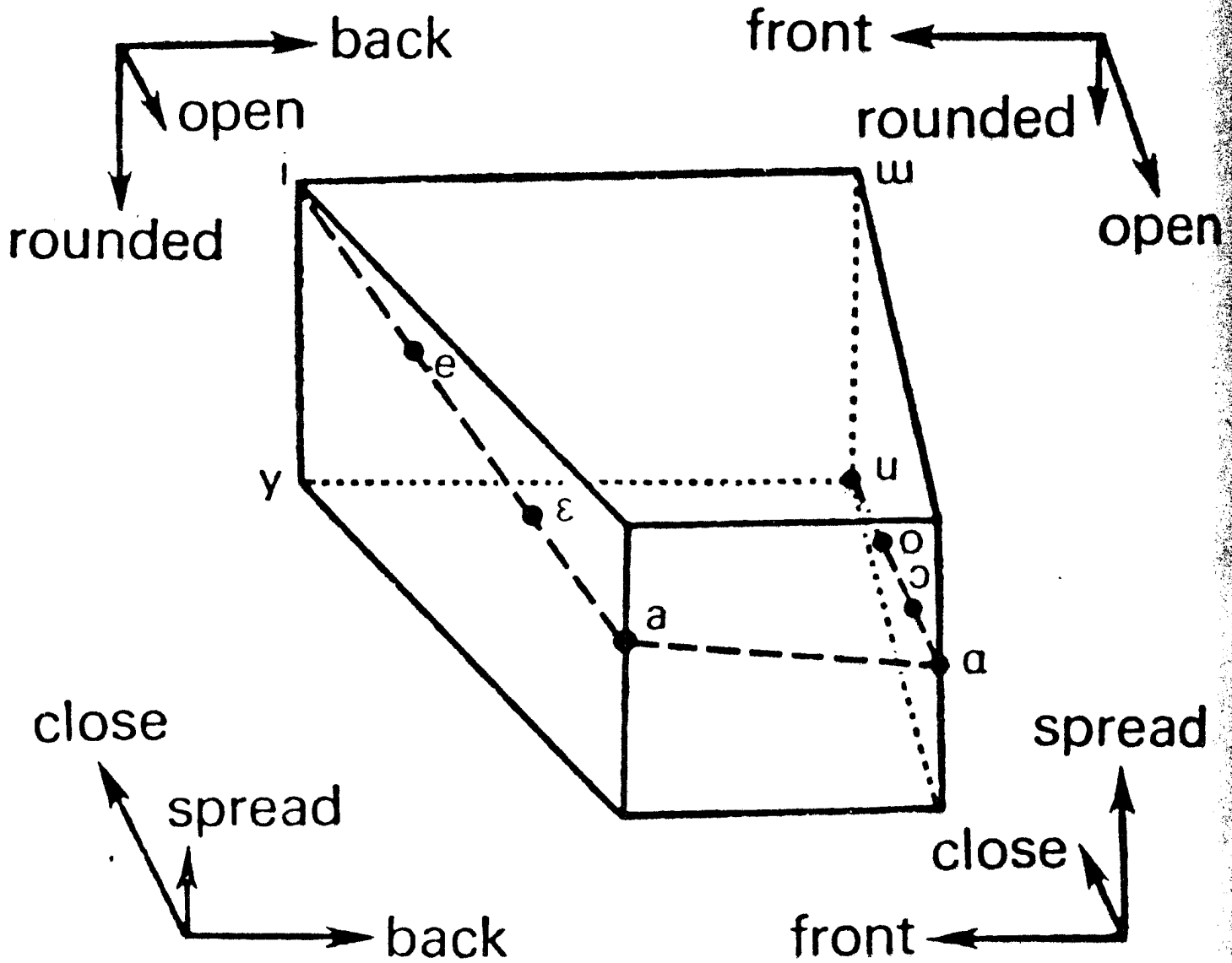


Figure 3. (reproduced from Ladefoged, 1967, p. 140) The relations of the eight vowels on three dimensions of vowel quality which have been hypothesized by phoneticians on the bases of auditory quality, tongue position, and other considerations. These hypothetical dimensions had not been successfully defined in terms of objective properties of the sound signal itself.

Additional information about the three factors is given by their loadings on persons and on formants. Since both the person mode and the formant mode showed less stability in the repeated short analyses than did the vowel mode, interpretation of their loading patterns should be made with caution. It is interesting to note, however, that the three columns of person loadings are positively correlated with one another. This suggests that the physiological or linguistic variations from person to person which affect a given aspect of vowel quality also tend in various degrees to affect the other two aspects of vowel quality. Furthermore, the fundamental frequency value (low vs. high pitched voice) for each person also correlates positively with that person's loadings on the three factors. People who speak in higher pitched voices tend to have smaller vocal tracts. The pattern of correlations given in Table 9 might be taken to suggest a physiological influence, such as vocal tract size or male-female differences in tract shape, to account for the individual differences in the relative emphasis of the three aspects of vowel quality. At this point, however, such interpretation is highly speculative.\*

The correlations shown in Table 9 are also useful in helping to explain the convergence difficulties encountered with this data set. It was discovered in the experiments with synthetic data that high positive correlations among the loadings for two factors in any mode will cause a significant slowdown in the process of convergence. The correlation of .90 between the person loadings for factors two and three is even higher than the artificially constructed correlations which slowed convergence with synthetic data. Therefore, it is not surprising that convergence was particularly difficult with this data set.

The formant mode factor loadings are perhaps the hardest to relate to external knowledge. They provide a definition of changes in the three dimensions of vowel quality in terms of changes in the pitch of the three formants. For this purpose a normalization of the data which gives the person loadings a mean of 1.0 and lets the formant loadings reflect values in mels is perhaps more revealing (as was done in Table 6B and Table 8B). Since the objective acoustic definition of the phonetician's three theoretical dimensions of vowel quality is still largely a mystery (Ladefoged, 1967, pp. 103, 140-141), phonetic theory cannot be used to directly evaluate the plausibility of these values. On the other hand, if the interpretation developed here for these three PP factors is further confirmed, and if the formant loadings for these factors can be well defined in future factor analyses, then an objective acoustic definition of the three theoretical dimensions of vowel quality might finally have been determined.

---

\* Questions could be raised, for example, about the sign of the correlations between the "backness" factor and the other factors. However, such details of interpretation will not be discussed in this article.

TABLE 9

Correlations Among Person Loadings and  
Fundamental Frequency of Voice

	Factor 1 ("Open - Closed"?)	Fundamental Frequency (F <sub>0</sub> )	Factor 3 ("Back"?)	Factor 2 ("Spread"?)
Factor 1 ("Open- Closed"?)		.42	.57	.55
Fundamental Frequency (F <sub>0</sub> )			.71	.73
Factor 3 ("Back"?)				.90

Correlations among person loadings from Table 8, plus fundamental frequency or voice pitch from Table 5. Columns have been arranged to show that the pattern of correlations is similar to that expected when there is a single influence responsible for most of the covariation.

## V. A UNIQUENESS THEOREM FOR PP FACTOR ANALYSIS

An empirical demonstration of uniqueness of solution for the three-mode PP model was presented in Section IV. This type of evidence for uniqueness leaves open the possibility that in some very unusual circumstances, not adequately covered in the empirical tests, the uniqueness might not hold. A formal mathematical proof of uniqueness would rule out such a possibility and provide the user with an important reassurance of the validity of the technique.

The following proof of uniqueness of the PP solution was discovered by Dr. Robert Jennrich.\* It requires that the factor loading matrices of each mode have a rank equal to or greater than the number of factors. This appears to be a stronger requirement for uniqueness than was found necessary in the empirical experiments (see Section IV). Therefore the theorem should be taken as formally establishing that, for  $N$  factors, an  $N$  by  $N$  by  $N$  data set can be sufficient for a unique solution. A further theorem will be required to determine the minimal conditions necessary for uniqueness.

### Jennrich's Basic Uniqueness Theorem\*\*

Theorem: If  $\sum_l O_{il} P_{jl} T_{kl} = \sum_l O'_{il} P'_{jl} T'_{kl}$  and if the matrices  $O, P, T$  each have rank  $L \leq I, J, K$ , then

$$(15) \quad O' = ORD_1, \quad P' = PRD_2, \quad T' = TRD_3$$

where  $R$  is a permutation matrix and  $D_1, D_2, D_3$  are diagonal matrices with  $D_1 D_2 D_3 = I$ .

Proof: Let  $x_{ijk} = \sum_l O_{il} P_{jl} T_{kl}$ , and let  $O_l P_l T_l$  denote the  $l$ th columns of  $O, P, T$  respectively. Then

---

\* UCLA Department of Biomathematics. The author is indebted to Dr. Jennrich both for his proof and for the basic form of the quick algorithm.  
 \*\*  $O, P,$  and  $T$  are here used to represent the weights designated by  $o, a,$  and  $F$  in equation 2 and  $O, I,$  and  $M$  in equations 8 and 9. The operator  $\otimes$  represents the direct or Kroneker product.

$$(16) \quad x = \sum_l O_l \otimes P_l \otimes T_l = \sum_l O'_l \otimes P'_l \otimes T'_l \quad .$$

Clearly the spaces spanned by  $O_1, O_2, \dots, O_l$  and  $O'_1, O'_2, \dots, O'_l$  are the same and, consequently, there exist  $a_{lr}$  such that

$$(17) \quad O'_l = \sum_r a_{lr} O_r \quad .$$

Similar statements hold for  $P$  and  $T$ , and hence

$$(18) \quad x = \sum_{lrst} a_{lr} b_{ls} c_{lt} O_r \otimes P_s \otimes T_t \quad .$$

From which it follows that

$$(19) \quad \sum_l a_{lr} b_{ls} c_{lt} = \delta_{rs} \delta_{st} \quad .$$

Using the fact that  $(c_{lt})$  is non-singular and considering the case when  $r \neq s$ , gives

$$(20) \quad a_{lr} b_{ls} = \sum_r \delta_{rs} \quad .$$

Consequently,  $(a_{lr})$  and  $(b_{ls})$  have exactly one non-zero element in each column. Moreover, the same statement holds for  $(c_{lt})$  and all non-zero elements are in the same positions in the three matrices. This establishes the first assertion. The second follows from

$$(21) \quad \sum_l a_{lr} b_{lr} c_{lr} = 1 \quad .$$

## VI. GENERALIZATION TO NON-LINEAR FACTOR MODELS

When non-linear patterns of factor influence underlie a set of real data, the solution obtained by fitting the linear PP model to that data will be a distorted description of those influences. In order to strengthen the accuracy and explanatory validity of a PP solution, such distortions should either be eliminated or recognized and taken into account.

The concepts of system-variation and proportional profiles were used in Section III to develop the linear PP factor model. In a similar fashion, they will be used here to develop PP models incorporating non-linear and interacting factors. There are two possible approaches to developing non-linear PP factor analysis: (1) to generalize the PP mathematical model to non-linear form, and "fit" this non-linear model to data; (2) to "fit" the linear model to data, but seek ways of recognizing when the results indicate the presence of non-linear influences, and, when such indications are present, seek ways of transforming the linear solution into a non-linear solution involving fewer factors. Both these approaches have been explored in a preliminary fashion, and some initial results have already been determined.

### A Non-Linear PP Model

The most direct approach to non-linear PP factor analysis is to change the factor model to include non-linear relationships. This can be accomplished simply by altering the equations (9), in which the data is estimated on the basis of the current parameters of the model, and then applying the same optimization techniques of Routine TBU to fit the parameters of this set of modified equations to the data. The first step in this process requires the development of an appropriate conceptual foundation for a non-linear three-mode PP factor model. It was decided to formulate a mathematical model that could describe "interaction" of factors.

In many areas of biology, psychology, economics, etc. concepts of basic processes or influences have been developed which would seem to call for interacting factors. For example, it might be expected in analyses of brain activity that certain underlying EEG "generators" will not simply add their effects to those of other "generators". Instead, it is likely that some brain activities will be involved in controlling, or moderating others.

Evidence for "moderator" variables has been presented by Saunders (1956) in relation to different psychological predictors of grade point average. In a study of "Method Factors", Campbell and O'Connell (1967) have uncovered systematic evidence of factors whose influence seems

multiplicative rather than simply additive, and they suggest that these results may cast doubt on the basic appropriateness of the additive factor analysis model for many applications (Campbell and O'Connell, 1967, pp. 424-425). More recently, Ladefoged has cited evidence for the necessity of a "moderator" relationship among some presumed latent dimensions or features of phonetic units (Ladefoged, in press).

A simple PP factor interaction model could be formulated as follows:

$$(22) \quad x_{kji} = o_{k1}a_{j1}F_{1i} + b_1(o_{k1}a_{j1}F_{1i})(o_{k2}a_{j2}F_{2i}) + o_{k2}a_{j2}F_{2i} \dots$$

where the "b" coefficients represent the degree of interaction between two factors. Given m factors, there would be  $m(m-1)/2$  such coefficients, as there are that many distinct factor pairs, not including the case where a factor is paired with itself. If we include terms where a factor "interacts with itself", the model becomes immediately extended to include a contribution from the squared factor value. There are many cases of square and inverse square relationships in the physical sciences, and there seems no basis on which to discount their likelihood elsewhere. In this expanded model there would be  $m(m+1)/2$  interaction terms and corresponding "b" coefficients.

Obviously, there are many other non-linear generalizations which could be made including logarithmic, exponential, and higher order polynomial terms. Unless some guidelines can be found for determining which models to apply to the data, the researcher seems to be faced with an embarrassing excess of possible models. A practical approach to handling this problem will be developed in relation to the linear analysis of non-linear data described below.

#### Analysis by "Fitting" Non-Linear Models

Synthetic error free data was created according to model (22) and then was analyzed by Routine TBU, using a modification of the PP analysis described in Section IV. The modifications consisted in altering the equations being fit to the data (9) to include interaction terms as in equation 22, and requiring TBU to estimate not only the O, I, and M weights (for the o, a, and F values) but also the set of "b" coefficients of interaction. The quick algorithm cannot be used, since it assumes a linear model. Only the more general TBU allows the parameters of any non-linear factor models to be directly fit to the data.

Difficulties of slow convergence were immediately encountered with this technique. Whereas linear solutions of two factors could be accomplished in 50 seconds on the IBM 360/91, convergence on the non-linear solution for two factors and one "b" weight took over 600 seconds. Therefore, only 8 of these non-linear experiments have been carried out. It



had been planned to use techniques of "group relaxation" to speed up the optimization routine, but these have been left undeveloped since more efficient means of accomplishing non-linear analyses seem possible using the technique of transformation from a linear solution, as will be discussed below.

Unique solutions were obtained with these non-linear models which exactly matched the latent non-linear structure used to create the data. Correct estimates of the loadings and of the size of the "b" interaction term were found for data sets generated by two factors and one interaction. Convergence difficulties precluded trying higher numbers of factors. Aside from convergence difficulties, the models and techniques seem to directly generalize to the non-linear factor-interaction case. But because of these convergence difficulties, no meaningful systematic test could be made of conditions of data "adequacy".

#### Linear Analysis of Non-Linear Data

When data created with a non-linear latent structure was analyzed with the linear model of Section IV, a number of interesting consequences were observed. The number of factors uniquely extracted became larger than the "true" number of latent (non-linear) factors. Four linear factors were extracted from data generated by two factors plus two non-linear terms (one "interaction" and one squared factor value). Data generated with a two factor version of the simple interaction model (22) looked, on linear analysis, as if it contained three factors. The best "fit" (least errors of prediction) was only obtained with three factors, and the three-factor solution was still unique.

An examination of the tables of loadings, (Table 10), revealed the most important property of all. The loadings on the third factor were all equal to the product of the loadings of the first two factors.

From a consideration of Jennrich's Uniqueness Theorem (Section V) these interesting results obtained by linear analysis of non-linear data structures could have been predicted in advance. The theorem imposes no restrictions besides minimum rank on the factor loading matrices. Therefore matrices of loadings can be considered in which the loadings of some certain factors are equal to a simple non-linear transformation of the loadings of other factors. Consider, for example, a three factor case where the loadings of one factor of the matrix are equal to the product of the corresponding loadings of the two other factors. When the linear model is used to synthesize data with this type of matrix, it will add the contributions of three "independent" factors, but the contribution of the third "independent" factor would always be exactly equal to the product of the contributions of the other two factors. Now if the third factor were dropped from the matrix, and only two factors were used, but a non-linear equation were used to generate the data according to the model of equation 22, three separate contributions

TABLE 10

## Linear Analysis of Data Generated by Two Interacting Factors

Occasion Loadings (Input)		
	Factor 1	Factor 2
1	0.857	0.857
2	1.285	1.842
3	0.857	0.300

Person Loadings (Input)		
	Factor 1	Factor 2
1	0.625	1.080
2	1.125	0.600
3	1.250	1.320

Test Loadings (Input)		
	Factor 1	Factor 2
1	1.027	0.972
2	0.0	0.972
3	0.840	0.0
	*	* * *

Occasion Loadings (Extracted)			
	Factor 3	Factor 2	Factor 1
1	0.856	0.854	0.650
2	1.287	1.846	2.125
3	0.856	0.300	0.225

Person Loadings (Extracted)			
	Factor 3	Factor 2	Factor 1
1	0.623	1.074	0.665
2	1.125	0.602	0.675
3	1.252	1.324	1.660

Test Loadings (Extracted)			
	Factor 3	Factor 2	Factor 1
1	1.030	1.020	1.068
2	-0.002	0.986	-0.012
3	0.840	0.002	-0.003

Results of linear analysis of non-linear data. The factor loadings of the two input factors were used to generate data according to the non-linear interaction model of equation (22) in Section VI. The normal linear PP model was then used to extract factors, and discovered three (rather than two) unique factors. The loadings on factor 1, however, turned out to be equal to the product of factors two and three.

would again be added to each data value, and these contributions would be equal to the individual contributions of the two factors, plus the product of their contributions. The resulting data for both cases would be identical.

The linear solution of such a data set would be unique for 3 factors (given "adequacy" of the data). The recovered sets of loadings would show two independent factors, and a third factor whose loadings were equal to the product of the loadings of the other two factors. This unique linear solution would be obtained regardless of whether the data were generated by a special three-factor "linear" matrix or instead by a two-factor matrix where the factors interacted in the generation of the data.

If real data were analyzed with a linear factor model and the loadings were found to show such a systematic relationship, the natural conclusion would be that there were in fact only two factors present, and that the third "factor" represented contributions to the data variance from an interaction of the two other factors. This conclusion would be justified because the likelihood of a real third independent linear factor having such a relation to the other two factors would be very small.

As has been noted by Bartlett (1953), and McDonald (1967a), such "product term factors" would arise in classical factor analysis if non-linear relationships were present underlying the data. But they could not be identified because of the rotational indeterminacy of the classic factor solution. Among the many possible rotations only one would reveal the systematic relationship between factor loadings which would correctly indicate the presence and nature of any non-linear influences. In the classical model there was no way to select this rotational position. Because of the uniqueness of the PP solution, this position is selected "automatically", or more correctly, the problem does not arise.

These results suggest that the loadings from any linear PP analysis should be plotted against one another to detect possible non-linear relationships. To detect more sophisticated relations, each mode should be taken separately, and in addition to plotting the loadings, the actual total factor contributions should be plotted against one another. This means that the products of the factor loadings should be computed as for estimation of the data (e.g.  $\sigma_{kl} a_{jl} F_{li}$ ), and that these products, which are here called factor contributions, should be plotted for pairs of factors. This would allow identification of other types of non-linear relationships (e.g. logarithmic or exponential functions of the factor contributions) which could not be as easily identified from plots of the single loadings.

Refined specifications of the relationships could then be computed by conventional curve-fitting techniques. This would allow precise

mathematical description of the form and degree of non-linearity, and give, at the same time, an expression of the degree of error of fit of a certain loading pattern to an assumption of any particular non-linear relationship. Eventually, statistical tests of the probability of any non-linearity hypotheses might be possible. Much work in this area obviously remains to be done before this can be accomplished, however.

This straightforward correspondence between types of linear and non-linear unique factor solutions allows a researcher to (1) recognize the presence of significant non-linear influences underlying his data; and (2) attempt to transform his linear solution into an undistorted non-linear representation involving fewer factors. It was noted earlier that there is an excessively great range of possible non-linear factor models which are available to the researcher. With this "linear" analysis technique, the question of which possible non-linear model to use is put off until the factor loadings are examined. Here the problem can be more easily handled. Visual inspection of the plots of factor loading relationships can directly suggest possible relationships, and can immediately rule out a great many of the relationships which might have seemed plausible before the initial linear analysis. The curve fitting techniques which can be used at this point are more familiar and perhaps easier to use than generalized non-linear PP analysis techniques. Certainly they use less computer time.

Since the quick solution algorithm, which depends on a linear model, provides a great improvement in speed and cost of computation, the ability to apply these linear techniques and still recover non-linear information has important practical significance.

There are a number of somewhat abstract and philosophical questions of possible interpretation of non-linear relations among columns of factor loadings which will not be discussed in detail in this article. Some of these will, however, be touched on in Section VII, in the discussion of the strong relation between the technique for handling non-linearities developed here, and the approach of McDonald.

VII. THE PP MODEL IN RELATION TO OTHER NON-LINEAR  
AND MULTI-MODAL FACTOR MODELS

The foundation of the proportional profiles model is due to Cattell (1944, 1955), but the model and technique which have been developed here from that base are strongly related to the important work of Tucker (1963, 1964, 1966) and Carroll and Chang (1969a, 1969b) in the development of multi-modal models. The work of McDonald (1967a, 1967b) with non-linear generalizations of the classical factor model is closely related to our non-linear generalization of the PP three-mode model, and contributes substantial rigor and detail to its mathematical justification. When these studies are taken together with other recent developments (Shepard and Carroll, 1966; Carroll, 1969; Gnanadesikan and Wilk, 1969; Evans, 1967), it seems that the fields of factor analysis and multi-dimensional scaling are undergoing a process of rapid evolution. The degree to which the results reported in this article are related to all of these other contributions suggests that the process is one of convergent evolution, and that the convergence is towards a small but very general set of powerful techniques for multi-dimensional analysis.

Tucker's General Multi-Modal Factor Model

The generalization of factor analysis from two-way to higher order data sets was first accomplished by L. R. Tucker in 1964 (Tucker, 1964). In contrast to the models developed in this article, he developed a three-mode mathematical model which is much more general, covering the system-variation and object-variation types of linear three-mode data variation in one mathematical description.

Tucker's Mathematical Model

Using notation conventions as adopted earlier, Tucker's mathematical expression of the value of a data point would be as follows.

$$\begin{aligned}
 (23) \quad \hat{x}_{kji} = & o_{k1} a_{j1}^F l_i g_{111} + o_{k2} a_{j1}^F l_i g_{211} + \dots + o_{km} a_{j1}^F l_i g_{m11} + \\
 & o_{k1} a_{j2}^F l_i g_{121} + o_{k2} a_{j2}^F l_i g_{221} + \dots + o_{km} a_{j2}^F l_i g_{m21} \\
 & \dots + o_{km} a_{jp}^F l_i g_{mp1} \\
 & \dots + o_{km} a_{jp}^F q_i g_{mpq}
 \end{aligned}$$

In this expression, a separate set of factors is extracted for each of the three modes of the data matrix, m for the occasion mode, p for

the measures, and  $q$  for the objects. In addition, there is a special set of coefficients or loadings which are defined for each possible combination of a factor from any given mode with a factor from each of the other two modes. These are the  $g_{mpq}$  terms. There are of course ( $m$  times  $p$  times  $q$ ) of these  $g$  terms. The data point is estimated by summing all combinations of the loadings for all factors, times the appropriate  $g$  values. Using more compact summation notation, Tucker's mathematical model can be expressed as follows:

$$(24) \quad \hat{x}_{kji} = \sum_m \sum_p \sum_q c_{km} a_{jp} f_{qi} g_{mpq}$$

or using his notation completely:

$$(25) \quad \hat{x}_{ijk} = \sum_m \sum_p \sum_q a_{im} b_{jp} c_{kq} g_{mpq}$$

Tucker's approach is to solve for a *separate set of factors in each mode*, and then generate a table of interactions or interrelations of these three sets of factors so that, taken together, they predict the individual data points. For a description of the details of Tucker's mathematical procedure and reasoning, the reader is referred to the original sources (Tucker, 1964, 1966; Levin, 1965). The discussion here will center on the relation between Tucker's model and the PP three-mode model.

Applied to the same set of data, Tucker's model and the PP model would not, in general, give the same results. Assuming an "adequate" set of system-variation data, an analysis by the PP model would give a unique explanatory solution. With the same data, Tucker's more general model would not give a unique solution. Clearly, then one could not make the same sort of argument for the "reality" or explanatory validity of the solution with Tucker's model. But an even more fundamental difference would exist between the solutions. The underlying conceptual models would differ so radically that the very meaning of the term "factor" would be different in the two cases. It is at this level of conceptual models that the most important comparisons can be made. The PP mathematical model can be considered a special case of Tucker's mathematical model, namely where the  $g_{mpq}$  terms are 1 for all "diagonal" cells in the core matrix (where  $m=p=q$ ) and zero everywhere else. The conceptual models, however, are altogether different.

#### Tucker's Conceptual Model

The most detailed discussion of Tucker's conceptual model is developed by one of his students, Joseph Levin, in an article dealing with the application of Tucker's techniques to two sets of real data (Levin, 1965, pp. 442-444).

Levin begins his exposition of "the structure and logical meaning" of Tucker's model by stating that

Three mode factor analysis is not a straightforward generalization of classical two-mode factor analysis. Two-mode analysis requires some modification before it can be generalized. It is, therefore, simpler, first to introduce a modified version of two-mode factor analysis and explain three-mode factor analysis as a straightforward generalization. (Levin, 1965, pp. 442-443)

Instead of distinguishing two different latent conceptual models in classical factor analysis and generalizing them separately, Tucker substituted a new two-mode conceptual model, one that was suitable for generalization to the three-mode case. In it, a concept of "factor" was developed which was somewhat different from the classical meaning of the term.

In classic factor analysis, there is a single set of latent variables or influences hypothesized to underlie the data. A factor analysis attempts to describe (by factor loadings) the degree to which this set of hypothesized entities influences the variables, and (by factor scores) the degree to which it influences the individuals being measured. This is diagrammed in Figure 4a.

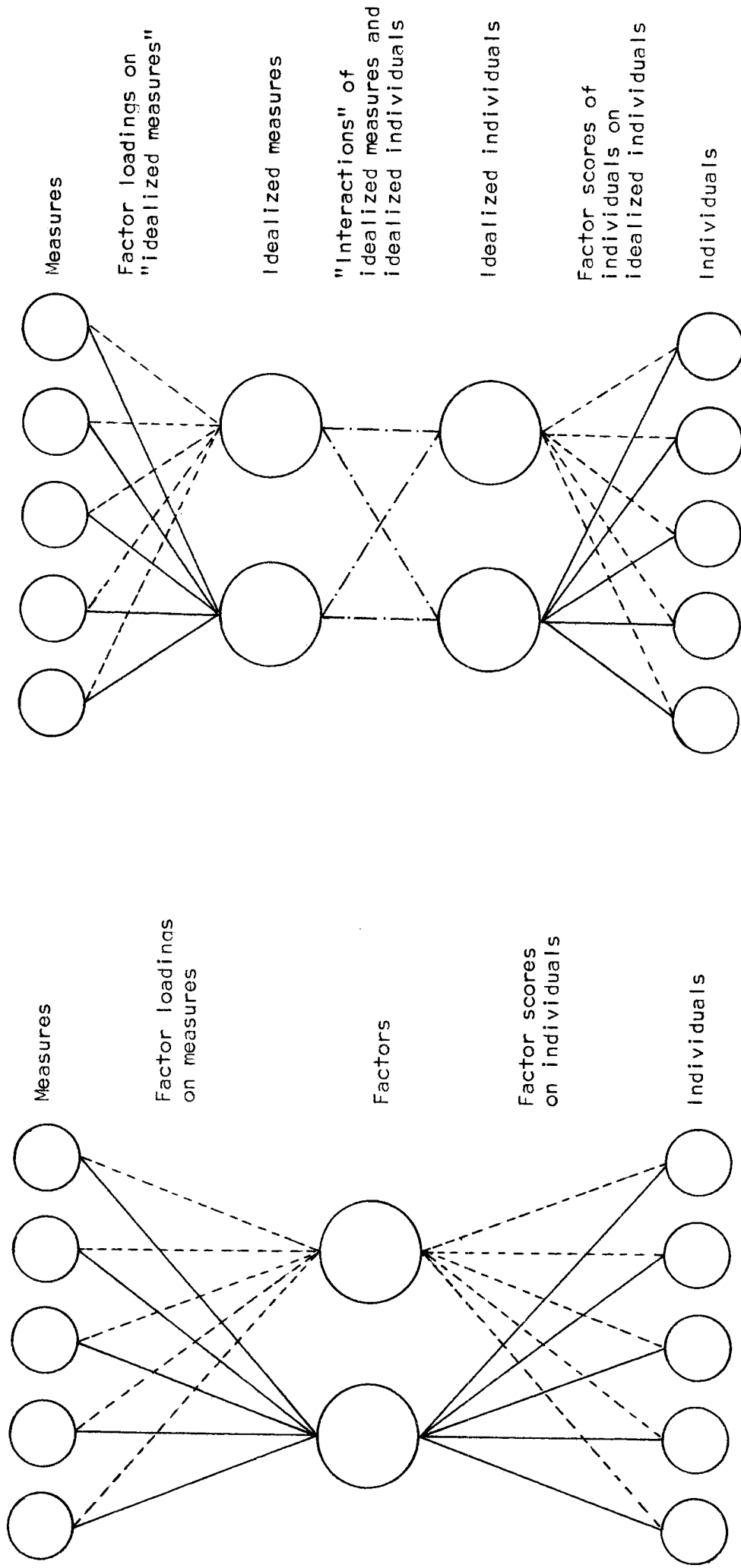
In Tucker's revised two-mode model, factor analysis would identify two sets of hypothetical entities; one set which is measured by the variables, and a *different* set present in the individuals. The entities underlying the variables are called "idealized measures", and those underlying the individuals are called "idealized individuals". The revised factor analysis then attempts to describe three, rather than two, patterns of influence: (1) the influence of the "idealized measures" on the real measures, (2) the influence of the "idealized individuals" on the real individuals, and (3) the "interactions"\* between the two sets of idealized entities. This model is diagrammed in Figure 4b.

Tucker's two-mode model has a straightforward generalization to the three-mode case. This involves a separate set of latent entities for each of the three modes, and a three-way matrix to describe their "interactions".

For system-variation data, Tucker's conceptual approach would seem more complex than necessary. In a case where the PP model would discover 4 factors in a set of three-mode data, Tucker's general model would discover 12 (4 for each mode). Consider the earlier example of factor

---

\* These "interactions" should not be confused with the "interactions" of factors of the non-linear model developed in Section VI. Tucker's "interactions" are between modes, and his model is still a strictly linear one.



Classic Two-Way Factor Analysis

Tucker's Modification of Two-Way Factor Analysis

Figure 4. Comparison of the concept of latent structure in classic factor analysis and in Tucker's modification.



analysis of the economy. The PP model would discover 4 factors with labels such as "inflationary pressure". Each factor would have characteristic influences or loadings on each business, each economic variable of that business, and on each occasion. Tucker's model would discover 4 "idealized businesses", 4 "idealized business measures", and 4 "idealized occasions". Each idealized business would contribute in varying degrees to the actual businesses, each idealized business measure would contribute in varying degrees to the actual variables, and each idealized occasion would contribute in varying degrees to the actual occasions. The core matrix would describe interactions among these idealized businesses, business measures, and occasions. Although both descriptions might be interesting and informative, Tucker's general approach, applied to simple system-variation data, clearly seems less parsimonious. Both models would predict or "fit" the data equally well, but Tucker's modification requires the postulation of three times as many hypothetical entities as the PP model. It also requires an additional set of relationships to be invoked, namely the "interactions" of the three sets of entities. The interpretation of entities such as "idealized businesses" and their interactions with "idealized occasions" seems somewhat more obscure than interpretation of system-variation factors, which are properties of the economy which differentially affect businesses, and which vary in activity or influence from one occasion to the next.

One might feel tempted to anticipate, then, that Tucker's conceptual model could be set aside in favor of the system-variation and object-variation models. Once techniques of analysis according to the object-variation model were developed to the same practical degree as PP factor analysis, one might feel that this model could cope with the cases not accessible by the PP model, and one way or the other an easily interpretable "explanatory" solution could be obtained for any "adequate" set of three-mode data.

Such, however, is not the case. Tucker's approach may seem less explanatory and parsimonious for simple cases of system-variation or object-variation. But other patterns of data variation can be envisioned for which neither of these two models (or even a combination of both models) would be completely adequate. For such cases of complex three-mode data variation, models must be developed that bear a strong relation to Tucker's pioneering general three-mode model.

To develop the idea of these complex patterns of three-way data variation, it will be useful to resume the discussion begun in Section III of the object-variation model and its relation to the PP model.

#### Object-Variation Model Applied to Non-Object-Variation Data

The object-variation concept was given mathematical form in equation 4. But this equation applies to other patterns of data variation besides the object-variation type. It is, in fact, a fully general

(linear) mathematical model for three-mode data variation. Compared to Tucker's general model it gives a representation of the data which invokes fewer hypothetical entities (since it finds one set of factors, i.e. those underlying the measures) but is much less compact than Tucker's three-mode description. In the object-variation model the factor score variations (the "v" values) for all individuals must be solved separately for each factor for each occasion. As a result, the solution not only generates a set of loadings for the factors on the variables, but a two dimensional array of "v" values for each factor. This yields a three-way matrix of reduced size, whose modes are factor by individual by occasion.

This loss of compactness in two modes might be justified if the model allows one to discover (by applying the continuity maximization constraint) a compact set of loadings for the third mode which represents a good estimate of the real latent structure underlying the variables. But such a result would not provide any information about the latent structure of the individuals or the occasions.

Sometimes, it might be possible to simplify or further analyze the tables of "v" values for each factor. A classic factor analysis of each factor's "v" table might reveal "secondary factors"\* indicating a strong latent structure in the variations of the factor scores across individuals and occasions. These secondary factors could take many forms. One can predict for example what such an analysis would reveal if the data were of system-variation or object-variation type.

If the original data had been of system-variation type, then an analysis of any factor's "v" table would give one secondary factor. The loadings of this one factor would be, in fact, the proportionality constants among occasions and among individuals which would have been discovered had the original data been analyzed using the PP model.

Interpretation of such a "v" table analysis would be straightforward. And since one-factor solutions are unique, even with traditional two-mode factor analysis, this roundabout system-variation solution would retain the uniqueness of the system-variation model.

Similarly straightforward results would be found if the original data had been a perfect example of object-variation. All the "v" values for one individual would be uncorrelated with those of any other individual. In this case the "v" tables would have as many factors as individuals, and there is no additional theoretical information to be gained by such analysis.

---

\* These are factors in the patterns of "v" values, not to be confused with "higher order" factors obtained by hierarchical factor extraction in classical two-mode factor analysis.

But now consider intermediate situations. Suppose that a secondary factor analysis of each primary factor's "V" table reveals two or three factors in each table. How does one interpret this systematic but more complicated pattern of variations in factor scores? Could such a pattern ever be expected to occur in real data?

The interpretation of intermediate data relationships in the "V" tables seems much more difficult than the two simple extreme cases of 1 factor or N factors (where there are N persons), but an analogy to the N factor case may be helpful. In the case of N factors for N individuals one would interpret this result as indicating that every individual is exposed to different circumstances which independently influence his variations in factor scores across occasions. In the case of personality measurements and factor types such as "hostility", one might suspect that the individual personal family and social circumstances of each person would change the degree of hostility in a person's makeup in a fashion independent of the changes in other individuals.

Now if the individuals being measured came from three large communal living groups (e.g. orphanages, "communes", religious communities, etc.) some of the variations in the amount of hostility might be correlated among those individuals coming from a given living group, i.e. as each group went through "good" and "bad" phases. Three secondary factors might then emerge from an analysis of a given factor's "V" table. Each secondary factor would correspond to a living group. If enough "V" tables displayed three secondary factors, the separate secondary solutions could be rotated to parallel proportional profiles for a possibly unique solution.

In the above interpretation, a separate set of factors has been isolated in each of two modes of the data matrix. Further analysis might sometimes reveal separate factors in the third mode. This concept of separate factors in each mode is similar to Tucker's concept of "idealized" entities in each mode. Interpretation of such interlocking latent structures can become very difficult. The approach to this problem through the analysis of "V" tables in the object-variation model may shed some light on possible criteria for unique explanatory solutions with these complex patterns of data variation. It may, for example, provide one approach to the discovery of rotation criteria for Tucker's three-mode model. As Tucker comments: "... the freedom of transformation permitted by [Tucker's] model is both important and the source of many problems. There is a lack of uniqueness. This gives rise to many problems yet to be solved" (Tucker, 1966, p. 291). In the difficult area of these complex types of three-mode data variation, Tucker's concepts and the concepts developed here have much in common and much to gain from one another (for recent work on meaningful rotation criteria see Tucker 1966b, 1967).

### McDonald's Non-Linear Factor Analysis

McDonald's discussion of the mathematical foundations for non-linear factor analysis (McDonald, 1967a, 1967b) relate directly to the technique of non-linear generalization by special analysis of a linear solution developed in the latter part of Section VI. His important monograph (McDonald, 1967a) should be consulted by those seeking formal mathematical development of the application of the basic factor model and equations to cases of non-linear latent influences.

Not only has McDonald laid an important mathematical foundation for non-linear factor analysis, but he anticipated aspects of the particular short-cut analysis technique which was developed in part VI to avoid the slow-convergence problems of the generalized optimization technique. This short-cut involved searching among the loadings or contributions of each pair of "linear" factors for systematic non-linear relationships. McDonald confirmed Bartlett's (1953) claim that a non-linear contribution would act as if it were an additional factor in the linear extraction process (McDonald, 1967a, 1967b).

There are several important differences, however, between McDonald's non-linear factor analysis and the non-linear generalizations of PP factor analysis developed in Section VI. The most crucial difference is that the results of non-linear PP factor analysis are unique and "explanatory". The indeterminacy which is present in solutions obtained by the application of McDonald's technique to normal two-mode factor analyses is quite serious. Not only are the results indeterminate in terms of identifying the "meaning" of the factors, but even the basic form of the non-linear relationships is indeterminate as well. McDonald demonstrates that one basic non-linear model, that of interacting factors (similar to our equation 22) can be transformed by rotation into a fundamentally different model involving non-interacting factors, calling only for contributions from the squares of the factor loadings. He comments that "The alternative interpretation of data in terms of model (19) [non-interacting polynomial components] would entail a radically different psychological theory, and a choice between model (1) [interacting factors] and model (19) [certain polynomial factors] may rest on psychological considerations, if, indeed, a rational choice can be made" (McDonald, 1967b, p. 214). Such indeterminacy, and pessimism about "rational choice" among the different possible solutions would seem to cast a serious shadow on any attempt to discover the "true" nature and structure of a set of non-linear latent relationships.

The whole question of rotation posed the gravest problems when McDonald's non-linear generalizations were applied to classical two-mode factor analysis:

... the concept of simple structure may be difficult or impossible to redefine in the context of non-linear factor analysis. The whole question of factorial invariance will need to be reappraised in this context ...

... Thurstone (1947) adopted the position that factors, conceived of as fundamental dimensions of ability, personality or the like, need not be thought of as orthogonal ... There is a good deal of impressive argument, by analogy with observable dimensions of physical entities, to justify this view. For our purposes, however, it proved necessary to stipulate that the latent variates or factors be not merely orthogonal, but completely mutually independent in the probability sense. If this stipulation is not made, there can be no distinction between linear and non-linear factor analysis. A fortiori, we cannot allow the factors to be correlated. (McDonald, 1967a, pp. 133-134)

Many of the dilemmas that confronted McDonald do not apply to the non-linear generalization of the PP three-mode model. Most importantly, the uniqueness of the solution provided by the PP model allows a rational choice to be made between the alternative interaction and polynomial models which were confused in the non-linear two-mode analysis. Furthermore, this uniqueness allows a much more direct and straightforward search for all the non-linear relationships that may be present among factors. McDonald's techniques were restricted basically to polynomial models (McDonald, 1967b, p. 133), although other relations could be theoretically derived by transformation from a series of polynomials.

Since no rotations are necessary (or possible) with the unique solutions of the PP three-mode model, the researcher is able to directly plot the loadings or contributions of each factor against the others, and to use curve fitting and optimization techniques to discover latent non-linear relationships of all kinds, including logarithmic, trigonometric, exponential, etc.

It is also important to note that the restriction to orthogonal factors is not present with PP analysis. As long as one assumes that highly specific non-linear patterns of correlation among factors is unlikely, then it would be easy (in principle) to distinguish plots in which there is essentially a linear correlation but a good deal of random scatter, from those in which there is less scatter, but a clearly defined non-linear relationship, such as would be unlikely unless one of the apparent factors represented a non-linear contribution of the other factor.

Clearly, however, there are ambiguities and philosophical difficulties in the interpretation of some of these relationships. A trivial one would be to determine whether one factor is the square of the second, or in fact the second is the "real" factor, and the first is a non-linear contribution which is a function of the square root of the second. More significant is the possibility of non-linear correlations among genuinely different factors. Perhaps some of these ambiguities

might be resolved in four-mode or higher order data sets. In other cases, one may be thrown back to the "most meaningful" interpretation. A detailed discussion of the philosophical and mathematical considerations involved in these decisions, insofar as they might matter (i.e. insofar as they might lead to different interpretations of the real phenomena underlying the data) must be deferred.

#### Carroll and Chang's Model for Multi-Dimensional Scaling

As an earlier version of this article was being prepared, the author was given copies of two unpublished papers by J. Douglas Carroll and J. J. Chang of Bell Telephone Laboratories (Carroll and Chang, 1969a, 1969b). It became clear from these most important articles that Carroll and Chang had independently discovered some of the most significant properties of the representation of data in terms of triple products, as in equation 9. Further, they had successfully applied this scheme to real data. For the interpretation of their results they developed a clear conceptual model which is related to the PP model, but which arises from considerations in multi-dimensional scaling rather than factor analysis.

Carroll and Chang deal with a three-way data matrix composed of judged similarities between stimuli. The dimensions of the matrix are (1) the stimulus, (2) the stimulus to which it is being compared, and (3) the individual doing the comparison. They conceive that there should be systematic differences between the similarity values not only based on which two stimuli are being compared but also based on which individual does the comparison. Assuming that the individuals, in general, share a number of latent "dimensions" on which they compare the stimuli, they nonetheless can be expected to differ on the relative weights that they give to different dimensions (Carroll and Chang, 1969b, p. 3).

The conditions of a system-variation conceptual model are fulfilled in the above description. But an adjustment is necessary before the final form of the mathematical model emerges. The theoretical expression for the judged distance between two stimuli is

$$(26) \quad d_{jk}^{(i)} = \sqrt{\sum_{t=1}^r w_{it}(x_{jt} - x_{kt})^2}$$

for the distance judgment of the  $i^{\text{th}}$  individual on the  $j^{\text{th}}$  and  $k^{\text{th}}$  stimuli (Carroll and Chang, 1969b, p. 3). It is the normal Euclidean distance formula modified by the weight ( $w_{it}$ ) that the  $i^{\text{th}}$  individual puts on the  $t^{\text{th}}$  dimension. The data is then transformed as follows:

The first step in the method of analysis is to convert the similarities into distance estimates. Under the linear assumptions we have made, this can be done by using one of the standard procedures described in Torgerson (1958). We then use the equations also described in Torgerson (1958, pp. 254-259) to convert the distance estimates for each subject into scalar products of vectors (to get the matrix of scalar products, we simply double center the matrix whose entries are  $-1/2 d_{jk}^2$ ). This gives us numbers  $b_{jk}^{(i)}$ , which, in the present case, can be regarded as scalar products between the vectors  $y_{jt}^{(i)}$ , i.e. (ignoring error terms):

$$(27) \quad b_{jk}^{(i)} = \sum_{t=1}^r y_{jt}^{(i)} y_{kt}^{(i)} = \sum_{t=1}^r w_{it} x_{jt} x_{kt}$$

The last expression results from substituting  $y_{jt}^{(i)} = w_{it}^{1/2} x_{jt}$ .

The rightmost expression in (27) above is equivalent to the mathematical form of the PP three-mode model (2) as modified to estimate data (9). That it shows up in the context of multi-dimensional scaling is further evidence of its generality and usefulness.

Carroll and Chang had made the important discovery that this procedure could give a unique solution, although they did not present any proof or any description of the conditions under which uniqueness would be found to hold or fail. Instead, their highly successful results with real data were used to demonstrate the explanatory validity of the unique solutions obtained with the model:

The model may not hold in every case, but if it does we gain a unique and hopefully psychologically meaningful orientation of axes, thus obviating the rotational problem and defining much stronger scales of measurement than is usual in multidimensional scaling. One example will be presented in a later section to support the argument that this in fact does happen. We have now collected many more cases (especially as yet unpublished data collected by Myron Wish) that lend credence to this notion. In essentially every case the dimensions have proved to be interpretable directly as they are derived from this analysis (i.e., without rotation). In cases where a set of *a priori* physical or theoretical dimensions were known, the recovered (unrotated) dimensions have always (to date) corresponded to them in an essentially one to one fashion. We therefore argue that it is appropriate to analyze data in terms of this very strong and specific model, and that only if this model fails to fit the data adequately should one have recourse to a more general model.

(Carroll and Chang, 1969b, pp. 6-7)

To analyze data according to this model, they devised an algorithm based on a procedure described by Eckart and Young (1936) which is similar to Jennrich's quick algorithm (before the addition of the relaxation factor). They have programmed their version to handle up to 7-mode data.

### Converging Models

Other recent developments in factor analysis seem to be pointing in a direction similar to the PP model and the work of Tucker, McDonald, and Carroll and Chang, discussed above. Evans (1967) has developed a special purpose factor model for three-mode "growth data" which shares many features with the object-variation model developed in Section III. It expands the factor loadings on measures rather than the individual's factor scores, and provides special features for arbitrary means and variances found in most psychological tests. Although his model has no unique solution, Evans discusses the desirability of invoking Cattell's principle of parallel proportional profiles for rotation to a final solution, and he presents a computer algorithm for maximizing such proportionality of factor loadings on measures across occasions. Horst (1965) has also been working with techniques for handling three-way data matrices. Non-linear generalizations of two-mode factor analysis have recently been formulated by Carroll (1969) and Gnanadesikan and Wilk (1969). But these non-linear techniques, like McDonald's, are subject to problems of rotational indeterminacy which do not arise with the PP model.

As has been noted earlier, there appears to be a rapid and convergent evolution taking place in certain areas of factor analysis and multi-dimensional scaling. As one result of this evolution, the model here called the proportional profiles model begins to emerge as one answer to the dual objectives of both greater generality (to incorporate multi-modal and non-linear data relationships) and greater strength or explanatory validity (to give stronger reasons for believing that a given solution corresponds in some close fashion to the "true" structure of the latent influences that generated the data relationships).

This evolutionary progress calls for reexamination of the fundamental conceptual models underlying the mathematical procedures which have been developed. Examination of the meaning and appropriateness of different types of mathematical description of data relationships, as was done in this article, can lead to better approaches to "explanatory" solutions.

Further conceptual, perhaps even philosophical epistemological questions need to be clearly formulated and the answers explored, before the modeling of non-linear and the more complex types of three-mode variations can be placed on a firmer "explanatory" foundation. (This situation is perhaps reminiscent of the philosophical questions that confronted the



early development of multiple-factor analysis itself.) Renewed exploration at this fundamental level, coupled with anticipated rapid advances in computing techniques, make it likely that a number of significant extensions of factor analysis, multi-dimensional scaling, and other related areas will emerge in the coming decade.

## References

- Bartlett, M. S. Factor analysis in psychology as a statistician sees it. In *Uppsala Symposium on Psychological Factor Analysis*, Nordisk Psykologi Monograph No. 3, 1953.
- Campbell, D. T. and O'Connell, E. J. Methods factors in multitrait-multimethod matrices: multiplicative rather than additive? *Multivariate Behavioral Research*, 1967, 2, 409-426.
- Carroll, J. D. Polynomial factor analysis. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 1969, 103.
- Carroll, J. D. and Chang, J. J. A new method for dealing with individual differences in multidimensional scaling. Unpublished manuscript, Bell Telephone Laboratories, Murray Hill, New Jersey, 1969a.
- Carroll, J. D. and Chang, J. J. Analysis of individual differences in multidimensional scaling via a n-way generalization of "Eckart-Young" decomposition. Unpublished manuscript, Bell Telephone Laboratories, Murray Hill, New Jersey, 1969b.
- Cattell, R. B. "Parallel Proportional Profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 1944, 9, 267-283.
- Cattell, R. B. *Factor analysis*. New York: Harper and Brothers, 1952.
- Cattell, R. B. and Cattell, A. K. S. Factor rotation for proportional profiles: analytical solution and an example. *British Journal of Statistical Psychology*, 1955, 8, 83-92.
- Comrey, A. L. Tandem criteria for analytic rotation in factor analysis. *Psychometrika*, 1967, 32, 143-153.
- Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211-218.
- Evans, G. T. Factor analytical treatment of growth data. *Multivariate Behavioral Research*, 1967, 2, 109-134.
- Fant, C. G. M. *Acoustic theory of speech production*. The Hague: Mouton, 1960.
- Gebhardt, F. A counterexample to two-dimensional varimax rotation. *Psychometrika*, 1968, 33, 35-36.

- Gnanadesikan, R. and Wilk, M. B. Data methods in multivariate statistical analysis. In P. R. Krishnaiah (ed.), *Multivariate analysis: II*. New York: Academic Press, 1966.
- Green, B. F. The computer revolution in psychometrics. *Psychometrika*, 1966, 31, 437-445.
- Harman, H. H. *Modern factor analysis*. Chicago: Univ. Chicago Press, 1960.
- Harman, H. H. and Jones, W. H. Factor analysis by minimizing residuals (MINRES). *Psychometrika*, 1966, 31, 351-368.
- Haverland, E. M. The application of an analytical solution for proportional profiles rotation to a box problem and to the drive structure in rats. Ph.D. thesis, University of Illinois Library, Urbana, Illinois, 1954.
- Horst, P. *Factor analysis of data matrices*. New York: Holt, Rinehart and Winston, 1965.
- Jöreskog, K. G. Testing a simple structure hypothesis in factor analysis. *Psychometrika*, 1966, 31, 165-178.
- Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- Ladefoged, P. *Three areas of experimental phonetics*. London: Oxford Univ. Press, 1967.
- Ladefoged, P. The measurement of phonetic similarity. *Statistical Methods in Linguistics*, in press (1970).
- Ladefoged, P. *Preliminaries to linguistic phonetics*. Forthcoming.
- Levin, J. Three mode factor analysis. *Psychological Bulletin*, 1965, 64, 442-452.
- McDonald, R. P. Nonlinear factor analysis. *Psychometric Monograph No. 15*, Chicago: Univ. Chicago Press, 1967a.
- McDonald, R. P. Factor interaction in nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1967b, 20, 205-215.
- Reichenbach, H. G. Routine TBU: routine for optimizing a set of variables. Report 62-21, Institute of Transportation and Traffic Engineering, University of California at Los Angeles, 1962.

- Reichenbach, H. G. Routine TBU: routine for optimizing a set of variables. (Unpublished expanded version of Reichenbach, 1962), 1969.
- Saunders, D. R. Moderator variables in prediction. *Educational and Psychological Measurement*, 1956, 16, 209-222.
- Shepard, R. N. and Carroll, J. D. Parametric representation of nonlinear data structures. In P. R. Krishnaiah (ed.), *Multivariate analysis*. New York: Academic Press, 1966.
- Thompson, G. H. and Ledermann, W. The influence of multivariate selection on the factorial analysis of ability. *British Journal of Psychology, General Section*, 1939, 24, 288-306.
- Thurstone, L. L. *Vectors of mind*. Chicago: Univ. Chicago Press, 1935.
- Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.
- Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (ed.), *Problems in measuring change*. Madison: Univ. Wisconsin Press, 1963, 122-137.
- Tucker, L. R. The extension of factor analysis to three-dimensional matrices. In N. Frederiksen (ed.), *Contributions to mathematical psychology*. New York: Holt, Rinehart and Winston, 1964.
- Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966a, 31, 279-311.
- Tucker, L. R. Learning theory and multivariate experiment: illustration by determination of generalized learning curves. In R. B. Cattell (ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally and Co., 1966b.
- Tucker, L. R. Three-mode factor analysis of Parker-Fleishman complex tracking behavior data. *Multivariate Behavioral Research*, 1967, 2, 139-151.
- Wrigley, C. and Neuhaus, J. O. The matching of two sets of factors. Contract Memorandum Report, A-32, Urbana, Illinois: Univ. Illinois, 1955.