



Universiteit
Leiden
The Netherlands

Datadozen: Analyse en historie

Kroonenberg, P.M.

Citation

Kroonenberg, P. M. (2005). *Datadozen: Analyse en historie*. Leiden: Universiteit Leiden.
Retrieved from <https://hdl.handle.net/1887/3494>

Version: Not Applicable (or Unknown)

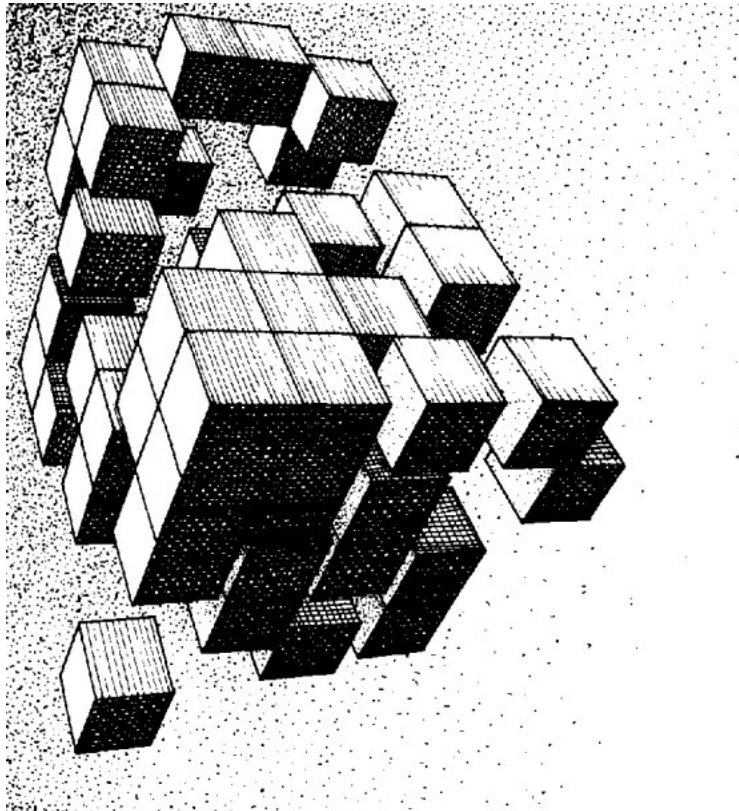
License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3494>

Note: To cite this publication please use the final published version (if applicable).

Prof. dr. P. M. Kroonenberg

Data boxes: Analysis and history



Universiteit Leiden

Cover illustration: Siep Kroonenberg (1983)

Data boxes: Analysis and history

Inaugural address by

Prof. dr. P. M. Kroonenberg

upon his accession to the
Special Chair of Multivariate Analysis (in particular of three-way data)
at Leiden University
on Friday, September 2nd, 2005

*Meneer de Rector Magnificus,
Leden van het bestuur van de Stichting ter Bevordering van de Wetenschap der
Datatheorie,
Leden van het Curatorium van deze bijzondere leerstoel,
Zeer gewaardeerde toehoorders,
including those of you who are unfortunately not able to understand Dutch, and who
will have to content themselves with enjoying this special occasion by looking at the
pictures in the hand-out and by inspecting these august surroundings.*

Once, as a young university lecturer, I visited a colleague at the University hospital in Amsterdam, where I chuckled at a nameplate on a door announcing that this was the office of an Assistant Professor of the Biochemistry of the Oral Cavity. I can imagine you experienced similar merriment at the esoteric description of my chair, "Multivariate analysis, in particular of three-way data". Multivariate analysis, or analysing several variables simultaneously, seems just about possible, and is common practice in many branches of science, but "especially of three-way data" might sound as abstruse to you as "of the oral cavity" sounded to me, long ago. Is it really necessary to institute a special chair for something so outlandish?

In this lecture I will try to convince you that my specialism is not nearly as esoteric as you might think. More in particular, I will argue that scientific research cannot do without three-mode analysis, even though not everybody knows this yet. In addition, I will show that three-way data are found everywhere, if one only looks; much like first-time prospective parents who suddenly notice pregnant women everywhere.

Before embarking on my explanation, I would like to say something about this room and the limitations it imposes on my lecture. Nowadays, scholars worth their salt cannot do without slick visual presentations full of moving images and exciting sound effects. However, the Leiden tradition, probably partly inspired by this medieval hall, forbids the use of audiovisual tools to embellish inaugural lectures. For a subject such as mine this is both a handicap and a challenge. Much of multivariate analysis strongly leans on presentations of data and results in the form of charts and three-dimensional images, and these may, of course, be shown perfectly on slides. As a compromise, I have made a booklet for you containing the illustrations to my lecture.¹

In this lecture, I will first explain what three-way data are and how they are tackled. By means of a detailed example I hope to give you an idea of why three-mode analysis fascinates me so much. Next, I will tell you something about the origin of three-mode analysis, its explosive development over the past ten years, and how I am planning to use my academic assignment in the next five years. In the third and last part of my lecture I will express my thanks to a number of people.

Three-way data and their analysis

What are three-way data?

To put it at its simplest, three-way data are data that no longer fit onto one index card, but need a box to contain them. You might still find that mystifying, but my first

illustration may help you here. Looking, for instance, at an example from Child Studies, we note that many data sets there take the form of the scores of a number of children on a number of variables, such as intelligence, and marks for English, Arithmetic and physical education (Gym), etc. (see Figure 1).

2000

	IQ	English	Arithm.	Gym
Barry	110	8	9	5
Mary	90	5	6	9
Anne	130	9	8	7
Charlie	95	7	6	8
Tim	100	7	8	9
Lea	101	6	5	5
Erica	97	6	6	9

Figure 1. Two-way data

2005

2004

2003

2002

2001

2000

	IQ	English	Arithm.	Gym
Barry	110	8	9	5
Mary	90	5	6	9
Anne	130	9	8	7
Charlie	95	7	6	8
Tim	100	7	8	9
Lea	101	6	5	5
Erica	97	6	6	9

Figure 2. Three-way data

Such data may be written on one index card, in the form of a table with the children in the rows and the variables in the columns. We now have two ‘ways’: children and variables. If the data form part of a longitudinal investigation in which the children are measured every year on the same variables, we have three-way data, the ‘ways’ being children, variables and, thirdly, years or occasions. One index card will now no longer suffice, but we need a separate card for each year, and the set of cards is kept in a box: three-way data fit in boxes, as shown in Figure 2.

Once we have got the hang of this, we suddenly notice data in boxes everywhere. A *plant breeder* has planted several varieties of peanuts (first way) in different locations (second way) and measures the characteristics of the harvested plants, such as yield, quality, the percentage of saturated and of unsaturated oils (third way).² A *chemist* has ten containers with different sugar solutions, sends light of varying wavelengths through the vats, and measures the excitation wavelengths emerging at the other side.³ A *church historian* wanting to describe the polemics on modernism in the middle of the 19th century might make a box with polemicists (first way), their opponents (second way), and arguments used (third way).⁴ A *medical pharmacologist* has derived the spatial structures of a number of cocaine variants from crystallographic measurements, and wants to know to what extent their spatial structures are the same, and to what extent they are different. The molecules form the first way, their constituent atoms the second, and the spatial coordinates the third.⁵

Ladies and gentlemen, you may sometimes have wondered how a radio mast, which receives signals not only from your mobile phone but also from countless other cell phones, knows exactly that the answer to your outpourings should be sent to you, rather than your neighbour who should not be privy to them at all? This process, too, may be described through three-way models.⁶

What is "Multivariate analysis, in particular of three-way data"?

Three-mode analysis is no more than the analysis of data that fit in boxes, but that is not going to help you much. It is probably better to rephrase the question (a familiar

trick among statisticians if they cannot or will not answer the original question). One possibility is: "What type of research questions can be tackled via three-mode analysis?" Such a question has the advantage that it is formulated in substantive rather than methodological terms.

Let us again look at the example of the children tracked over a number of years. What questions would the researchers have had in mind when they started to collect data? There are of course many possibilities, but in this lecture I want to confine myself to those questions that may be handled via three-mode analysis. In this case, the *central questions* might be:

1. What are the relations between the variables?
2. What trends may be discovered over time?
3. Are there different types of children?

These are three questions, one for each way. Although this type of question is interesting, they only apply to one way at a time. Three-mode analysis has been devised especially to deal with more complex questions such as:

4. Do the relations between the variables change over time? For instance, it is a well-known fact that in very young children intelligence is still very amorphous, but that as they get older some children develop better on some aspects of intelligence than others. In other words, time brings a change in the structure, i.e. the interrelations between the various parts, of an intelligence test.
5. An even more complex question is: Does the structure of the variables change over time in a different way for different groups of children, for instance for boys and for girls, or for children with different levels of mental handicap?

With such complex questions, involving all three aspects, or ways, of the data, three-mode analysis really comes into its own. *Plant breeders*, for instance, are interested in the specific adaptation of crops, in other words, they like to know which varieties of a plant will perform well on specific aspects in locations with specific characteristics. In concrete terms, where should one grow what type of peanut plants, in order to obtain peanuts that are specifically suitable for making peanut butter? Do specific schools of *modernist theologians* attack specific other schools, using specific types of arguments?

One's first acquaintance with techniques for three-mode analysis is often a bit of a shock, because of the complications involved in the correct understanding and interpretation of those techniques. However, this is unavoidable, as we are dealing with complex techniques intended to solve complex questions. Testing differences in average lengths between boys and girls is child's play for the average student, but three-way questions just are more intricate. However, it is this complexity that makes research more rewarding.

I am always surprised to find how often researchers content themselves with univariate answers in situations that are inherently multivariate. No wonder, to paraphrase the Dutch columnist Jan Blokker, that most answers from social sciences research have long been known to the Wise Woman of the Western Woods, and that many people upon reading these results think: "Well, that's nothing new." The snake in the grass here is that if the results had been different, others, or even the same people, would also have said that they knew this all along.

As a proof of this I may mention a trick of A.J.F. Köbben's,⁷ who in his Cleveringa lecture of 1980 presented some research results regarding the behaviour of immigrants. These brought nods of approval and agreement in the auditorium. A few minutes later, he returned to the subject and asked the audience whether they had really believed him: actually, the results were the complete opposite of what he had told them earlier.

The crux of the matter is that, as a result of the multivariate character of phenomena, under one set of circumstances one specific univariate result will be obtained, whereas the opposite result may be found in other circumstances. In short, univariate results are often misleading in a multivariate world, and the multivariate character of a particular phenomenon needs to be probed by means of techniques that do justice to this complexity, such as those in the toolbox of the three-way data analyst.

Figure 3 shows a typical example of such a three-way data analyst, looking in some despair at the complexity of the data cube he has been given. This woodcut by M.C. Escher, by the way, would fit perfectly in Jan Luyken's 1694 collection of emblemata, *Het Menselyk Bedryf* (The Human Enterprise), together with the fisherman, the baker, and the blacksmith.

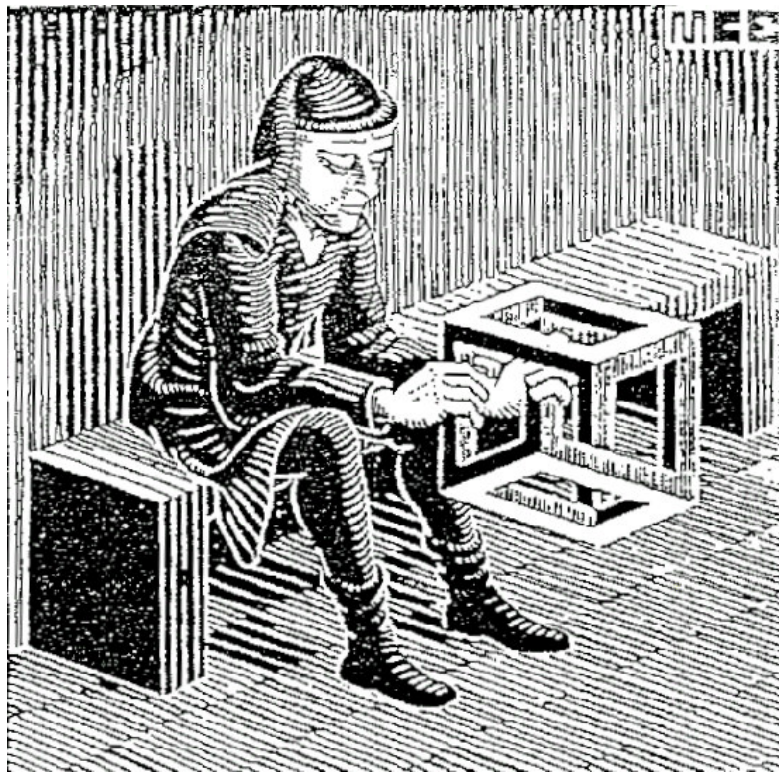


Figure 3. M. C. Escher, *Man with cuboid*
© 2005 The M.C. Escher Company B.V., Baarn - Holland.
<http://www.mcescher.com/>. All rights reserved.

How were three-way data analysed before the arrival of three-mode analysis?

To put it briefly: by flattening the box, or stringing out its contents. In both cases the idea is to make three-way data into two-way data by eliminating one of the ways. Instead of looking at the interactions between three types of units (or ways), one then only needs to analyse two.

Flattening is typically done by taking the averages over all cards in the box, so that one is left with one card containing means. These may be averages over all years, so that one loses sight of trends over time, but one may also take averages over all subjects, so that individual differences disappear below the horizon.

Stringing out is done by either laying out all the cards in one long row, so that the relation between similar variables at different time points is neglected, or laying out all cards in one tall column, so that the connections between people's scores at different moments is lost. In all these cases, the data and their analysis are shrunk from three-way to two-way.

Sometimes this may do no harm, because it is possible that a three-mode analysis leads to the conclusion that no three-mode analysis is necessary. For instance, if nothing changes over time, or if all subjects may be viewed as having been randomly drawn from one single population. Or if all theologians attack all opponents with the same intensity with all relevant arguments. However, if this is not the case, the flattening or stringing out of three-way data leads to an unnecessary and sometimes unacceptable simplification.

So... what should we do with three-way data?

Since the beginning of the 1960s, a series of techniques has been devised specifically aimed at doing justice to three-way data. They bear such intriguing names as three-mode principal component analysis, three-mode factor analysis, three-mode cluster analysis, parallel factor analysis, multiway covariance analysis, multidimensional scaling techniques for individual differences, generalized Procrustes analysis, multivariate longitudinal analysis, and many more of this kind.

Most of these methods have a strongly *exploratory character*, which means that one tries to find the patterns among the elements of the three ways, without a priori postulating specific configurations, and without applying tests to these patterns. This is partly because it is difficult to specify such patterns beforehand, and partly because 'hypothesis testing' supposes that something is known about the distributions of the scores, which for three-way data is only very rarely the case. It is, however, perfectly possible to determine the stability of the estimated values of the parameters via repeated sampling from the sample in question ('bootstrap'-method), but these developments are still in their infancy in three-mode analysis.⁸

How do Japanese students judge Chopin's Preludes?

In order to give you some idea of the sort of results obtained via three-mode techniques, I would like to present an analysis of data produced by Japanese students, in reaction to listening to (parts of) the 24 preludes by the Polish composer Chopin played by a French pianist. What makes these data special is that they combine psychology, semantics, cross-cultural studies, musicology, and statistics. An example close to my heart, as some of you may affirm who have earlier heard me speak about this study, in a slightly different form. Moreover, in this example a type of data is used, called 'semantic differentials', that actually initiated the development of three-mode analysis, as I will explain in more detail later. The analyses have been produced in close collaboration with Professor Takashi Murakami of Nagoya University, Japan⁹.

The *research questions* were the following:

1. Music may be characterised in *technical* terms such as key, tempo, and mode, i.e., major/minor. To what extent are *laymen* sensitive to these aspects when we ask them to describe music in everyday adjectives such as loud, gloomy, tender, etc.?
2. Is it possible at the same time to establish, by means of the same adjectives, their preference for specific types of music, not expressed in those adjectives but in the technical terms relating to key signature, tempo and mode?

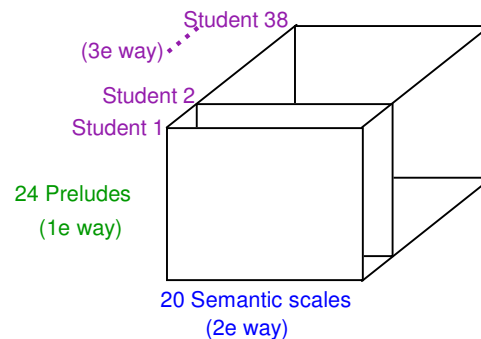


Figure 4. Data box for the Chopin study

The study was conducted among 38 Japanese students who were familiar with classical music (the first way). The 24 preludes were played to them (the second way), and after every prelude (or the first 80 seconds of it) they were asked to fill out a set of 20 so-called semantic rating scales (the third way); these are given in Table 1. Note that these scales consist of two complementary concepts (restless – calm; fast – slow, strong – weak, etc.). As an exercise, you may make up your own mind on prelude 22, which will now be played by Casper de Jonge [*organ example*¹⁰]. You may hand in your answers next time in the data box provided.

Table 1. Semantic differentials
(English translation of the Japanese terms;
scores from 1 to 7)

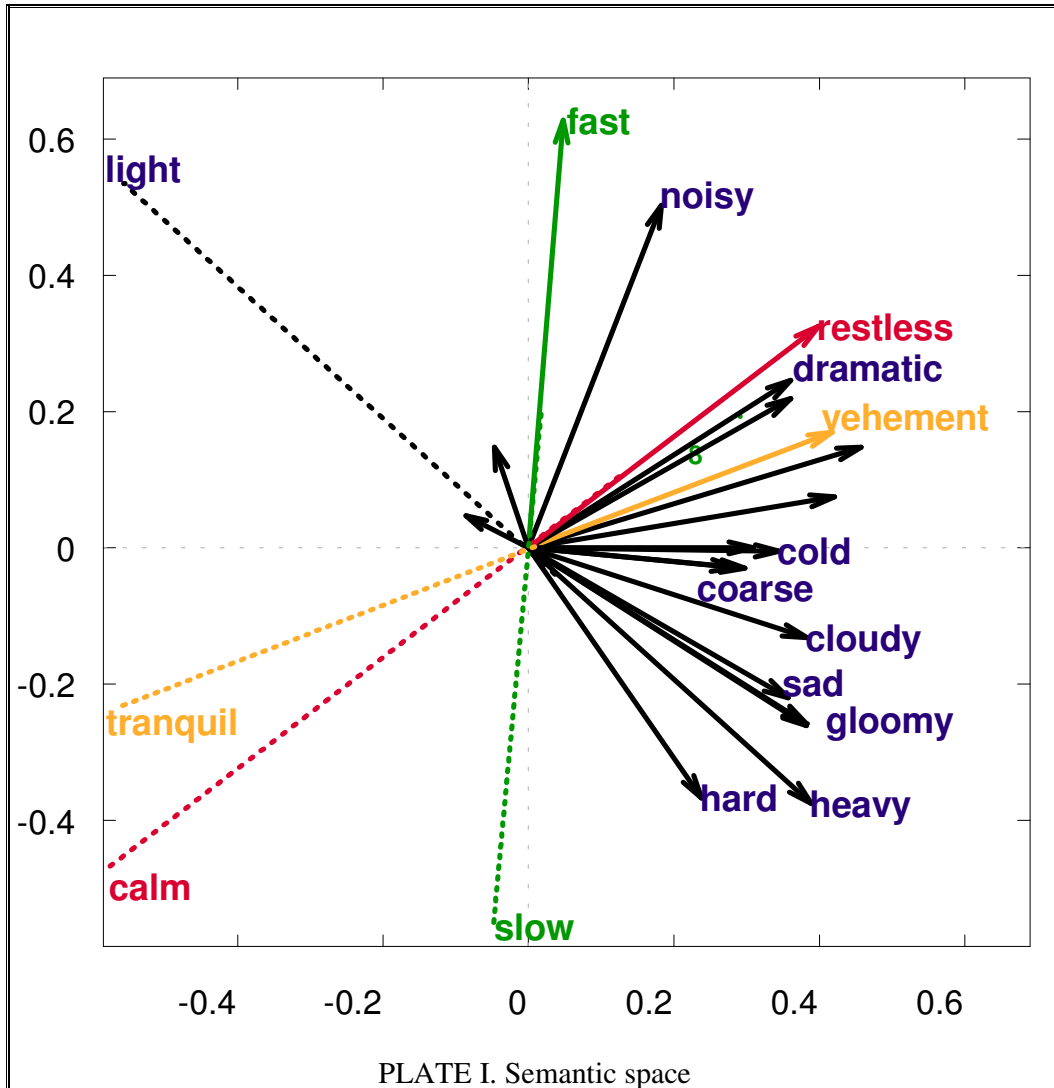
Calm (= 1)	-	Restless (= 7)
Gentle	-	Severe
Quiet	-	Noisy
Lyrical	-	Dramatic
Tranquil	-	Vehement
Weak	-	Strong
Slow	-	Fast
Still	-	Loud
Light	-	Heavy
Cheerful	-	Gloomy
Bright	-	Dark
Soft	-	Hard
Happy	-	Sad
Clear	-	Cloudy
Warm	-	Cold
Small	-	Large
Delicate	-	Coarse
Thin	-	Thick
Unattractive	-	Attractive
Uninteresting (= 1)	-	Interesting (= 7)

What we are trying to find is the connection between the preludes and the rating scales as related to the individual differences between students. The main result of the investigation was that the students largely agreed on the technical, cognitive aspects of the preludes, but differed as to the affective elements, i.e., the sort of music they preferred. I will now explain this in more detail.

Consensus on the musical-technical aspects of the preludes

Scales

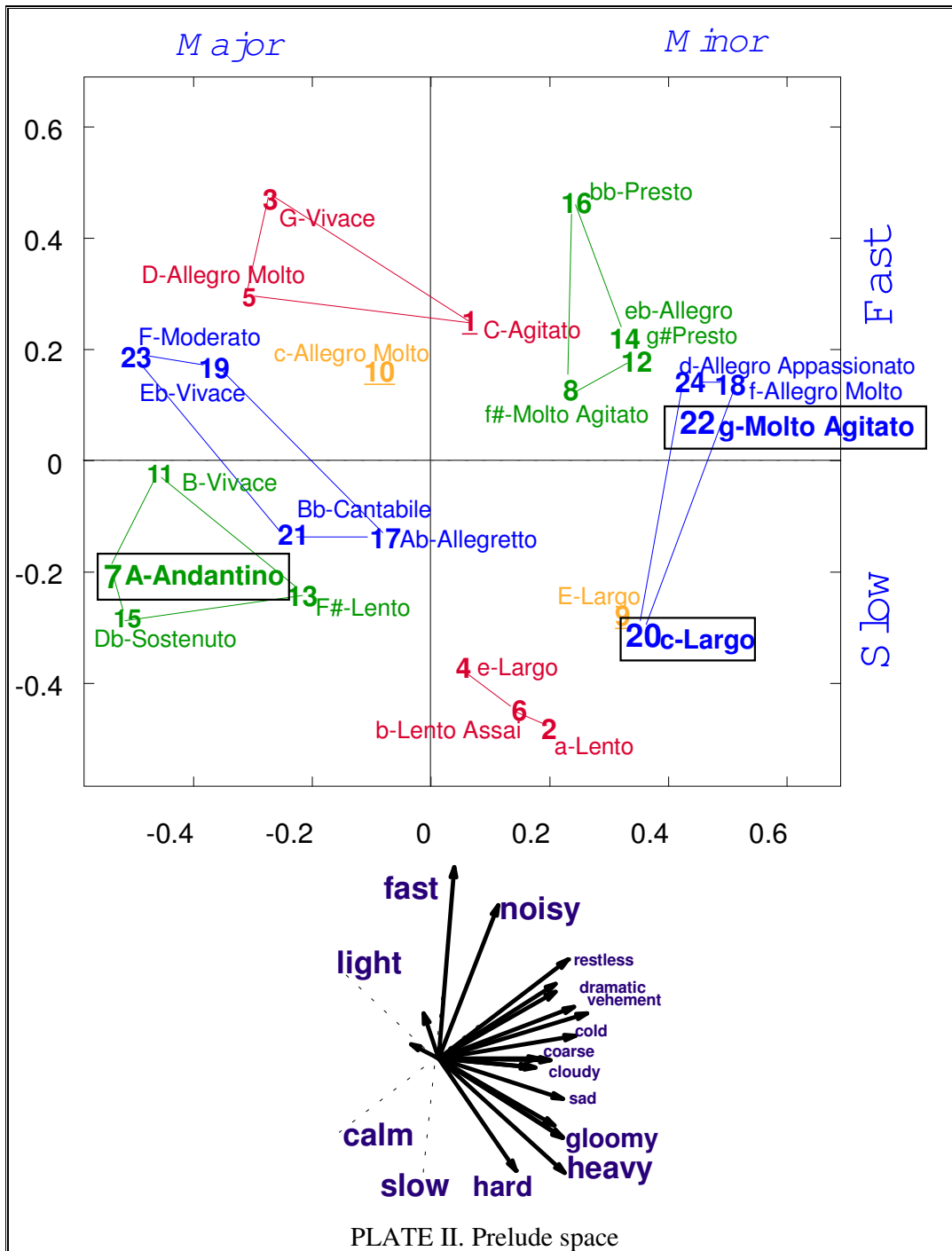
Let us look at the *Semantic Space* (Plate I), in which we see the relations between the twenty bipolar scales as reflected in the students' assessments. Every scale is represented by an arrow, with one adjective at the tip (*restless*) and its antonym on the other side (*calm*). In the interest of clarity there are only four arrows of which both poles have been drawn. What does it mean that some arrows are close together, and that others are at wide angles to each other? Let us first look at the arrows marked *restless* and *dramatic*. These are close together, because students generally gave preludes similar scores on these two scales. Hence, *restless* preludes are also *dramatic*, and *calm* preludes are also *lyrical*. When the arrows are at right angles to each other, such as for instance *fast* and *cold*, this means that according to the students those scales have no relation to each other at all: *fast* preludes may be hot as well as cold, and the same applies to *slow* preludes.



Preludes

Plate II, the *Prelude Space*, shows the configuration of the preludes as a reflection of the scores the students assigned to them. At the bottom of the Plate, I have given the scales again; for a correct interpretation you should imagine them as an overlay on top of the preludes in the figure.

We can now see that the prelude you just scored, no. 22 in g, tempo indication *molto agitato*, was judged especially loud and vehement. Prelude no. 20 in c, *largo*, you will find in the bottom right-hand corner of the Prelude Space. The students judged this prelude coarse, heavy, sad and gloomy [*organ example*]. Prelude no. 7 in A, *andantino*, in the bottom left-hand corner of the Prelude space, was rated calm, lyrical and tranquil [*organ example*]. Unfortunately, I cannot treat you to really fast preludes, as these are unsuitable for the organ. In this respect you might like to know that preludes 4, 6 and 20, which are close together in the Prelude Space, were all three played on organ at Chopin's funeral.¹¹



Technical merits

If we now judge the configuration of the preludes in the Prelude Space on their technical merits, we note that the students have made a clear division in Tempo (fast and slow), and in Mode (major and minor). In other words, on the basis of the semantic scales the students have arranged the preludes in a pattern that is found to correspond to an arrangement based on musical-technical aspects. This is not all, however; closer inspection reveals that the arrangement of key signatures over the area largely corresponds to that of the circle of fifths (see Figure 5). We now even

note two anomalies. These are nos. 9 and 10: 9 is in a major key and is situated in the ‘minor’ area, and 10 is in a minor key and is found in the ‘major’ part of the Prelude Space. For the moment, I will leave aside the question why these preludes received such deviant assessments.¹²

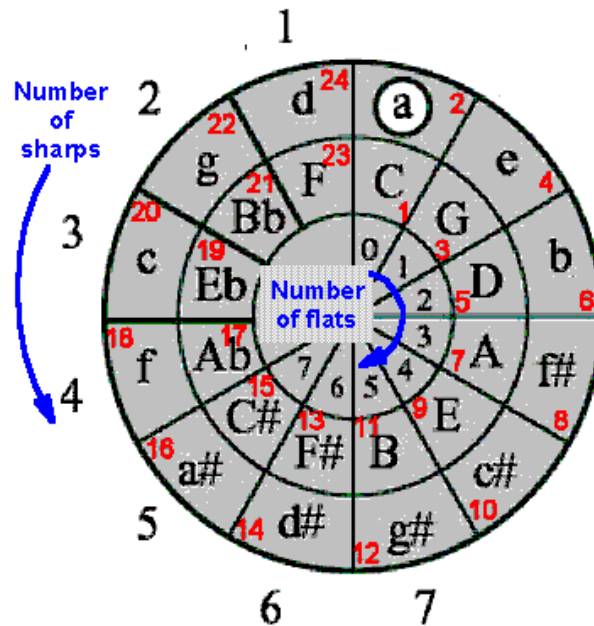


Figure 5. Circle of fifths

Note to Figure 5:

The numbers on the outside refer to the number of flats in the key in question; those in the innermost circle to the number of sharps. The numbers in the boxes are those they have in Chopin’s score.¹³

Where the students differed

What still needs to be discussed is to what extent the students *differed*. I have no figure to illustrate this, because it may easily be said in words. The students especially liked either fast pieces in a major key, or slow pieces in a minor key. Of course, you knew this all along. Don’t worry: you have not been set up for a trick *à la* Köbben.

The history of three-mode analysis

The history of techniques for three-mode data analysis starts with the late Ledyard R Tucker (he died very recently, in 2004), who, among other places, worked at the University of Illinois and the Educational Testing Service in Princeton. Another prominent godfather is Douglas Carroll, who spent a large part of his academic career at Bell Laboratories and is now on the staff at Rutgers University. A third founding father of three-mode techniques is Richard Harshman, permanently based at the University of Waterloo in London, Canada. In this lecture I will concentrate on Tucker’s work because it has been the basis for practically all of mine.

Although of course mathematics is the mother of all sciences and (mathematical) statistics her daughter, a mathematician might still be surprised to see how a changeling discipline such as psychometrics, the science of quantification in psychology, has given rise to so many important statistical techniques and innovations: not only factor analysis, multidimensional scaling techniques, item-response techniques, structural equation models, but also three-mode analysis.

How I would have loved to have been a student at the University of Illinois at the beginning of the 1960s, when I, all innocence and ignorance, was just starting grammar school. The Department of Psychology must have been a breeding ground of brilliant ideas. Almost concurrently, the semantic differential was developed by Charles Osgood,¹⁴ the behavioural differential by Harry Triandis,¹⁵ and three-mode analysis by Ledyard Tucker.¹⁶ Many of their colleagues produced other important contributions to psychometrics, and to psychology itself. Someone with inside knowledge of what was happening there could write a successful academic novel -- *Many Beautiful Minds*.

How did three-mode analysis originate?

In an interview with Neil Dorans in 2004¹⁷ Tucker had the following to say on this topic:

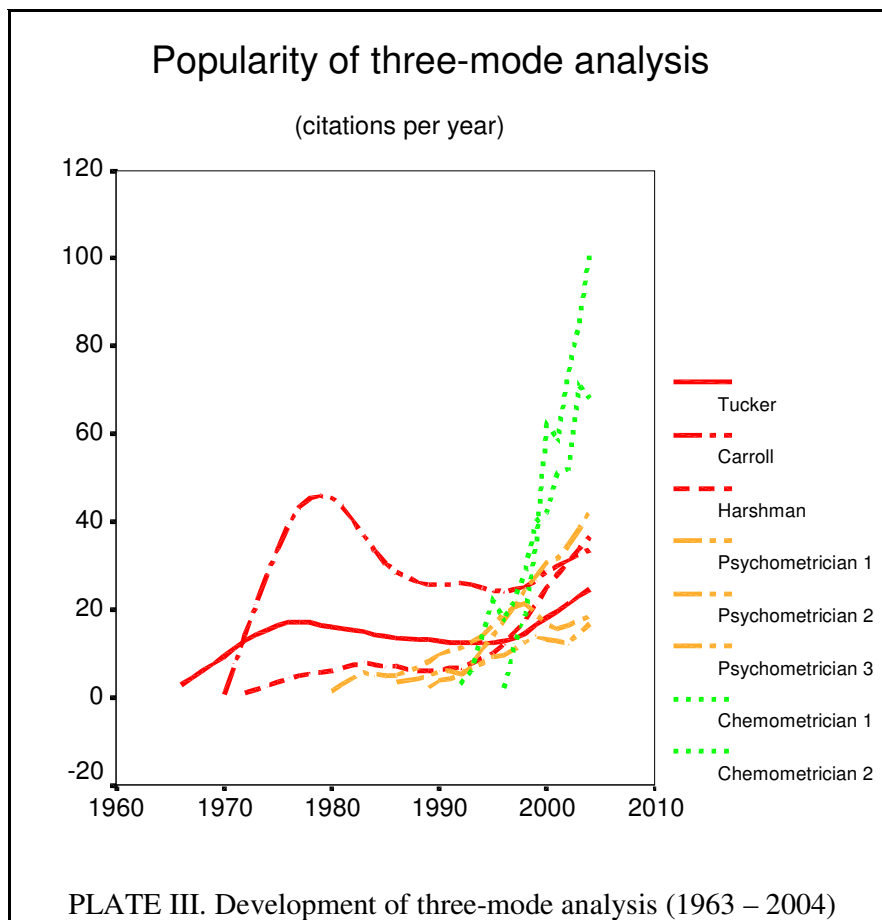
Three-mode factor analysis grew out of this multidimensional scaling work. While I was at ETS, I had observed that Charles Osgood of the University of Illinois had collected data from three modes—concepts, scales, and subjects—in his semantic differential research. I thought that the data should be analyzed differently than it was. He collapsed over people and threw away individual differences data [i.e., he flattened his data box]. So I developed the 3-mode factor analysis approach, a very general model for evaluating individual differences data. It was able to deal with the variety of people (p. 8).

Tucker illustrated his objection to ‘flattening’ with the example of the car manufacturer Chrysler, who designed a car for the ‘average customer’ without appreciating the importance of individual differences. Unfortunately, the car did not sell, because there was no such thing as the *average* customer, and Chrysler almost went bust on this enterprise. The necessity of paying attention to individual differences is still the guiding principle for almost all work in the area of three-mode analysis.

What is the current status of three-mode analysis?

A full and detailed survey of the development of three-mode analysis would be out of place in this lecture, all the more so because I would have to go into the technique more deeply than I intend to do today. However, in order to give you at least some idea of its fortunes over the years I have drawn a diagram (Plate III) showing the roughly calculated citation curves for the main protagonists, divided into founding

fathers (Tucker, Carroll, and Harshman), psychometricians, and chemometricians. The last two groups consist almost exclusively of scientists from the Netherlands. This representation has not been prompted by chauvinism: it is a fact that during the 1980s and 90s various Dutchmen have done pioneering work in this area.



What is most striking in this diagram is the steep increase in the number of citations during the 1990s, and the fact that this escalation is almost exclusively due to the chemometricians. Moreover, the rising number of citations for the psychometricians is actually also due to the stormy developments in chemometrics. As soon as chemometricians realised the power of three-mode analysis, especially the models developed by Richard Harshman, there was no stopping them. One of the primary causes for this success was the fact that in (analytical) chemistry models were used that already had the form of three-way models. All that was needed was to estimate the parameters, instead of finding out whether the model was applicable and if so, in which form. Also, important commercial applications immediately offered themselves.¹⁸

An interesting aspect of this development is the fact that psychometricians also started to publish in chemometrics journals, because that was where their biggest customers were located.¹⁹ Does this mean that as from now, three-mode analysis is exclusively a matter for chemists? Certainly not. Applications in other branches of

science, such as agriculture, signal processing, medicine, and mathematics, regularly appear, not in overly great numbers but here, too, there is a growing interest.²⁰

Have the social and behavioural sciences now become irrelevant? Is there still a use for three-mode methods in these disciplines? Can't we just leave all the hard work to others, and simply use their techniques? Is it really necessary for the social and behavioural sciences to have a chair of multivariate analysis, in particular of three-way data? As you would expect, my answer is an emphatic "Yes!".

It is exactly the establishment of this chair that will make it possible to accommodate the special character of data in the social and behavioural sciences. As my closest colleagues in the field of data theory never cease to emphasize, many of these data are characterised by being categorical rather than continuous, unlike data in most scientific disciplines. The Leiden Data Theory Group has made important contributions towards the analysability of large collections of categorical data, and of course it will be my job to continue this line in the field of three-mode analysis.

Another long-established trend is the design of large-scale longitudinal studies of the development of children. These are not only carried out overseas, but also here in Leiden, in particular in my own department of Child & Family Studies and Data Theory. Here, too, researchers are confronted with complex sets of multivariate data, often of varying measuring levels and with large differences between (groups of) individuals. Three-mode analysis and longitudinal data analysis with serious consideration for individual differences were made for each other. In this respect, too, I am able to lean on work carried out in the Netherlands, and in this respect I especially want to mention my Groningen colleagues.²¹

We may conclude from this that an intensive focus on three-mode analysis, especially directed at data that are characteristic for the social and behavioural sciences, will be of crucial importance for the substantive development of those sciences and of data analysis itself.

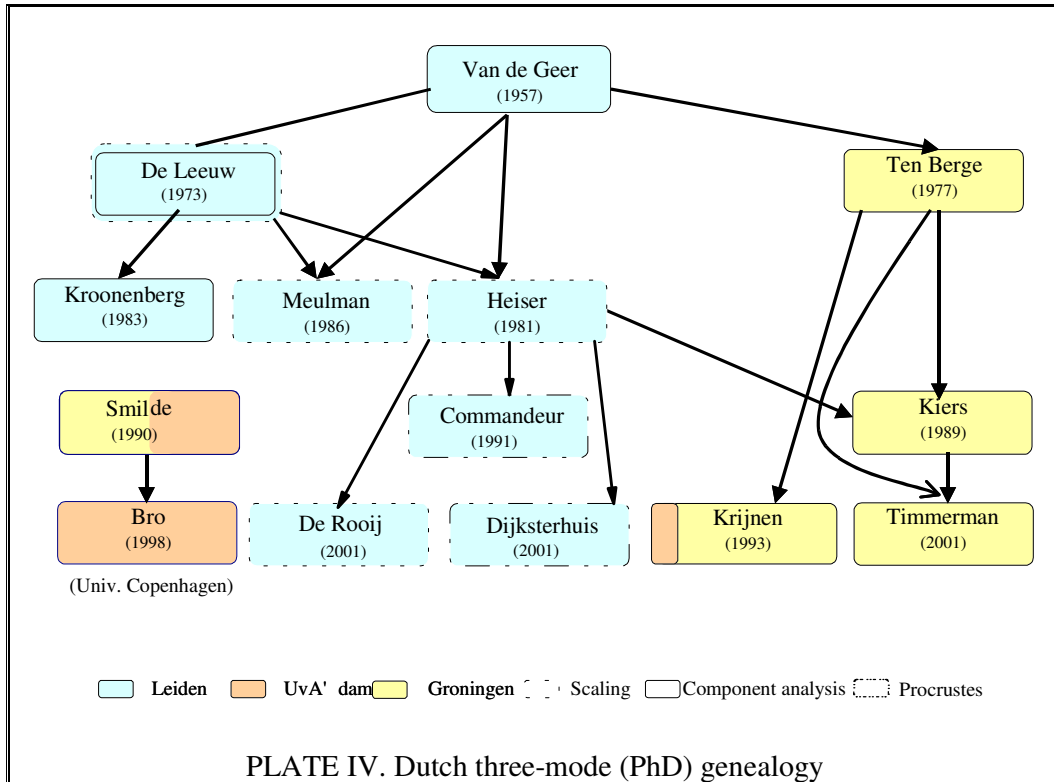
How did the Netherlands come to occupy such a prominent position within three-mode analysis?

Professor Van de Geer, the initiator of the Foundation that established this chair, was the great pioneer of data analysis in the Netherlands, and especially in Leiden. His daughter, Professor Sara van de Geer, put this in her inaugural lecture five years ago as follows: "Data Theory in Leiden is your brain child".²²

In addition, Van de Geer was also the founding father of three-mode analysis in the Netherlands. This sounds a bit like 'world-famous on the Rapenburg', but it is not intended that way. Van de Geer wrote a couple of texts on three-mode analysis, which unfortunately have never been published, but which helped me as a student to understand what it was all about.²³

To illustrate Van de Geer's impact on three-mode analysis in the Netherlands I have constructed a 'Ph.D. genealogy' (Plate IV), which clearly shows him as its nestor. I will not discuss the diagram in detail here, but just point out that he stood at the basis

of both the Leiden and the Groningen schools. The lateral connections show that there actually is such a thing as a Dutch school of three-mode analysis. I could have made a similar diagram of persons who have published articles together, but that would have resulted in an impossible tangle of lines. If you think this is all rather incestuous, I beg to differ. It is exactly this intensive cooperation that has stimulated the current developments. Especially the structure offered by the Inter-University Research School for Psychometrics and Sociometrics (*Interuniversitaire Onderzoekschool voor de Psychometrie and Sociometrie, IOPS*) has proved very fruitful, and demonstrates the strength of such national research schools rather than the increasingly popular local variants.



When as a student of applied mathematics I decided I wanted to create mathematics for a change, instead of just reproducing it, I asked one of my lecturers for a topic on which to write my M.A. paper. If I understood him correctly, he was a bit surprised at my request and answered (this was 1975) that I surely could not expect them to have a topic ready on the shelf for every interested student. I then went to Jan de Leeuw, my boss in my function as a student assistant, who thrust some sheets of graph paper at me, each containing a few formulas related to a ‘ping-pong’ algorithm for three-mode analysis²⁴. The rest is history – I have never been able to let this topic rest, and here I am addressing you today.

Word of thanks

In the third and last part of my lecture I would like to thank those who have contributed to the fact that I am standing here before you.

Those who contributed to the establishment of the chair

First, of course, thanks are due to all those who contributed to the establishment of this chair: especially the Board of the Foundation for the Advancement of Data Theory as a Science (*Stichting ter Bevordering van de Wetenschap der Datatheorie*) under the inspiring leadership of Professor Willem Heiser, and of course the Board of this university, which agreed to establish the chair in the Faculty of Social and Behavioural Sciences. The Curatorium of the chair, particularly professors Heiser, Meulman, Smilde and Ten Berge, will keep an eye on my activities. I do not see them as a group of schoolmasters intent to keep me focused on my work, as much as a group of colleagues who will stimulate my attempts to further expand three-mode analysis, and make it a useful tool for substantive disciplines especially in the social and behavioural sciences.

Those who contributed to my data-theoretical education

One of the most important people in my development as a statistician has been Albert Verbeek, who died far too young: friend, mentor, and colleague. He taught me what statistics is really about, and I am still trying every day to pass on his insights to my students.

As I mentioned earlier, most of my training in the area of multivariate analysis, especially of three-way data, I owe to Jan de Leeuw in person, and Professor Van de Geer through his written texts.

Professor Meulman, dear Jacqueline, in the field of research we have until now worked together on only a modest scale, but all the more on Ph.D. courses.²⁵ I hope that in the future we will find, especially in the area of optimal scaling and three-mode analysis, a good symbiosis that may lead to a number of elegant techniques, attractive applications and interesting publications. The only attempt made so far in that direction was that of Richard Sands, a student at the University of North Carolina.²⁶ He is now Chairman and Chief Executive Officer of Constellation Brands, the world's largest wine company. Not bad, but of course not a patch on the delights of three-mode analysis.

Those who contributed to my development as a social scientist

When I applied for a post in the Department of Child and Family Studies in 1975, I was given preference over another candidate because, as I heard later, the selectors thought it would be easier for a statistician to train himself in behavioural sciences than the other way round. Whether I have satisfied this expectation I leave to others to judge, but I have to say, possibly to my disgrace, that after roughly thirty years I still feel more a statistician than a social scientist. I am still the odd one out, but by now have come to look upon the Department as a home away from home.

Professor Van IJzendoorn, dear Rien, one of the things I have always particularly valued is your stimulating attitude towards me and the leeway you have given me to

give my specialism free rein. You had, and still have, great patience with my analyses of cotton crops,²⁷ peanut plants,²⁸ cocaine molecules,²⁹ and sick blue crabs.³⁰ Sometimes you must have had the unhappy feeling that I analysed anything as long as the data fitted into a box and did not have anything to do with the social and behavioural sciences. Fortunately you regularly managed, by putting nice data sets about attachment in my way, to prevent me from wandering off at a tangent.³¹ It is largely thanks to you that I have never felt an urgent need to take my three-mode business permanently elsewhere. This seems to me a typical example of secure attachment and "secure-base behaviour". My many shorter and longer explorations abroad always took me back sooner or later to my *alma mater*.

Students: What you may expect of me

Dear students of the Department of Education and Child Studies. In your daily practice of being a student you have not had much to do with my specialism, and I am afraid that in the future this will not change. This is partly because the data that you work with, rarely need boxes. What I can offer you, is my expertise in handling all sorts of data, three-way or otherwise. The fact that an enormous variety of data has passed through my hands over the past years will, I expect, prove useful when you have arrived at the stage of your Master's project and wonder, perhaps desperately, whatever you are going to do with the massive amounts of data you have collected..³²

Professor Van de Geer said in his inaugural lecture of 1964 that "in lectures such as these, on average 2.01% of the total number of words is addressed directly to [students], a percentage whose meagreness contrasts sharply with daily practice" (p. 26).³³ I have to confess that in this lecture, I, too, whether unintentionally## or unknowingly, have not progressed much beyond this sad percentage.

Family: The link with the past that ensures the future

It is sad that neither of my parents have lived to see this ceremony. I also regret that my mother's brother, Lipke Holthuis, is for health reasons unable to be here. He has to a large extent contributed to my academic career, not only in his capacity as a scholar, but also as a rich uncle during my lean student years. In addition, his example has inspired me with an irrepressible urge to go out and explore the world. Fortunately, of my parents' generation we have my father's brother Kees and his wife Lydia in the audience. However, the youngest generation is present in great number, and I am hoping that some day they may get around to putting data in boxes.

If they do, who knows whether four- and multi-way boxes may not have become their staple diet. These hyperboxes have been around for a while, and are being extensively explored, but on that topic I hope to have the opportunity to speak to you some other time.

Thank you.

Notes

- ¹ The illustrations in that booklet have all been included in this English version of the text.
- ² This example refers to various three-way applications in agriculture, see for instance: Basford, K. E., Kroonenberg, P. M., & DeLacy, I. H. (1991). Three-way methods for multiattribute genotype by environment data: An illustrated partial survey. *Field Crops Research*, 27, 131-157.
- ³ An introductory survey containing a clear explanation of the Parafac model, with applications in chemistry, may be found in: Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38, 149-171.
Smilde, A.K., Bro, R., & Geladi, P. (2004). *Multi-way analysis. Applications in the chemical sciences*. Chichester, UK: Wiley.
- ⁴ This example is imaginary, but based on the work carried out in a research project 'Christianity and Modernity' at the Leiden Faculty of Theology under the direction of Prof. Dr. E. G. E. van der Wall.
- ⁵ Kroonenberg, P. M., Dunn III, W. J., & Commandeur, J. J. F. (2003). Consensus molecular alignment based on generalized Procrustes analysis. *Journal of Chemical Information and Computer Science*, 45, 69-97.
- ⁶ Sidiropoulos, N. D., Bro, R., & Giannakis, G. B. (2000). Parallel factor analysis in sensor array processing. *IEEE Transactions on Signal Processing*, 48, 2377-2388.
- ⁷ Köbben, A.J. F. (1980). Het heilig vuur. *Intermediair*, 16 (49), 1-7 (also mentioned in his lecture upon his accession to the Cleveringa chair at Leiden University, under the same title). The trick has probably been taken from P.F. Lazarsfeld (1949). "The American Soldier". An expository review. *Public Opinion Quarterly*, 13, 377-404 (esp. p. 380). This article is quoted in J. Goudsblom (1974). *Balans van de sociologie*. Utrecht: Het Spectrum (Aula pocket no. 518). I am grateful to Prof. Leo van der Kamp for these references.
- ⁸ Kiers, H.A.L. (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics*, 18, 22-36.
- ⁹ Murakami, T., & Kroonenberg, P.M. (2003). Individual differences in semantic differential studies and their analysis by three-mode models. *Multivariate Behavioral Research*, 38, 87-96.
- ¹⁰ During the lecture, (parts of) Chopin's Preludes 22, 20 and 7 were played by University organist Drs. C. C. de Jonge, with slight adaptations by himself. None of the examples was longer than 1 minute.
- ¹¹ Source (seen: 28/8/2005): <http://www.geocities.com/darnbi/english/prel.htm>. About Prelude Nr.4 : "It was fittingly played, with Nos. 6 and 20, by the famous organist Lefébure-Wely, at Chopin's funeral service at the Madeleine Church in Paris".
- ¹² The probable cause is that in some longer preludes key changes occur, and the students were offered only the first 80 seconds, which may have given them a distorted impression of these preludes.
- ¹³ Adapted version of the original design, which can be found at <http://www.uncletim.com/answers/circle.htm> (seen: 19/7/2005); according to the website derived from Tim Gillespie, *Uncle Tim's Building Blocks*, Fifth printing 2004, ISBN 0-9647059-7-4
- ¹⁴ Osgood, C.E., Suci, G. J., & Tannenbaum, P. *The measurement of meaning*. Urbana, IL: University of Illinois Press.
- ¹⁵ Triandis, H. C. (Ed.) (1972). *The analysis of subjective culture*. New York: Wiley
- ¹⁶ Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279-311.
- ¹⁷ Dorans, N. (2004). *A conversation with Ledyard R Tucker*. Princeton: Educational Testing Service (www.ets.org/research/dload/tucker.pdf; seen 6/9/2005; see also the obituary by Dorans in *Journal of Educational and Behavioral Statistics*, 29, 377-378).
- ¹⁸ See the book by Smilde, Bro en Geladi (note 3, above).
- ¹⁹ Especially Kiers, Ten Berge and Harshman, and recently also Kroonenberg, have published several articles in the *Journal of Chemometrics* and *Chemometrics and Intelligent Laboratory Systems*.
- ²⁰ There are already two specialist, small-scale conferences focusing on three-way and multi-way analysis: TRilinear methods in Chemistry And Psychology (TRICAP), since 1993, and the Tensor workshops, since 2004.
- ²¹ The Groningen researchers Ten Berge and Kiers have been awarded, together with the Leiden researchers Heiser and Kroonenberg, a five-year grant for the project: "Multi-way analysis of multivariate longitudinal data" (1996-2000), which has resulted in dissertations by Timmerman and De Rooij, plus several articles on this subject.

-
- ²² Van de Geer, S. A. (2000). *Een zeker toeval*. Rede uitgesproken ter gelegenheid van de aanvaarding van het ambt van hoogleraar in de Kansrekening en Statistiek aan de Universiteit Leiden (inaugural address, December 1, 2000).
- ²³ Van de Geer, J. P. (1974). Toepassing van drieweganalyse voor de analysis van multiple tijdreeksen.[Application of three-mode analysis to multiple time series]. In C. J. Lammers (Ed.), *Groei en ontwikkeling van de ziekenhuisorganisaties in Nederland* [Growth and development of Dutch hospital organisations]. Internal report, Institute of Sociology, Leiden University.
- ²⁴ Kroonenberg, P. M., & De Leeuw, J. (1977). *TUCKALS2: A principal component analysis of three mode data*. Research Bulletin No. RB 001-' 77, Department of Data Theory Leiden University, Leiden, the Netherlands. (A slightly edited version of my Master's thesis).
- ²⁵ I am referring in particular to the course "Applied multivariate data analysis", which was taught several times in the Ph.D. program by Willem Heiser, Jacqueline Meulman and myself, to the delight of participants and instructors alike.
- ²⁶ Sands, R., & Young, F. W. (1980). Component models for three-way data: ALSCOMP3, an alternating least squares algorithm with optimal scaling features. *Psychometrika*, 45, 39-67.
- ²⁷ Basford, K.E., Kroonenberg, P.M., DeLacy, I.H., & Lawrence, P.K. (1990). Multiattribute evaluation of regional cotton variety trials. *Theoretical and Applied Genetics*, 79, 225-234.
- ²⁸ Kroonenberg, P.M., Harch, B.D., Basford, K.E., & Cruickshank, A. (1997). Combined analysis of categorical and numerical descriptors of Australian groundnut accessions using nonlinear principal component analysis. *Journal of Agricultural, Biological, and Environmental Sciences*, 2, 294-312.
- ²⁹ See note 5.
- ³⁰ Kroonenberg, P. M., Basford, K. E., & Gemperline, P.J. (2004). Grouping three-mode data with mixture methods: The case of the diseased blue crabs. *Journal of Chemometrics*, 18, 508-518.
- ³¹ "Van IJzendoorn, M.H., & Kroonenberg, P.M. (1988). Cross-cultural patterns of attachment: A meta-analysis of the Strange Situation. *Child Development*, 59, 147-156" is without doubt our most successful enterprise so far, with "Van IJzendoorn, M. H., Kroonenberg, P. M., Out, D., Randsdorp, Y., Lehmann, A., Van der Maas, H. (submitted). Does more non-maternal care lead to aggression? The NICHD Study of Early Child Care and Youth Development on quantity of non-maternal care and aggression" as the most recent product of our long and productive cooperation.
- ³² A year and a half ago, the Department of Education and Child Studies set up a Thesis Lab, which offers students in the Master's programme the opportunity to work on their thesis or Master's project, and benefit from advice from experienced tutors on writing a thesis and everything this involves, and on the processing and analysis of their data.
- ³³ Van de Geer, J. P., *Kans en Onzekerheid*. Rede uitgesproken bij de aanvaarding van het ambt van gewoon hoogleraar in de Experimentele Psychologie en de Statistische Methode in de Psychologie aan de Rijksuniversiteit te Leiden (inaugural address, May 1, 1964), p. 26.